A novel attention mechanism for noise-adaptive and robust segmentation of microtubules in microscopy images

Achraf Ait Laydi[1,2 [0009-0004-8898-1903],] Louis Cueff[1 [0009-0003-6798-0920]], Mewen Crespo[1,3 [0000-0003-3834-9462]], Yousef El Mourabit[2 [0000-0001-7851-3816]] and Hélène Bouvrais[1 [0000-0003-1128-1322]]

[1] CNRS, Univ. Rennes, IGDR (Institut de Génétique et Développement de Rennes), UMR6290, 35000 Rennes, France
[2] TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay Slimane Univ., Morocco.
[3] IRMAR – UMR 6625, 263 avenue du Général Leclerc, 35042, Rennes, France (current affiliation).

Abstract

Background: Segmenting cytoskeletal filaments in microscopy images is essential for studying their roles in cellular processes such as cell division and intracellular transport. However, this task is highly challenging due to the fine, densely packed, and intertwined nature of these structures. Imaging limitations—noise, low contrast, and uneven fluorescence—further complicate analysis. While deep learning has advanced segmentation of large, well-defined biological structures, its performance often degrades under such adverse conditions. Additional challenges include obtaining precise annotations for curvilinear structures and managing severe class imbalance during training.
Results: We introduce a novel noise-adaptive attention mechanism that extends the Squeeze-and-Excitation (SE) module to dynamically adjust to varying noise levels. Integrated into a U-Net decoder with residual encoder blocks, this yields ASE_Res_UNet, a lightweight yet high-performance model. To address annotation challenges, we developed a synthetic dataset generation strategy that ensures accurate annotations of fine filaments in noisy images, producing a synthetic dataset with two difficulty levels for segmentation benchmarking. We systematically evaluated loss functions and metrics to mitigate class imbalance, ensuring robust performance assessment. ASE_Res_UNet effectively segmented microtubules in noisy synthetic images, outperforming its ablated variants. It also demonstrated superior segmentation compared to models with alternative attention mechanisms or distinct architectures, while requiring fewer parameters, making it efficient for resource-constrained environments. Evaluation on a newly curated real microscopy dataset and a recently reannotated dataset highlighted ASE_Res_UNet's effectiveness in segmenting microtubules beyond synthetic images. For these datasets, ASE_Res_UNet was competitive with a recent synthetic data-driven approach that shares two cytoskeleton pretrained models. Importantly, ASE_Res_UNet generalised well to other curvilinear structures (blood vessels and nerves) across diverse imaging conditions.
Conclusions: This work advances microtubule segmentation through three key contributions: (1) Providing two benchmark datasets (synthetic and real), addressing a critical gap in standardised evaluation resources for this task; (2) Introducing ASE_Res_UNet, a lightweight yet robust model combining noise-adaptive attention with residual learning; (3) Validating competitive performance across synthetic and real microscopy data. Additionally, we demonstrated generalisation to diverse curvilinear structures, showcasing potential for broader applications in biological research and medical diagnosis.

Keywords
Microtubule segmentation, Noise-adaptive attention, Microscopy fluorescence images, Low-contrast and noisy image segmentation, Synthetic image datasets, Generalisation to vessels and nerves.

Corresponding author
Hélène Bouvrais: helene.bouvrais@univ-rennes.fr

1 Background

Microtubules are essential cytoskeletal filaments involved in a wide range of cellular processes, including tissue modelling, cell division, migration, morphogenesis, and intracellular transport [1-5]. Besides, they serve as therapeutic targets in several diseases, such as cancers and neurodegenerative disorders [6-10], making them a central focus of both fundamental and translational research. Structurally, microtubules are dynamic, semi-flexible filaments composed of tubulin dimers that assemble into a cylindrical structure with a diameter of 25 nm and lengths reaching several tens of micrometres [11]. Their dynamic and mechanical behaviours, such as growth, shrinkage, and bending, are governed by intrinsic properties and regulated by extrinsic factors, including microtubule-associated proteins (MAPs) [12]. To investigate microtubule functions, dynamics, and mechanics in cells, fluorescence microscopy is widely used [13]. In live cells, microtubules can be visualized using fluorescently tagged proteins (e.g., EB proteins to label plus-ends or fluorescent tubulin to label the lattice), while fixed cells are typically stained using anti-tubulin antibodies. Tracking microtubule plus-ends using EB proteins is a well-established method [14, 15] that has enabled detailed characterization of microtubule dynamics and led to the discovery of numerous regulatory MAPs at microtubule extremities [16, 17]. In contrast, the microtubule lattice has received less attention. Recent works, however, have highlighted its regulatory role in processes such as tubulin self-repair and post-translational modifications, which can influence key mechanical properties like flexural rigidity [18-21]. It calls for additional investigations to discover MAPs regulating these processes.

Quantitative analysis of microtubules often requires accurate segmentation to extract geometric and topological features, such as curvature, orientation, and length. However, segmenting microtubules in fluorescence microscopy images remains a major challenge due to several confounding factors. These include high image noise, low contrast, intensity inhomogeneity, and background artifacts. Furthermore, microtubules appear as thin, curved, and sometimes discontinuous filaments, often forming dense, overlapping networks with intensity variation along their length, which complicate their precise delineation. These challenges are exacerbated in live-cell imaging, where low-light acquisition conditions (to avoid phototoxicity) result in low signal-to-noise ratio (SNR) images [22, 23]. Despite the importance of microtubules, relatively few computational tools are available for their full-length segmentation [24]. Existing methods are often multi-step, tailored to specific datasets or applications, require hand-crafted parameter adjustments, and can be difficult to adapt to different microtubule networks or imaging conditions [25-31]. Interestingly, microtubules share structural similarities with other biological curvilinear structures, such as blood vessels, fibrils, or nerve fibres.

Early approaches to curvilinear structure segmentation relied on traditional image processing techniques, particularly filtering methods designed to enhance line-like patterns. One influential method is the vesselness filter introduced by Frangi *et al.*, based on the eigenvalues of the Hessian matrix [32]. Later enhancements included orientation- and symmetry-based filters, matched filters, and multi-scale Gabor filters [33-37]. Active contours and geodesic path-based algorithms have also been applied to delineate elongated structures by leveraging local continuity and smoothness [29, 38-42]. However, these methods depend heavily on hand-crafted parameters and often perform poorly on diverse datasets due to limited robustness and generalisability.

The rise of deep learning, particularly convolutional neural networks (CNNs), has significantly advanced biomedical and microscopy image segmentation. The introduction of U-Net by Ronneberger *et al.* was a breakthrough, enabling accurate, end-to-end segmentation with limited annotated data [43, 44]. Its encoder-decoder architecture with skip connections facilitates precise localisation while preserving contextual information. Numerous U-Net variants have since been proposed, including models adapted to curvilinear structure segmentation [45]. For instance, Residual U-Net introduces residual connections to capture long-range dependencies [46, 47], while Attention U-Net incorporates gating mechanisms to focus on informative features [48-50]. However, these deep learning advances have had only limited impact on the analysis of cytoskeletal filaments in fluorescence microscopy [31, 51-53], especially compared to their widespread application in other biomedical domains. Several key challenges may explain this gap. (i) High-quality annotated datasets are scarce due to the difficulty of labelling fine structures, particularly in live-cell images. (ii) Severe

class imbalance, where curvilinear structures occupy only a small fraction of the image, complicates model training. (iii) Many existing models are modality-specific and do not generalise well across imaging conditions. (iv) Incorporating residual or attention mechanisms improves model performance but increases parameter count and computational demand, limiting usability in resource-constrained environments. (v) Segmenting faint, noisy curvilinear structures in low-SNR fluorescence images remains inherently difficult.

To address these challenges, we presented the following contributions. (1) We propose a novel noise-adaptive attention mechanism to improve segmentation of thin structures in noisy images. (2) We introduce ASE_Res_UNet, a deep-learning architecture based on U-Net with residual blocks in the encoder and Adaptive Squeeze-and-Excitation (ASE) attention blocks in the decoder. (3) We generated original synthetic datasets of fluorescence microscopy images of microtubules (*MicSim_FluoMT*) with two difficulty levels and high-quality ground truth masks that are rarely available in real noisy microscopy images. ASE_Res_UNet accurately segmented microtubules in these synthetic images and it outperformed standard U-Net, its own architectural variants, and several architectures based on alternative attention mechanisms. (4) We demonstrated strong generalisation capabilities of ASE_Res_UNet on diverse curvilinear structures and imaging modalities, including fluorescent microtubule images (providing a novel dataset, *MicReal_FluoMT*), retinal blood vessels (*DRIVE* dataset) [54], and corneal nerves (*CORN-1* dataset) [55]. (5) We tackled the class imbalance problem by selecting suitable loss function and evaluation metrics, avoiding misleading performance estimates due to dominant background class. Overall, this work makes a significant contribution to the field of computational bioimage analysis by providing ASE_Res_UNet—a robust and generalisable deep learning framework designed for the accurate segmentation of curvilinear structures. It not only enhances the quantitative analysis of cytoskeletal filaments, thereby advancing our fundamental understanding of various cellular processes, but also holds promise for accelerating disease diagnosis by enabling precise delineation of diagnostically relevant structures in biomedical images.

2 Methods

2.1 Synthetic datasets for segmenting microtubules

To address the challenges of microtubule segmentation in noisy fluorescence microscopy images, we created two synthetic datasets of fluorescently labelled microtubules, along with corresponding binary masks (ground truth) using a two-step pipeline (details in Supplemental Material 1A). (1) *Cytosim* was used to mimic the astral microtubule network of *Caenorhabditis elegans* zygotes (Fig. S1A) [56]. (2) *ConfocalGN* converted the simulations into realistic confocal-like fluorescent images (Fig. S1A-E) [57]. This pipeline yielded two large, fully annotated datasets with distinct fluorescence intensity distributions along cytoskeletal filaments, leading to different difficulty levels for segmentation (Fig. S2) [58]. In the *MicSim_FluoMT easy* dataset, microtubules display uniform intensity along their length (Fig. S2A, D). In the *MicSim_FluoMT complex* dataset, intensity decreases toward the cell periphery, making microtubule extremities harder to distinguish from the cytoplasm background (Fig. S2B, E). Each dataset contains 1192 images with diversity in microtubule density, curvature, and crossing patterns. We used 953 images for training, 119 for validation, and 120 for testing. We also benchmarked our architecture on the recently published SynthMT dataset, which comprises 6600 synthetic interference reflection microscopy (IRM) images containing microtubules [31]. The dataset was split as follows: 5280 images (80%) for training, 660 (10%) for validation, and 660 (10%) for testing.

2.2 Real image datasets for testing model generalisation

To evaluate performance on real microscopy data, we created the *MicReal_FluoMT* dataset consisting of 49 fluorescence microscopy images of *C. elegans* zygotes, where microtubules were stained with anti-tubulin antibodies. Annotation was performed using a previously developed three-step pipeline [59] (details in Supplemental Material 1B). The dataset was split into 19 training, 10 validation, and 10 test images [60]. Some masks included annotations that were distinct from microtubules, due to non-specific staining or reflecting the annotation-pipeline limitations. We also used in-house image samples to test the model's robustness, namely

live *C. elegans* zygotes expressing GFB::TBB-2 imaged using the Nikon Nsparc microscope. Finally, we benchmarked our architecture on an independent microtubule dataset, the Higaki 2024 dataset [52], whose test split was recently reannotated to higher mask accuracy [53]. This update included 98 confocal images of *Nicotiana tabacum* BY-2 cells labelled with YFP-tubulin. We used 90 for 5-fold cross-validation and reserved 8 for segmentation predictions.

To further challenge the model's generalisation ability, we tested it on two public datasets, selected for their structural similarities to cytoskeletal filaments. The *DRIVE* (Digital Retinal Images for Vessel Extraction) dataset comprises 40 colour fundus images of retinal vessels [54], widely used in vessel segmentation benchmarks [61-63]. Challenges include low-contrast, illumination variations, complex vessel branching, and diverse vessel widths. We trained the model on 20 images, and we used 10 images for validation and 10 for testing. The *CORN-1* dataset contains 1516 confocal microscopy images of the corneal nerves with manual annotations [55] and was used in a few deep-learning based segmentation studies [61, 64]. These images pose challenges due to noise, background inhomogeneity and variability in nerve thickness and intensity. Training was conducted using 80% of the dataset, with 10% used for validation and 10% for testing.

2.3 ASE_Res_UNet architecture including a novel noise-adaptive attention mechanism.

We developed ASE_Res_UNet (Fig. 1), an U-Net variant whose originality lies in the use of a novel attention module–Adaptive Squeeze-and-Excitation (ASE)–which enables adaptation to noise [65, 66]. The placement of modules within the U-Net backbone was carefully designed to maintain a reasonable number of trainable parameters.

The encoding path consists of four down-sampling blocks, starting with 32 filters and doubling them at each stage. Each block comprises two convolutional layers, each followed by batch normalization and ReLU activation, and a residual block that incorporates a skip connection [67] (Eq. 1). The later preserves low-level spatial information and facilitates gradient flow during training, which was crucial for learning fine, elongated patterns in noisy environments. By incorporating residual units early in the network, we ensured that critical filament features were retained despite successive down sampling operations.

$$y = \mathrm{ReLU}\left(\mathrm{BN}\left(W_2 * \mathrm{ReLU}\big(\mathrm{BN}(W_1 * x)\big)\right)\right) + x \qquad (1)$$

where *x* denotes the input, *W1* and *W2* are the weights of the first and second convolutional layers, respectively, and BN refers to Batch Normalisation.

The decoding path consists of four up-sampling blocks, mirroring the encoder, which includes transposed convolution layers for up-sampling, concatenation with the corresponding encoder feature map, the Adaptive Squeeze-and-Excitation (ASE) attention module, and two convolutional layers with batch normalisation and ReLU activation (Fig. 1). We modified the traditional Squeeze-and-Excitation attention mechanism, which statically recalibrates channel-wise responses [65], into an adaptive version that dynamically adjusts attention weights based on the input's noise characteristics, making the decoding process context-sensitive.

In more detail, the noise level *N* is estimated from the input using a convolutional layer followed by a sigmoid activation (Eq. 2). Then, the noise level is averaged over spatial dimensions and batch (Eq. 3). The reduction ratio $r_{adaptive}$ is dynamically adjusted based on the averaged noise level (Eq. 4). A global spatial average pooling computes the channel-wise statistics (Eq. 5). Two fully connected layers with a ReLU and sigmoid activation compute the channel-wise weights e_c (Eq. 6). Finally, the input *X* is rescaled using e_c (Eq. 7).

$$N = Sigmoid\big(Conv2d(X)\big), N \in R^{B\times H\times W} \qquad (2)$$

$$\widehat{N} = \frac{1}{BxHxW}\sum_{b=1}^{B}\sum_{i=1}^{H}\sum_{j=1}^{W} N_{b,i,j} \qquad (3)$$

$$r_{adaptive} = max\big(1, \lceil 16 \cdot (1 + \widehat{N})\rceil\big) \qquad (4)$$

$$s_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j}, \forall c \in \{1, \dots, C\} \quad (5)$$

$$e_c = Sigmoid\left(FC_2^{(r)} \cdot ReLU\left(FC_1^{(r_{adaptive})} \cdot s_c\right)\right) \quad (6)$$

$$\widetilde{X_{c,i,j}} = X_{c,i,j} \cdot e_c, \forall c, i, j. \quad (7)$$

ASE modules were applied only in the decoder, where spatial details are reconstructed. This design should promote continuity of thin structures while suppressing background noise during up sampling.

To our knowledge, ASE_Res_UNet is the first architecture to incorporate noise-adaptive channel attention into a U-Net -based model specifically tailored for curvilinear structure segmentation.

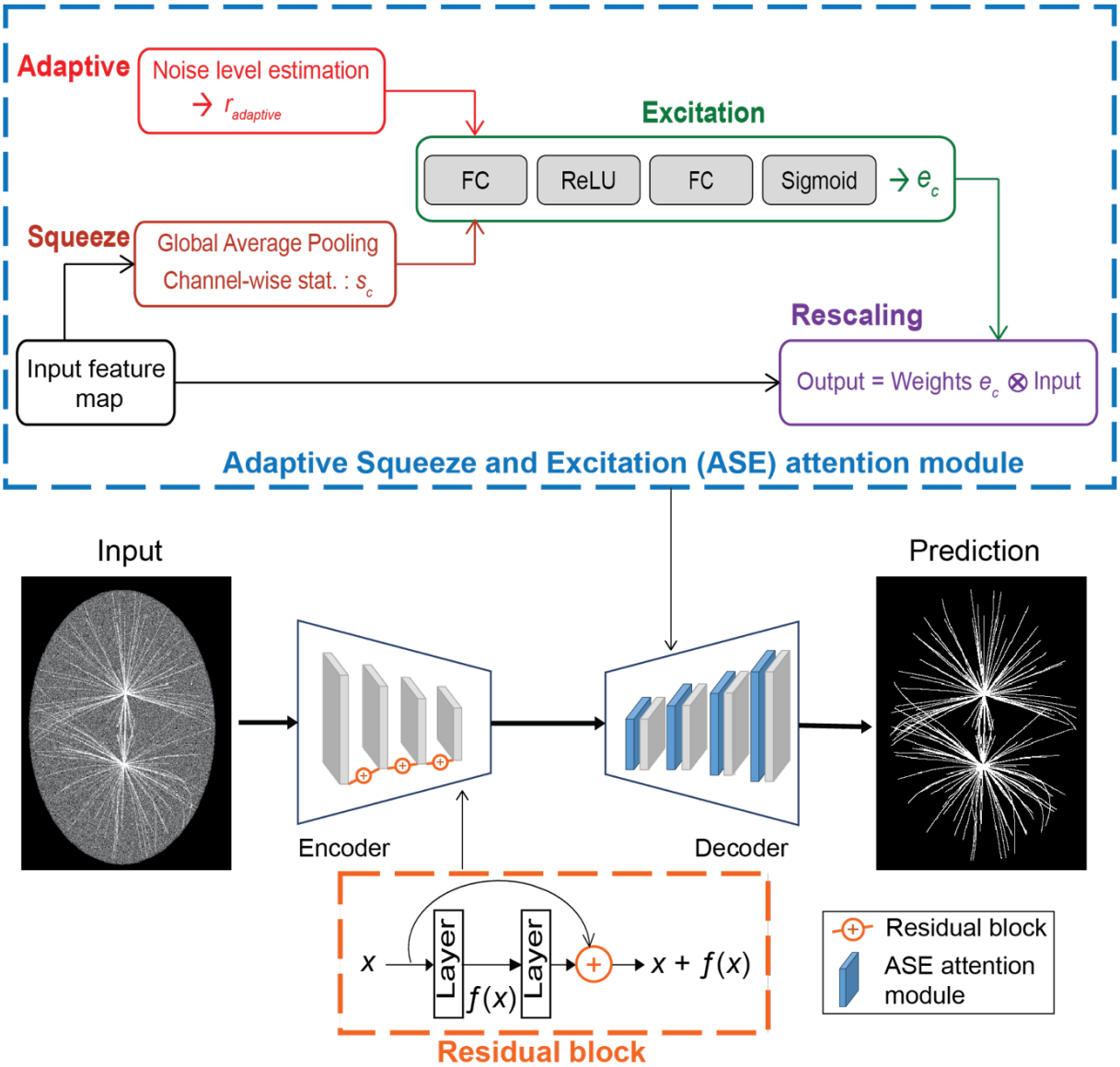

Fig. 1: ASE_Res_UNet includes noise-adaptive attention mechanisms in the decoder.
Proposed architecture of ASE_Res_UNet, including (orange) the residual blocks in the encoder and (blue) the novel Adaptive Squeeze-and-Excitation (ASE) attention modules in the decoder. Each ASE module includes: (brown) a squeeze step which uses global average pooling per channel to compute channel-wise statistics s_c, and (red) an adaptive step, which assesses the noise level to determine the reduction ratio $r_{adaptive}$. This enables (green) the computation of the channel-wise weights e_c, which are used (purple) to rescale the input map.

2.4 Handling background class dominance in metric and loss function selection

To assess the segmentation performance, we employed six quantitative metrics (Table S1), each capturing different aspects of prediction quality: Dice Similarity Coefficient, Intersection over Union (IoU), Sensitivity, Precision, Matthews Correlation Coefficient (MCC), and Area Under the Precision-Recall Curve (PR AUC) [68, 69]. These metrics were appropriate for highly imbalanced datasets [70, 71]. By combining both threshold-dependent and threshold-independent metrics, we ensured a more robust and holistic performance assessment. Statistical significance was evaluated using two-tailed Student's *t*-test with Welch–Satterthwaite correction for unequal variance. We reported significance as follows: ◊: $0.01 < p \leq 0.05$; *: $0.001 < p \leq 0.01$; **: $1x10^{-4} < p \leq 0.001$; ***: $p \leq 1x10^{-4}$; and ns (non-significant): $p > 0.05$.

To handle class imbalance, we evaluated five loss functions on the *MicSim_FluoMT complex* dataset, under identical training conditions (section 2.5): Binary Cross Entropy (BCE), Dice Loss, Focal Loss, Hausdorff Distance Loss, and Weighted Cross Entropy (WCE) [72, 73]. WCE yielded the best overall performance (Table 1). Foreground-to-background weights were empirically tuned, with 1.0 for microtubules and 0.25 for background producing the best results.

Model	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
WCE	**0.7825 ± 0.0156**	**0.9688 ± 0.0011**	**0.7345 ± 0.0233**	0.8377 ± 0.0108	**0.7762 ± 0.0148**	**0.8730 ± 0.0147**
BCE	0.7625 ± 0.0169 ***	0.9687 ± 0.0010	0.6589 ± 0.0221 ***	0.9051 ± 0.0090 ***	0.7647 ± 0.0154 ***	0.7764 ± 0.0271 ***
Dice	0.6898 ± 0.0340 ***	0.9607 ± 0.0033 ***	0.5775 ± 0.0607 ***	0.8666 ± 0.0357 ***	0.6966 ± 0.0261 ***	0.7432 ± 0.0216 ***
Focal	0.7196 ± 0.0173 ***	0.9654 ± 0.0012 ***	0.5814 ± 0.0209 ***	**0.9447 ± 0.0071** ***	0.7336 ± 0.0149 ***	0.8308 ± 0.0152 ***
Hausdorff Distance	0.7422 ± 0.0166 ***	0.9611 ± 0.0015 ***	0.7354 ± 0.0195	0.7494 ± 0.0196 ***	0.7320 ± 0.0166 ***	0.8193 ± 0.0187 ***

Table 1: The Weighted Cross Entropy loss outperforms the other loss functions on segmenting microtubules.
Microtubule segmentation performance of ASE_Res_UNet under different loss functions including the Weighted Cross Entropy (WCE) loss, the Binary Cross Entropy (BCE) loss, the Dice loss, the Focal loss and the Hausdorff loss, evaluated on the *MicSim_FluoMT complex* dataset using various metrics (mean ± standard deviation over 120 test images). Bold values indicate the best performances. Statistical differences between WCE model and other models using different loss functions are indicated only when significant (***: $p \leq 0.0001$).

2.5 Training

During training, we tuned several hyperparameters to balance effective multi-scale feature extraction with moderate computational cost. The best configuration included: an initial learning rate of 0.001, dynamically reduced based on validation performance; a batch size of 2, balancing memory efficiency and convergence speed; the Adam optimizer configured with β_1 and β_2 equal to 0.9 and 0.999; and a 3x3 filter size, chosen for its low sensitivity to noise and effectiveness in capturing fine details. We applied data augmentation during training, including random rotations, horizontal and vertical flips, and slight colour jittering. Training was terminated when the validation loss stabilised over 50 consecutive epochs, indicating convergence. This

resulted in a model with approximately 8 million trainable parameters, more compact than models such as TransUNet or CAR-UNet models [74, 75].

3 Experimental results

3.1 ASE_Res_UNet effectively segments microtubules from synthetic microscopy images mimicking the microtubule networks of *C. elegans* zygotes.

Various U-Net variants incorporating residual and attention modules have been developed or used for segmenting retinal blood vessels [45, 46, 48-50, 61, 62, 76-79]. We hypothesised that a similar approach could be effective for segmenting microtubules, given geometrical similarities both being tiny and curvilinear structures. To design a U-Net-based architecture tailored for microtubule segmentation in noisy fluorescence microscopy images, we trained on synthetic image datasets with reliable ground-truth annotations, allowing robust performance evaluation (cf. Section 2.1 and Supplementary Material 1A). We developed a novel model, ASE_Res_UNet, which combines residual encoder blocks with ASE attention decoder modules—a newly proposed attention mechanism designed to account for noise variability in microscopy images (cf. section 2.3, Fig. 1).

We first asked whether ASE_Res_UNet would be effective in segmenting microtubules in fluorescence microscopy images using *MicSim_FluoMT easy* dataset. This dataset simulates cytoplasmic fluorescence background, such as that arising from fluorescently tagged tubulin (e.g., GFP::tubulin), which can interfere with accurate microtubule segmentation. The model was trained using the WCE loss function (cf. sections 2.4 & 2.5) with no image preprocessing—training conditions that were maintained across all models in this study. ASE_Res_UNet successfully segmented microtubules along their full length (Fig. 2B). Quantitative evaluation confirmed strong performance: Dice = 0.9308 $\pm$ 0.0066, IoU = 0.9894 $\pm$ 0.0005, sensitivity = 0.9295 $\pm$ 0.0070, and precision = 0.9322 $\pm$ 0.0077 (mean $\pm$ standard deviation (SD) across 120 test images, cf. section 2.4). These results indicated a well-balanced segmentation, with neither false positives nor false negatives being predominant. Given the class imbalance in our images, we also computed the Matthews Correlation Coefficient (MCC = 0.9281 ± 0.0067) and the area under the Precision-Recall curve (PR AUC = 0.9845 ± 0.0029), both of which further supported the model's excellent performance. Importantly, segmentation quality remained high even under challenging conditions rarely addressed in previous work on blood vessel segmentation, such as high filament density in central regions and filament crossings in periphery. These results demonstrated that ASE_Res_UNet is a robust and accurate model for segmenting microtubule networks in fluorescence microscopy images, even in the presence of cytoplasmic background fluorescence.

To further assess the model under more challenging conditions, we created a second synthetic dataset, *MicSim_FluoMT Complex*, which simulates decreasing fluorescence intensity toward microtubule ends. This mimics the limited visibility of peripheral microtubules in confocal microscopy, caused by focal plane limitations and microtubule bending. These images contain regions where microtubules are visually indistinguishable from cytoplasmic background, presenting extremely low SNR scenario. These variations in signal intensities were of particular interest to rigorously test the noise-adaptability of ASE_Res_UNet. The model was trained on this dataset and then evaluated on the corresponding test images. While segmentation remained accurate in the central regions of embryos, errors were frequently observed at microtubule extremities (Fig. 2C). Performance metrics reflected the increased difficulty: Dice = 0.7825 ± 0.0156, IoU = 0.9688 ± 0.0011, sensitivity = 0.7345 ± 0.0233, precision = 0.8377 ± 0.0108, MCC = 0.7762 ± 0.0148, and PR AUC = 0.8730 ± 0.0147 (mean $\pm$ SD over 120 test images). Interestingly, ASE_Res_UNet tended to limit false positives, as indicated by higher precision relative to sensitivity. This conservative segmentation behaviour was desirable in biological applications, where microtubules are typically the primary structures of interest and false positives could be misleading.

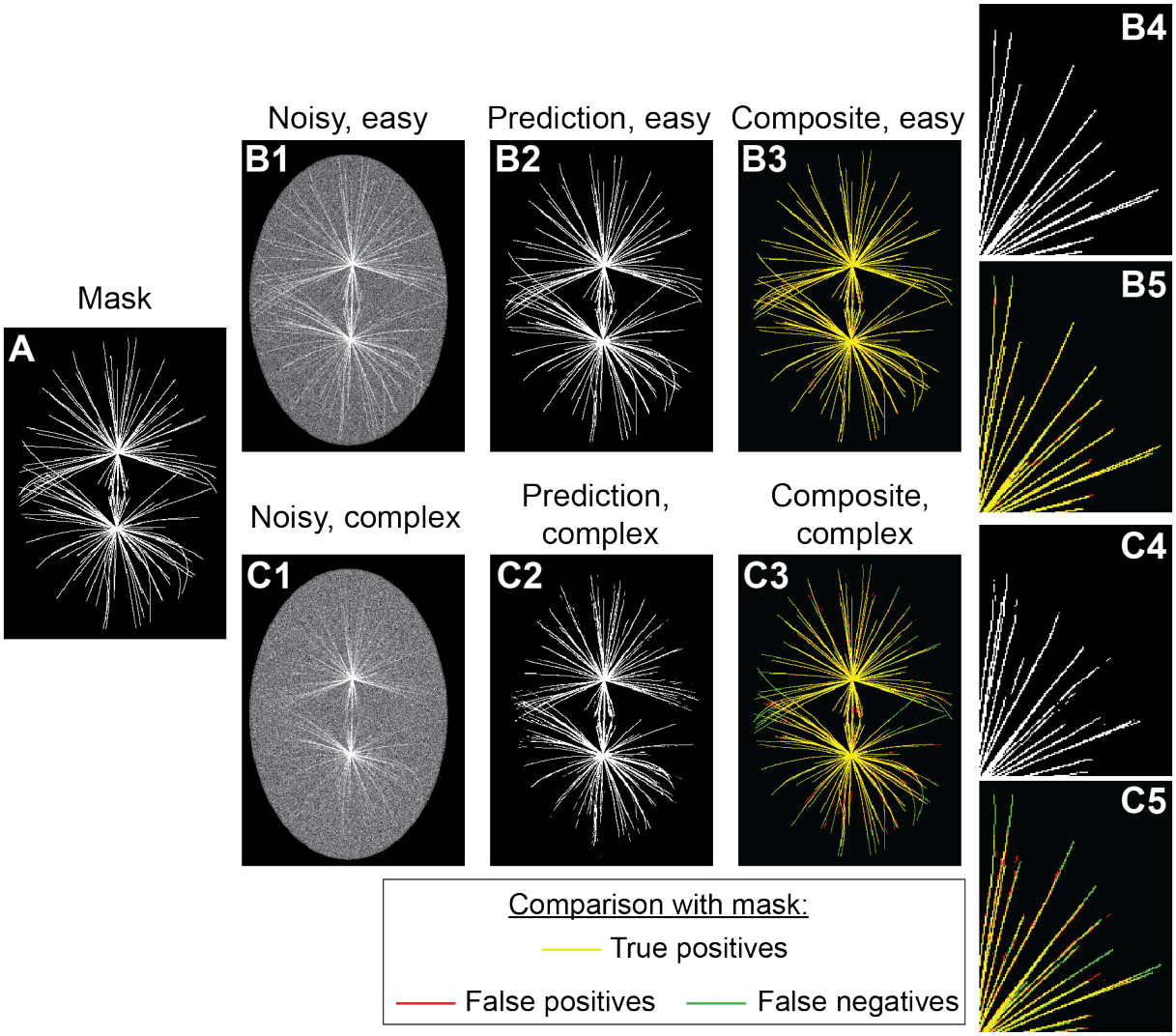

Fig. 2: ASE_Res_UNet architecture accurately segments microtubules in noisy fluorescence images, though performance declines in regions with extremely low signal-to-noise ratios.
Microtubule segmentation results obtained using ASE_Res_UNet on a test image from the *MicSim_FluoMT* datasets: (**A**) ground truth; (**B1, C1**) *easy* and *complex* images; (**B2, C2**) corresponding predicted images; (**B3, C3**) composite images used to visualise segmentation accuracy: true positives are shown in yellow, false negatives in green, false positives in red, and true negatives in black; (**B4, C4**) zoomed-in regions of interest (ROI); (**B5, C5**) composite ROI images.

We therefore concluded that the proposed ASE_Res_UNet architecture was well-suited for accurately segmenting microtubules in noisy fluorescence images. However, its performance declined in extreme cases where microtubules were barely distinguishable from the fluorescence background. These findings motivated us to further investigate the contribution of each architectural component of ASE_Res_UNet to its overall segmentation performance.

3.2 Both residual blocks and ASE modules enhance segmentation of low-intensity microtubule extremities.

To evaluate the contribution of each component in ASE_Res_UNet, we conducted an ablation study by designing three variant architectures for comparison: U-Net, the baseline architecture; Res_UNet, incorporating residual blocks in the encoder, mirroring ASE_Res_UNet's encoder path; and ASE_UNet, integrating ASE attention modules in the decoder, mimicking ASE_Res_UNet's decoder path. We first applied this ablation study to the *MicSim_FluoMT easy* dataset. The baseline U-Net already achieved high performance across all evaluation metrics (Table 2), suggesting that the U-Net backbone, the loss function, and the training setup were well optimised. Adding residual blocks (Res_UNet), ASE attention modules (ASE_UNet), or both (ASE_Res_UNet) did not significantly improved performance. Visual inspection of the predictions confirmed these results (Fig.

S3), indicating that this relatively easy dataset did not sufficiently challenge the architectures. In this case, the standard U-Net was capable of capturing most microtubule structures, leaving little room for further gains.

Model	Loss ↓	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
U-Net	**0.0035 ± 0.0002**	**0.9315 ± 0.0060**	0.9894 ± 0.0004	**0.9343 ± 0.0084** ***	0.9289 ± 0.0088 *	**0.9288 ± 0.0060**	**0.9849 ± 0.0028**
Res_UNet	0.0038 ± 0.0003	0.9287 ± 0.0071 ◊	0.9891 ± 0.0005 ***	0.9245 ± 0.0100 ***	**0.9329 ± 0.0084**	0.9258 ± 0.0071 ◊	0.9835 ± 0.0035
ASE_UNet	0.0036 ± 0.0002	0.9314 ± 0.0059	0.9894 ± 0.0004	0.9317 ± 0.0065 *	0.9311 ± 0.0066	0.9286 ± 0.0059	0.9843 ± 0.0029
ASE_Res_UNet	0.0036 ± 0.0002	0.9308 ± 0.0066	0.9894 ± 0.0005	0.9295 ± 0.0070	0.9322 ± 0.0077	0.9281 ± 0.0067	0.9845 ± 0.0029

Table 2: All ASE_Res_UNet variants accurately segment microtubules in the *MicSim_FluoMT easy* dataset.
Microtubule segmentation performance of ASE_Res_UNet and its variant architectures on the *MicSim_FluoMT easy* dataset, evaluated using various metrics (mean ± standard deviation over 120 test images). Bold values indicate the best performances. Statistical differences between ASE_Res_UNet and its variant architectures are indicated only when significant (◊: $0.01 < p \leq 0.05$; *: $0.001 < p \leq 0.01$; ***: $p \leq 0.0001$).

We hypothesised that the more challenging *MicSim_FluoMT Complex* dataset—with variable fluorescence intensity along microtubules—would better reveal performance differences. Indeed, ASE_Res_UNet outperformed the baseline U-Net across all metrics, except for precision (Table 3). This indicated that combining residual blocks and ASE modules enhanced the model's ability to segment microtubules in noisy and variable-intensity conditions. We interpreted the slightly lower precision as a trade-off: reducing false negatives came at the cost of a slight increase in false positives. When only residual blocs were added to the encoder (Res_UNet), we observed a modest performance improvement over U-Net (Table 3). Adding only ASE attention modules into the decoder (ASE_UNet) resulted in a more substantial boost, highlighting the importance of adaptive attention for detecting faint or low-contrast filaments (Table 3). Notably, ASE_Res_UNet outperformed ASE_UNet, confirming that residual connections and attention mechanisms provided complementary benefits for robust segmentation under complex scenarios. To complement the quantitative analysis, we visually examined the segmentation outputs across several test images. ASE_Res_UNet showed better recovery of microtubule extremities compared to U-Net (Fig. 3C1-C3), consistent with confusion matrix results (Table S3). U-Net missed microtubule pixels in filament regions with lower intensity near the periphery. To explore this spatial effect further, we computed performance metrics separately for two embryo regions: the periphery, where microtubules extremities are faint, and the central region, where microtubules are more prominent (Fig. S4). For ASE_Res_UNet, segmentation quality was consistently lower in peripheral regions compared to central ones—as reflected by reduced Dice, sensitivity, MCC, and PR AUC scores (Table S4). When comparing ASE_Res_UNet to U-Net across these two regions, the performance gain was more pronounced in the periphery, suggesting that the added modules particularly benefited segmentation of low-intensity microtubule ends. We next examined whether the individual components—residual blocks or ASE modules—were sufficient to improve peripheral segmentation. Visually, Res_UNet underperformed compared to ASE_Res_UNet (Fig. 3D1–D3). Although it slightly improved segmentation of some extremities over U-Net, it failed to generalise across all noisy regions. This was supported by the spatial analysis, which showed only minor improvements over U-Net in the periphery (Table S4). In contrast, ASE_UNet showed stronger enhancement of extremity segmentation (Fig. 3E1–E2), albeit with an increase in false positives (Table S3). Spatial metrics confirmed that ASE modules significantly improved performance in the periphery (Table S4). Still, ASE_UNet

underperformed relative to ASE_Res_UNet (Fig. 3E vs. 3F), suggesting that combining residual and attention modules yielded better results than either component alone.

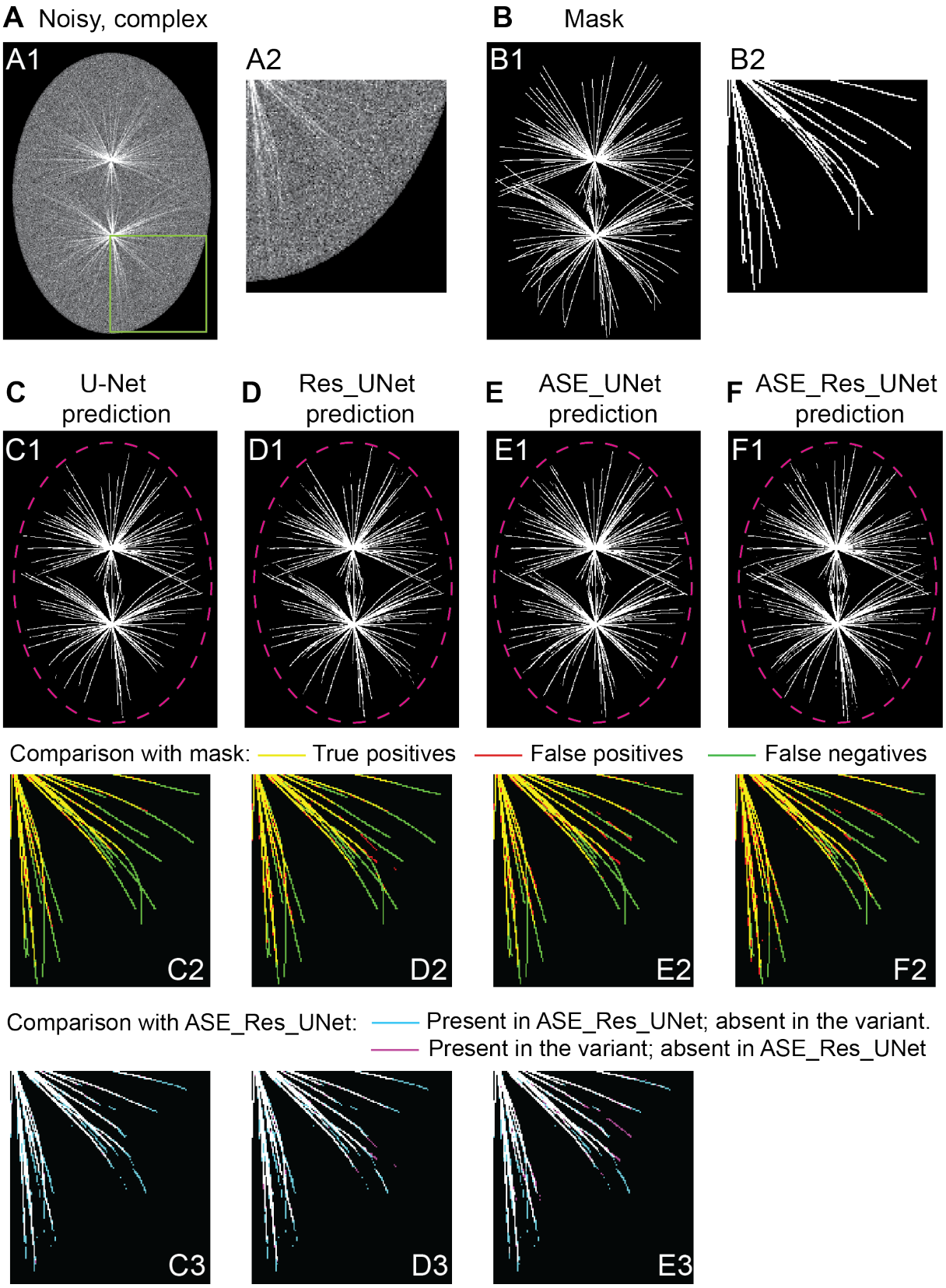

Fig. 3: ASE_Res_UNet recovers longer microtubules of the *MicSim_FluoMT complex* dataset compared to its variants.
Microtubule segmentation results obtained using ASE_Res_UNet and its variants on a sample test image from the *MicSim_FluoMT complex* dataset. (**A**) Input image; (**B**) Corresponding ground truth; (**C-F**) Predicted segmentations from (C) U-Net, (D) Res_UNet, (E) ASE_UNet, and (F) ASE_Res_UNet models. (A1-F1): whole simulated embryo; (A2-F2): zoomed-in regions of interest (ROI) to better highlight differences between model predictions and ground truth, with cropping window shown in green on panel A1; (C2-F2) composite ROI images showing true positives in yellow, false negatives in green, false positives in red, and true negatives in black; (C3-E3): composite ROI comparison between (magenta) each variant and (cyan) ASE_Res_UNet. Therefore, microtubules segmented by both architectures appear in white, those segmented only by the variant architecture appear in magenta, and those segmented only by ASE_Res_UNet appear in cyan. Dashed pink line indicates the embryo contour.

Model	Loss ↓	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
U-Net	0.0136 ± 0.0011 ***	0.7362 ± 0.0190 ***	0.9662 ± 0.0022 ***	0.6175 ± 0.0293 ***	**0.9132 ± 0.0147** ***	0.7429 ± 0.0158 ***	0.8657 ± 0.0140 **
Res_UNet	0.0132 ± 0.0007 ***	0.7383 ± 0.0185 ***	0.9660 ± 0.0016 ***	0.6292± 0.0271 ***	0.8943 ± 0.0127 ***	0.7419 ± 0.0160 ***	0.8557 ± 0.0162 ***
ASE_UNet	0.0125 ± 0.0007 ***	0.7694 ± 0.0155 ***	0.9679 ± 0.0014 ***	0.7025 ± 0.0369 ***	0.8544 ± 0.0383 ***	0.7658 ± 0.0142 ***	0.8635 ± 0.0157 ***
ASE_Res_UNet	**0.0116 ± 0.0004**	**0.7825 ± 0.0156**	**0.9688 ± 0.0011**	**0.7345 ± 0.0233**	0.8377 ± 0.0108	**0.7762 ± 0.0148**	**0.8730 ± 0.0147**

Table 3: ASE_Res_UNet outperforms its architectural variants in microtubule segmentation on the *MicSim_FluoMT complex* dataset.
Microtubule segmentation performances of ASE_Res_UNet and its variant architectures on the *MicSim_FluoMT complex* dataset, evaluated using various metrics (mean ± standard deviation over 120 test images). Bold values indicate the best performances. Statistical differences between ASE_Res_UNet and its variant architectures are indicated only when significant (**: $0.0001 < p \leq 0.001$; ***: $p \leq 0.0001$).

Overall, both quantitative and qualitative analyses demonstrated that ASE_Res_UNet improved segmentation accuracy and robustness to noise, particularly for faint microtubule extremities. We proposed that the combination of residual blocks and ASE attention modules enhanced feature extraction and noise handling, allowing the model to better distinguish microtubules from background fluorescence. Since our results highlighted persistent challenges in highly degraded regions, where microtubule signals are barely detectable, we questioned whether alternative model architectures might offer improved performance under these extreme conditions.

3.3 ASE_Res_UNet outperforms other advanced architectures in segmenting low-intensity microtubule extremities.

Deep learning models have been developed for curvilinear structure segmentation such as retinal blood vessels—morphologically similar to microtubules—and have thus inspired a variety of architectural approaches from which we could draw [48, 61, 62, 64, 75, 76, 80-82]. We benchmarked ASE_Res_UNet against several deep-learning architectures that differ in the attention mechanisms or in their core component. We used the *MicSim_FluoMT complex* dataset, which represents a highly challenging segmentation scenario. All models were trained under identical conditions (sections 2.4 & 2.5) and evaluated using both quantitative metrics and qualitative visual inspection.

We first focussed on critically evaluating key architectural design choices: (1) our custom attention mechanism, and (2) the placement of residual and attention blocks within the network. To do this, we compared

ASE_Res_UNet against three deep learning architectures that also employed attention mechanisms and residual components within an U-Net backbone: CAR-UNet [75], AG_Res_UNet [83], and SE_Res_UNet [65]. CAR-UNet (Channel Attention Residual U-Net), originally developed for segmenting retinal vessels with strong performance [75], integrates Channel Attention Double Residual Blocks (CADRB) in both the encoder and decoder paths, and Modified Efficient Channel Attention (MECA) modules in the skip connections (Fig. S5A). In contrast to ASE_Res_UNet, CAR-UNet applies both residual and attention components at every layer, resulting in a larger model (~16 million parameters). AG_Res_UNet incorporates a grid-based self-attention gating module, previously applied in biomedical segmentation tasks [83]. In this comparison, we retained the same backbone as ASE_Res_UNet, and modified only the attention mechanism to focus on our novel mechanism. This resulted in a model with a similar number of parameters (Fig. S5B). Similarly, SE_Res_UNet was created by replacing the ASE module with a Squeeze-and-Excitation (SE) module [65] (Fig. S5C), allowing us to isolate the contribution of our adaptive SE module. SE mechanisms have also been applied in retinal vessel segmentation [76, 77]. Among these three models, ASE_Res_UNet achieved the best performance across multiple metrics, including Dice coefficient, IoU, sensitivity, MCC, and PR AUC (Table 4). The most notable gain was in sensitivity, indicating improved detection of low-intensity microtubule extremities. This came with a minor reduction in precision due to increased false positives (Table S5), yet overall performance was superior without increasing model size. Qualitative analysis of segmentation predictions further supported these findings. CAR-UNet and AG_Res_UNet both produced shorter segmented filaments (Fig. 4D, E), reflecting their lower sensitivity and higher false negative counts. In contrast, SE_Res_UNet produced longer filaments (Figure 4F), demonstrating the utility of SE blocks for enhancing low-intensity detection. Still, it lagged behind ASE_Res_UNet overall. Altogether, these comparisons validated two key design choices in ASE_Res_UNet: (1) ASE modules, which integrate noise estimation, outperformed AG and SE modules in noisy images. (2) Placing residual blocs in the encoder and attention mechanisms in the decoder proved sufficient, enabling high performances with a compact model suitable for resource-constrained environments.

Second, we investigated whether architectures based on fundamentally different design paradigms could outperform ASE_Res_UNet in segmenting microtubules in noisy fluorescence images. We selected an architecture with a Generative Adversial Network (GAN) module alongside a U-Net, i.e. pix2pix model [84] (Fig. S5D). GAN-based methods have previously shown superior performance over standard U-Net or Residual U-Net architectures in tasks such as retinal vessel or corneal nerve segmentation [64, 80-82]. We also included TransUNet, an architecture that integrates transformer layers with a U-Net [74, 85] (Fig. S5E), recently applied to various biomedical image segmentation tasks [74, 85-88]. While Pix2pix showed poor performance, TransUNet produced results closer to ASE_Res_UNet, though at the cost of having approximately four times more parameters (Tables 4 and S5). Specifically, TransUNet had a true-positive count most similar to that of ASE_Res_UNet, albeit with slightly higher false-positive and false-negative counts (Table S5). Visually, Pix2pix produced predictions with more hallucinated structures and shorter microtubules (Fig. 4G), consistent with its lower and higher numbers of true positives and false positives, respectively (Table S5). In contrast, the predictions from TransUNet closely resembled those of ASE_Res_UNet, particularly in the accurate segmentation of microtubule extremities (Fig. 4H). These results suggested that the GAN-based approach was not well suited for segmenting microtubules in noisy images, while the transformer-based architecture in TransUNet showed promising performance, although inferior to ASE_Res_UNet and requiring substantially more computational resources.

All models occasionally hallucinated structures in the peripheral regions, where microtubule intensity approached background noise. To assess this systematically, we computed performance metrics separately for central vs. peripheral regions (Table S6). ASE_Res_UNet consistently outperformed all others in both regions, with the largest margin in the periphery. TransUNet was the second-best performer, but with significantly reduced precision, indicating more hallucinations. AG_Res_UNet achieved the highest precision overall (fewest false positives), but underperformed in peripheral detection due to higher false negatives.

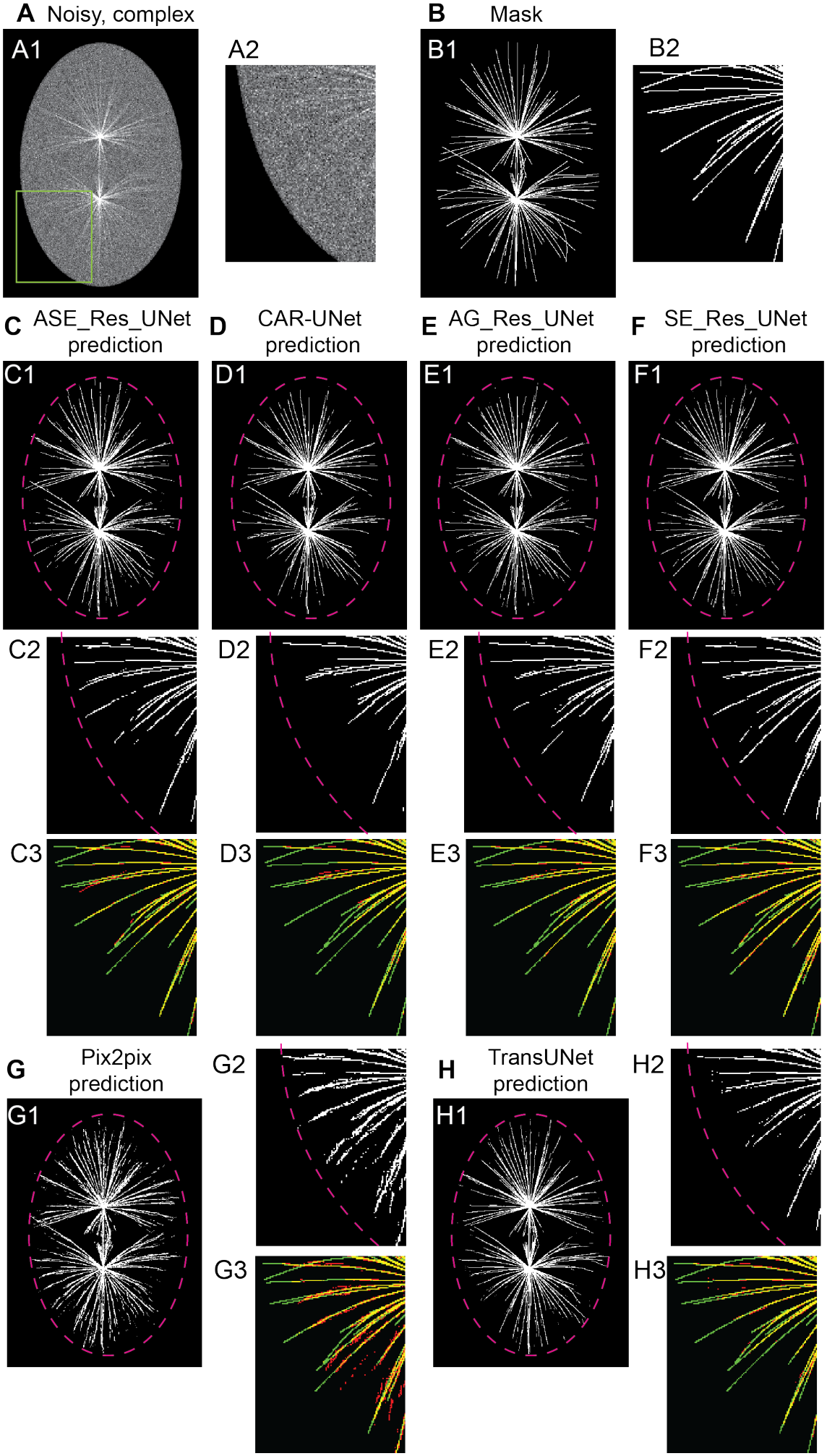
A Noisy, complex
A1
A2
B Mask
B1
B2
C ASE_Res_UNet prediction
C1
C2
C3
D CAR-UNet prediction
D1
D2
D3
E AG_Res_UNet prediction
E1
E2
E3
F SE_Res_UNet prediction
F1
F2
F3
G Pix2pix prediction
G1
G2
G3
H TransUNet prediction
H1
H2
H3

Fig. 4: ASE_Res_UNet produces the most accurate segmentation predictions compared to state-of-the-art architectures.
Microtubule segmentation results obtained using ASE_Res_UNet and five other advanced architectures on a sample test image from the *MicSim_FluoMT complex* dataset. (**A**) Input image; (**B**) corresponding ground truth; (**C-H**) predicted segmentations from (C) ASE_Res_UNet, (D) CAR-UNet, (E) AG_Res_UNet, (F) SE_Res_UNet, (G) Pix2pix, and (H) TransUNet models. (A1-H1): whole simulated embryo; (A2-H2): zoomed-in regions of interest (ROI) to better highlight differences between model predictions and ground truth, with cropping window shown in green on panel A1; (C3-H3) composite ROI images showing true positives in yellow, false negatives in green, false positives in red, and true negatives in black. Dashed pink line indicates the embryo contour.

Model	Parameter number ↓	Loss ↓	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
CAR-UNet	15.83 millions	0.0127 ± 0.0005 ***	0.7605 ± 0.0147 ***	0.9676 ± 0.0013 ***	0.6732 ± 0.0211 ***	0.8743 ± 0.0145 ***	0.7591 ± 0.0137 ***	0.8620 ± 0.0153 ***
AG_Res_UNet	8.02 millions	0.0122 ± 0.0006 ***	0.7631 ± 0.0155 ***	0.9679 ± 0.0012 ***	0.6765 ± 0.0207 ***	**0.8758 ± 0.0188** ***	0.7617 ± 0.0148 ***	0.8676 ± 0.0164 ***
SE_Res_UNet	8.81 millions	0.0122 ± 0.0005 ***	0.7655 ± 0.0175 ***	0.9681 ± 0.0011 ***	0.6826 ± 0.0258 ***	0.8718 ± 0.0109 ***	0.7634 ± 0.0161 ***	0.8656 ± 0.0167 ***
TransUNet	32.6 millions	0.0123 ± 0.0005 ***	0.7721 ± 0.0170 ***	0.9675 ± 0.0011 ***	0.7214 ± 0.0271 ***	0.8309 ± 0.0081 ***	0.7657 ± 0.0160 ***	0.8607 ± 0.0169 ***
Pix2pix	30.6 millions	0.0179 ± 0.0007 ***	0.6889 ± 0.0170 ***	0.9534 ± 0.0014 ***	0.6805 ± 0.0158 ***	0.6979 ± 0.0263 ***	0.6767 ± 0.0170 ***	0.7434 ± 0.0220 ***
ASE_Res_UNet	7.96 millions	**0.0116 ± 0.0004**	**0.7825** ± 0.0156	**0.9688** ± 0.0011	**0.7345** ± 0.0233	0.8377 ± 0.0108	**0.7762** ± 0.0148	**0.8730 ± 0.0147**

Table 4: ASE_Res_UNet outperforms other advanced architectures in microtubule segmentation on the *MicSim_FluoMT complex* dataset.
Microtubule segmentation performances of ASE_Res_UNet and other advanced architectures on the *MicSim_FluoMT complex* dataset, evaluated using various metrics (mean ± standard deviation over 120 test images). Bold values indicate the best performances. Statistical differences between ASE_Res_UNet and other advanced architectures are all significant (Student t-test per metric: ***: $p \leq 0.0001$).

ASE_Res_UNet provided the best balance between detecting low-intensity microtubule extremities and minimising hallucinations, while also maintaining the lowest computational cost. These results confirmed that ASE_Res_UNet is well-suited for challenging segmentation tasks involving noisy, low-contrast fluorescence microscopy images.

3.4 ASE_Res_UNet achieves near-perfect microtubule segmentation on the SynthMT dataset

To further evaluate ASE_Res_UNet, we tested its performance on the SynthMT dataset [31], an independent synthetic microtubule dataset that differs from *MicSim_FluoMT* in two key aspects (cf. section 2.1): it features individual *in vitro* microtubules (rather than *in vivo* cytoskeletal networks) and exhibits a higher SNR. Additionally, this dataset provides a large training set (5280 images). Given these characteristics, we foresaw that microtubule segmentation would be comparatively easier, yielding strong performance. Visual inspection confirmed this hypothesis: ASE_Res_UNet's predictions closely aligned with ground-truth masks (Fig. 5). Quantitative metrics further validated its accuracy: Dice = 0.9807 ± 0.0171, IoU = 0.9981 ± 0.0020, Sensitivity =

0.9854 ± 0.0099, Precision = 0.9764 ± 0.0282, MCC = 0.9803 ± 0.0172, and PR AUC = 0.9985 ± 0.0024 (mean ± SD over 660 test images). ASE_Res_UNet outperformed foundation models from the SynthMT study, including the recently introduced SAM3Text (average precision of 0.95) [31]. These results demonstrated that ASE_Res_UNet, originally developed for noisy microtubule segmentation, also excels in high-SNR conditions.

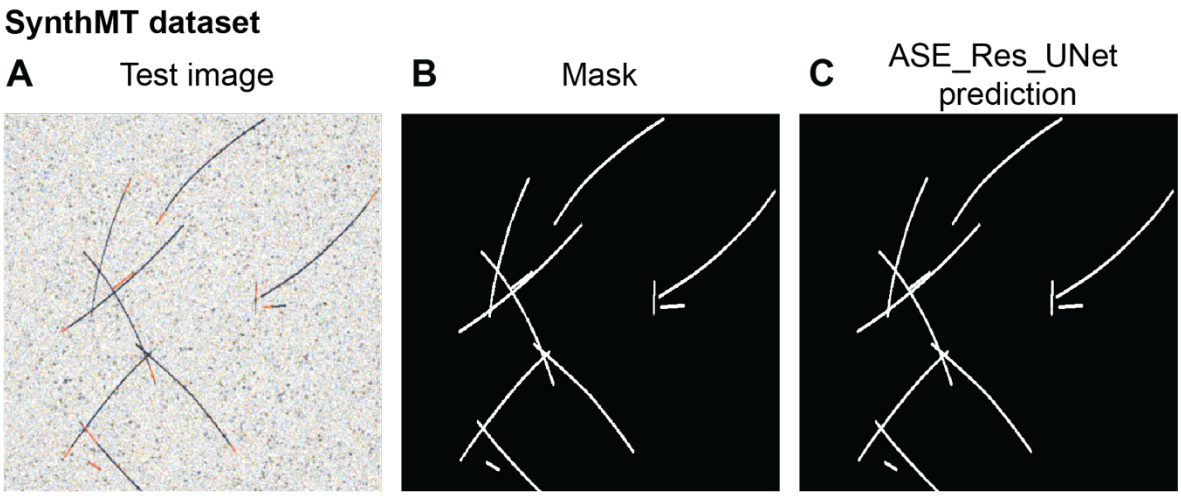

Fig. 5: ASE_Res_UNet predictions closely match ground-truth masks in the SynthMT dataset
(**A**) Exemplar synthetic test image, (**B**) Ground-truth mask, and (**C**) ASE_Res_UNet prediction (model trained on SynthMT).

3.5 ASE_Res_UNet enables to segment microtubules in real microscopy images.

We next questioned whether ASE_Res_UNet would maintain its high performance on real microscopy images, which often pose additional challenges not fully captured in synthetic datasets. Real images may exhibit biological and experimental variability, imaging artifacts, or background structures that complicate learning. Moreover, the limited size of annotated real-world datasets—due to the difficulty and time cost of manual annotation—can further hinder training. To address this, we evaluated ASE_Res_UNet on the *MicReal_FluoMT* dataset, which contains 49 fluorescence microscopy images of *C. elegans* fixed embryos stained with anti-tubulin antibodies (cf. Section 2.2). This dataset represents a realistic biological context with naturally occurring imaging noise, variability in filament orientation and density, and heterogeneous fluorescence intensities (e.g., bright at centrosomes and faint at the cell periphery). Despite the limited training data, ASE_Res_UNet accurately segmented astral microtubules (Fig. 6). Notably, it successfully detected microtubule extremities that were barely visible in the raw images (Fig. 6A, C), and captured filaments with weak or uneven staining. Interestingly, the model produced segmentations that were often smoother and more continuous than the corresponding manual annotations (Fig. 6C vs. 6B), suggesting that the model may surpass manual annotations in terms of consistency and completeness. Quantitative metrics supported these observations: IoU = 0.9010 ± 0.0284, accuracy = 0.9477 ± 0.0158, specificity = 0.9823 ± 0.0070, precision = 0.7795 ± 0.0604, and PR AUC = 0.7779 ± 0.0502 (mean ± standard deviation over 10 test images). We compared ASE_Res_UNet predictions with segmentations generated using standard image-processing tools in Fiji, a widely adopted software in biological research [89]. Among local thresholding methods, the Bernsen algorithm (suitable for edge detection [90]) and the Phansalkar algorithm (optimised for low contrast images [91]) achieved the highest performance (Fig. 6E-F). However, these methods produced shorter microtubules and omitted some microtubules compared to ASE_Res_UNet. Overall, ASE_Res_UNet demonstrated robust segmentation performance in fixed *C. elegans* embryos, effectively capturing microtubules in both dense and low-signal regions, despite the limited size of the training dataset. These results highlighted its potential as a valuable tool for biologists, offering high-quality and efficient segmentation of cytoskeletal filaments in fluorescence microscopy images.

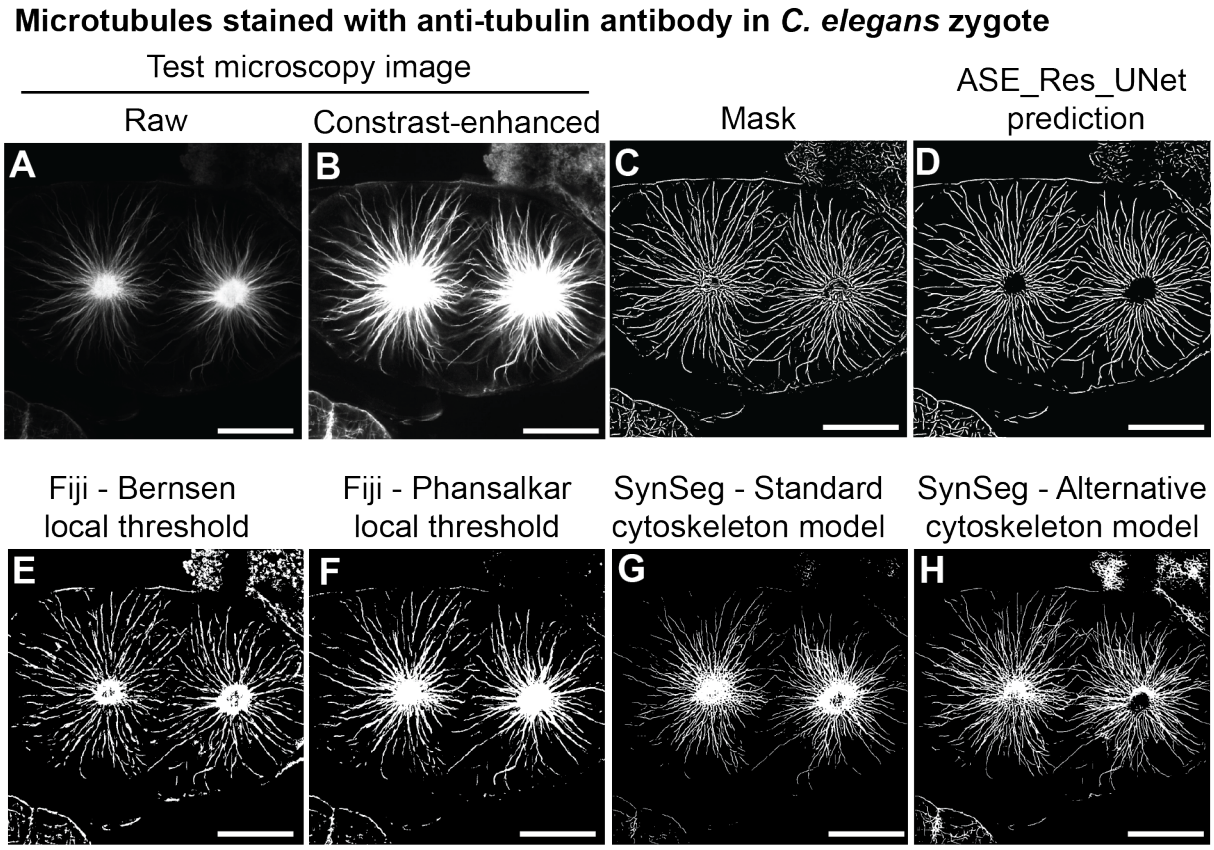

Fig. 6: ASE_Res_UNet successfully segments microtubules in fixed embryos, including dense centrosomal regions and faint peripheral areas.
Microtubule segmentation in a fixed *C. elegans* zygote stained with anti-tubulin antibody: (**A, B**) Test microscopy image: (A) raw, and (B) contrast-enhanced (to see low-contrast filaments and filament extremities), (**C**) Ground-truth mask, (**D**) ASE_Res_UNet prediction (model trained on the *MicReal_FluoMT* dataset), (**E, F**) Fiji-based segmentations using local thresholding: (E) Bernsen or (F) Phansalkar, and (**G, H**) SynSeg predictions from pre-trained cytoskeleton models with smart auto-resize: (G) standard model, and (H) alternative model. Scale bar indicates 10 μm.

To further evaluate our method's performance, we compared ASE_Res_UNet with SynSeg, a state-of-the-art deep-learning framework for subcellular structure segmentation—including microtubules—that employs a synthetic data-driven approach [53]. SynSeg offers two pre-trained models for cytoskeleton segmentation (termed standard and alternative), which we benchmarked on the *MicReal_FluoMT* dataset. Visual inspection revealed that both SynSeg models produced lower-quality segmentations than ASE_Res_UNet, with the standard model performing particularly poorly (Fig. 6 G, H). SynSeg's outputs exhibited shorter microtubules, discontinuous filaments, and missed detections. Quantitative metrics further confirmed ASE_Res_UNet's superiority. The alternative SynSeg model achieved an IoU of 0.8641 ± 0.0350, accuracy of 0.9268 ± 0.0201, specificity of 0.9728 ± 0.0096, precision of 0.6251 ± 0.0430, and PR AUC of 0.5705 ± 0.0436 (mean $\pm$ standard deviation over 10 test images). Collectively, these findings established ASE_Res_UNet as a highly competitive—if not superior—approach for microtubule segmentation compared to this cutting-edge framework.

We next evaluated ASE_Res_UNet on a previously published confocal microtubule image dataset (Higaki dataset) that was recently partially reannotated by Guo *et al.* [53] (cf. section 2.2). Using 90 images of *Nicotiana tabacum* BY-2 cells, we performed five-fold cross-validation and achieved strong performances (Fig. 7A-D). ASE_Res_UNet outperformed SynSeg predictions from pretrained cytoskeleton models (Fig. 7E, F), which yielded an averaged precision of ~0.65 and an averaged Dice of ~0.70 (see Fig. 6 in [53]). These results demonstrated ASE_Res_UNet's competitiveness in segmenting microtubules in real images with diverse filament networks and varied microscopy modalities.

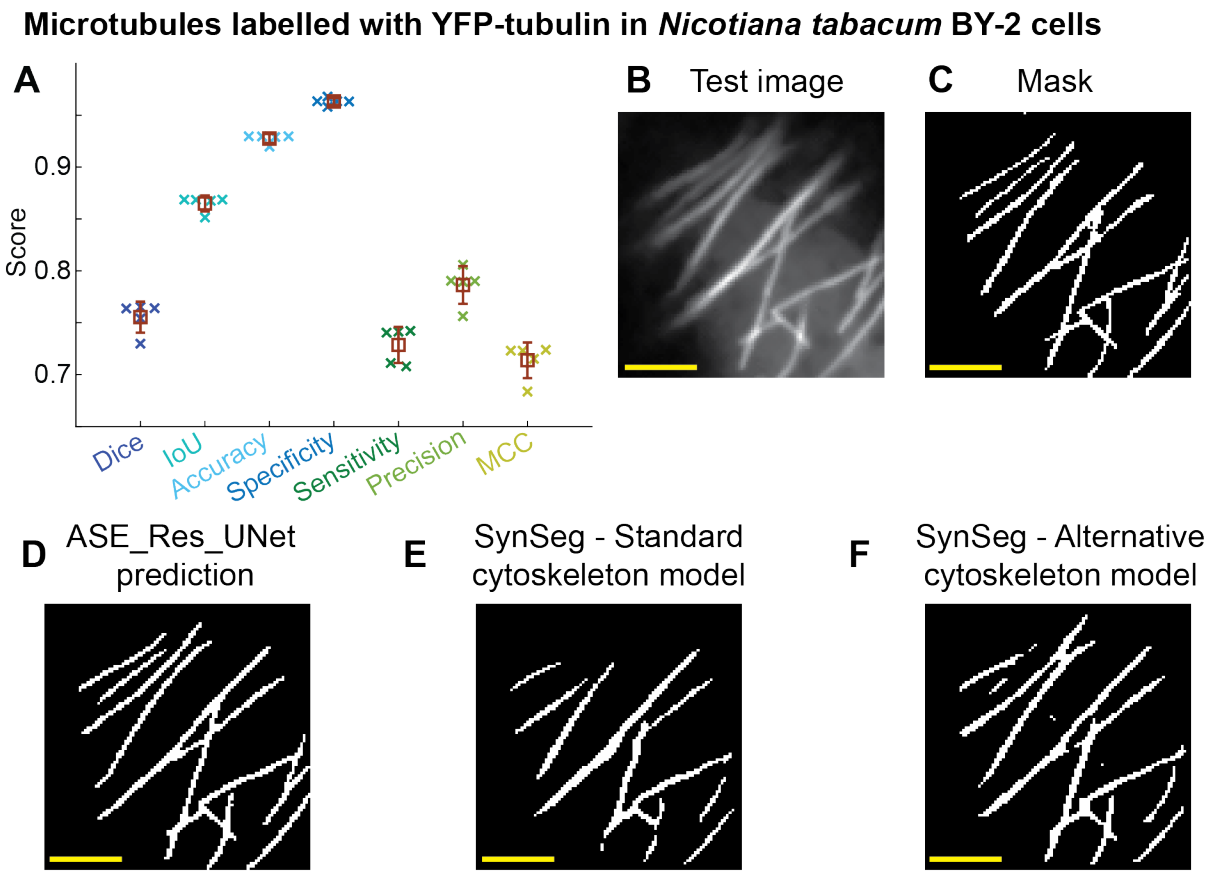

Fig. 7: ASE_Res_UNet accurately segments microtubules in *Nicotiana tabacum* BY-2 cells.
(**A**) Performance metrics for ASE_Res_UNet across 5-fold cross-validation (CV) on 90 re-annotated images from the Higaki dataset. Crosses represent individual fold score, while brown squares and error bars indicate the means ± standard deviation across fold. (**B**) Test image (unseen during CV), (**C**) corresponding mask (re-annotated by Guo *et al.*), (**D**) ASE_Res_UNet prediction (trained on a given fold), (**E, F**) predictions from SynSeg pre-trained cytoskeleton models with smart auto-resize: (E) standard model, and (F) alternative model. Yellow scale bar indicates 5 μm.

To further assess ASE_Res_UNet's robustness, we applied the model trained on *MicReal_FluoMT* to live *C. elegans* mitotic embryos expressing GFP::TBB-2, without fine-tuning, transfer learning, or preprocessing. Since astral microtubules are very dynamic in the *C. elegans* zygote, it calls for high-speed imaging, leading to low SNR and challenging segmentation conditions (Fig. 8A). ASE_Res_UNet successfully segmented most microtubules along their entire length, demonstrating strong generalisation (Fig. 8B). However, faint microtubules were detected with lower accuracy, suggesting that semi-supervised learning with partial annotations could improve performance in such cases. ASE_Res_UNet outperformed both the Bernsen and Phansalkar local thresholding algorithms (Fig. 8C, D), and SynSeg predictions from standard and alternative cytoskeleton models (Fig. 8E, F). Notably ASE_Res_UNet improved microtubule continuity compared to SynSeg's alternative cytoskeleton model. These results suggested that ASE_Res_UNet holds promise for applications across diverse biological samples, including dynamic imaging conditions.

Overall, ASE_Res_UNet performed robustly on real microscopy images, including datasets with different imaging modalities and cell types. It offered a novel tool for biological image analysis, enabling accurate and efficient microtubule segmentation across a range of experimental conditions.

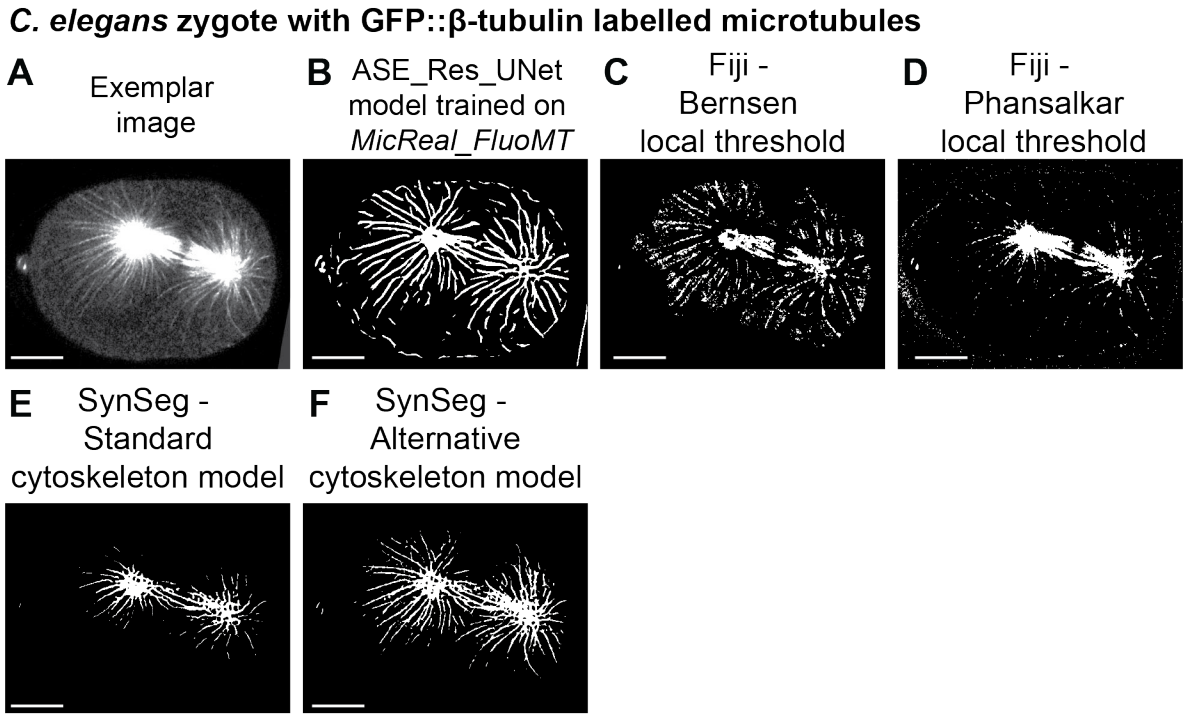

Fig. 8: ASE_Res_UNet trained on the *MicReal_FluoMT* dataset generalizes to live *C. elegans* embryos. Microtubule segmentation in a *C. elegans* zygote expressing GFP::β-tubulin: (**A**) exemplar fluorescence image, (**B**) ASE_Res_UNet prediction (model trained on the *MicReal_FluoMT* dataset without fine-tuning or transfer learning), (**C, D**) segmentations via Fiji's local thresholding methods: (C) Bernsen and (D) Phansalkar, and (**E, F**) predictions from SynSeg pre-trained cytoskeleton models with smart auto-resize: (E) standard model, and (F) alternative model. Scale bar indicates 10 μm.

3.6 ASE_Res_UNet demonstrates strong generalisation for segmenting vessels and nerves in biomedical images

Finally, we asked whether ASE_Res_UNet possessed generalisation capabilities for segmenting curvilinear structures beyond microtubules. To test this, we first evaluated the model on the *DRIVE* dataset, which contains retinal fundus images with annotated blood vessels that differ significantly in appearance, scale and imaging modality (cf. section 2.2). ASE_Res_UNet trained on *DRIVE* was able to accurately capture complex vessel patterns, including thin and low-contrast vessels (Fig. 9A-C). Quantitative evaluation further confirmed the model's strong performance (Table 5). To assess competitiveness, we compared ASE_Res_UNet with CAR-UNet, a model specifically designed for retinal vessel segmentation. Despite using approximately half the number of trainable parameters, ASE_Res_UNet achieved slightly better performance (Table 5), and successfully segmented some faint vessels missed by CAR-UNet (Fig. 9D vs. 9C). We extended this comparison to AG_Res_UNet and SE_Res_UNet, which differed from ASE_Res_UNet only in their attention mechanisms. ASE_Res_UNet outperformed both models in terms of precision (Table 5), which was consistent with qualitative observations (Fig. 9E, F). Last, we compared ASE_Res_UNet to TransUNet, which had previously shown competitive performance on the microtubule segmentation task. ASE_Res_UNet again slightly outperformed TransUNet on the DRIVE dataset and provided a better balance between false positives and false negatives (Table 5, Fig. 9G). Importantly, this was achieved with significantly fewer computational resources. Overall, ASE_Res_UNet demonstrated robust generalisation by achieving state-of-the-art performance on a segmentation task it was not originally designed for.

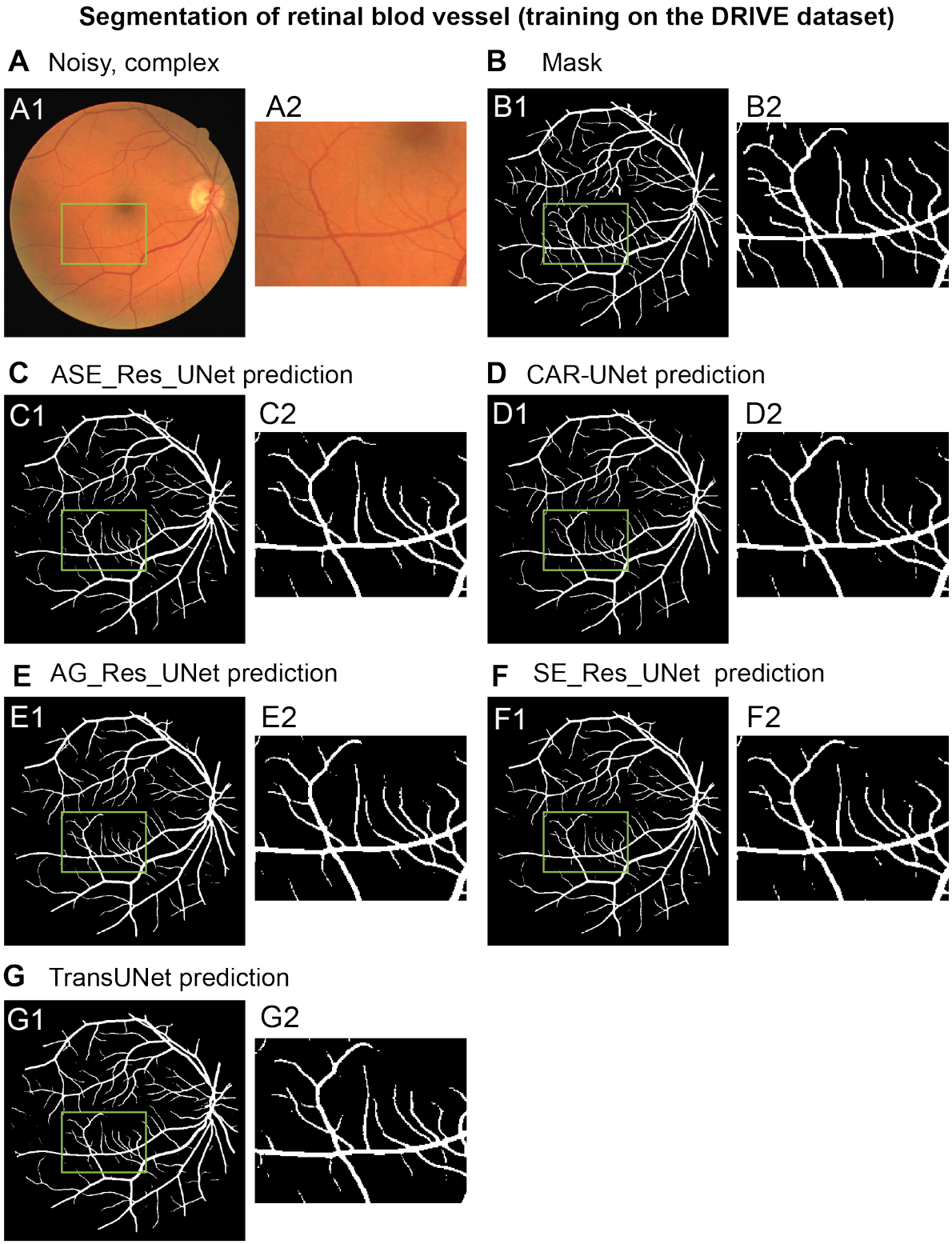

Fig. 9: ASE_Res_UNet successfully segments retinal blood vessels, including thin and low-contrast vessels, while limiting false positives.
Retinal vessel segmentation results obtained using ASE_Res_UNet and four advanced architectures on a sample test image from the *DRIVE* dataset. (**A**) Input image; (**B**) corresponding ground truth; (**C-G**) predicted segmentations from (C) ASE_Res_UNet, (D) CAR-UNet, (E) AG_Res_UNet, (F) SE_Res_UNet, and (G) TransUNet models. (A1-G1): whole images; (A2-G2): zoomed-in regions of interest (ROI) to better highlight differences between model predictions and ground truth, with cropping window shown in green on panels A1 and B1.

Model	Parameter number ↓	Loss ↓	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
CAR-UNet	15.83 millions	0.1563 ± 0.0210	0.8159 ± 0.0165	0.9414 ± 0.0059	0.8036 ± 0.0427	**0.8312 ± 0.0314**	0.8003 ± 0.0171	0.9032 ± 0.0122
AG_Res_UNet	8.02 millions	0.1601 ± 0.0216	0.8171 ± 0.0158	0.9401 ± 0.0053	0.8289 ± 0.0472	0.8084 ± 0.0295	0.8011± 0.0162	0.9024 ± 0.0147
SE_Res_UNet	8.81 millions	0.1636 ± 0.0218	0.8174 ± 0.0191	0.9414 ± 0.0065	0 .8117 ± 0.0417	0.8254 ± 0.0286	0.8016 ± 0.0204	0.9019 ± 0.0171
TransUNet	32.3 millions	0.1557 ± 0.0227	0.8150 ± 0.0129	0.9389 ± 0.0049	**0.8347 ± 0.0378**	0.7983 ± 0.0291	0.7987 ± 0.0138	0.9004 ± 0.0152
ASE_Res_UNet	8.00 millions	**0.1481 ± 0.0221**	**0.8236 ± 0.0167**	**0.9433 ± 0.0055**	0.8185 ± 0.0432	**0.8311 ± 0.0294**	**0.8084 ± 0.0178**	**0.9114 ± 0.0150**

Table 5: ASE_Res_UNet achieves state-of-the-art performance on segmenting retinal blood vessels. Retinal vessel segmentation performances of ASE_Res_UNet and other advanced architectures on the *DRIVE* dataset, evaluated using various metrics (mean ± standard deviation over 20 test images). Bold values indicate the best performances. Statistical differences between ASE_Res_UNet and other architectures are non-significant.

To further assess the generalisation ability, we evaluated of ASE_Res_UNet on the *CORN-1* dataset, which contains fluorescence microscopy images of corneal nerves (cf. section 2.2). ASE_Res_UNet successfully segmented the nerve structures, including those with weak fluorescence signals (Fig. 10C vs. 10B) and it performed better than U-Net (Fig. 10D). Quantitative metrics aligned with these findings: ASE_Res_UNet outperformed U-Net across all metrics (Table 6). Its success on three distinct curvilinear structure segmentation tasks—microtubules, blood vessels, and corneal nerves—underscored the robustness and adaptability of ASE_Res_UNet's architectural design. In particular, the combination of residual encoding and noise-adaptive ASE attention modules enables effective feature extraction and denoising across diverse biomedical imaging contexts.

Model	Loss ↓	Dice ↑	IoU ↑	Sensitivity ↑	Precision ↑	MCC ↑	PR AUC ↑
U-Net	0.0416 ± 0.0017 ◊	0.6993 ± 0.0123 *	0.9837 ± 0.0004 ◊	0.6647 ± 0.0267 ◊	0.7523 ± 0.0135	0.7023 ± 0.0112 *	0.7895 ± 0.0107 ◊
ASE_Res_UNet	**0.0370 ± 0.0026**	**0.7298 ± 0.0140**	**0.9851 ± 0.0008**	**0.7015 ± 0.0135**	**0.7726 ± 0.0200**	**0.7321 ± 0.0138**	**0.8214 ± 0.0176**

Table 6: ASE_Res_UNet outperforms U-Net for segmenting corneal nerves. Retinal nerve segmentation performances of ASE_Res_UNet and U-Net on the CORN-1 dataset, evaluated using various metrics (mean ± standard deviation over the 5-fold cross-training). Bold values indicate the best performances. Statistical differences between ASE_Res_UNet and U-Net are indicated only when significant (◊: $0.01 < p \leq 0.05$; *: $0.001 < p \leq 0.01$).

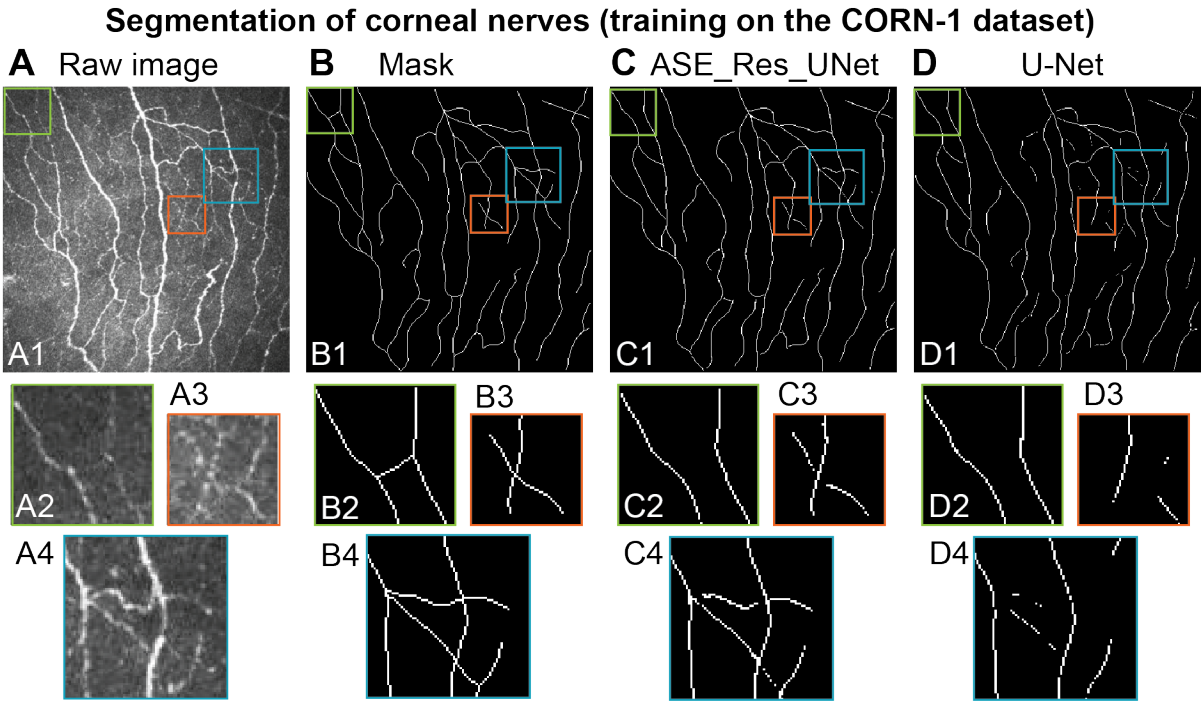

Fig. 10: ASE_Res_UNet successfully segments corneal nerves in noisy fluorescence microscopy images. Retinal nerve segmentation results obtained using ASE_Res_UNet and U-Net trained on the *CORN-1* dataset. (**A**) Exemplar test image; (**B**) corresponding ground truth; (**C-D**) predicted segmentations from (C) ASE_Res_UNet, and (D) U-Net models. (A1-D1): whole image; (A2-D2): first zoomed-in region of interest (ROI) indicated in orange; (A3-D3) second zoom-in ROI indicated in green; and (A4-D4) third zoom-in ROI indicated in blue.

4 Discussion

This work positions ASE_Res_UNet as a novel computationally efficient tool for the segmentation of microtubules in fluorescence microscopy, even under challenging conditions such as background fluorescence noise, inhomogeneous backgrounds, low contrast, and uneven fluorescence intensity. On the *MicSim_FluoMT easy* dataset, it successfully handles background fluorescence often present in microscopy (Table 2; Fig. 2B and S3F). On the *MicSim_FluoMT complex* dataset, it segments filaments with low SNR, as encountered in live-cell imaging (Table 3; Fig. 2C, 3F, 4C). On the *MicReal_FluoMT* dataset and Higaki reannotated dataset, it effectively segments filaments with varying contrast and fluorescence (Fig. 6 & 7). ASE_Res_UNet also demonstrates strong generalisation capabilities. It successfully segments distinct curvilinear structures, retinal blood vessels (*DRIVE* dataset) and retinal nerves (*CORN-1* dataset), despite differences in imaging modalities and structural topology (Tables 5, 6; Fig. 9C, 10C). This highlights model's applicability beyond microtubule segmentation.

A key innovation is the Adaptive Squeeze-and-Excitation (ASE) mechanism, which explicitly incorporates image noise into the attention process. This improves performance, especially in noisy conditions. Indeed, ASE_Res_UNet outperforms SE_Res_UNet on the *MicSim_FluoMT complex* dataset (Table 4; Fig. 4C, F) and produces cleaner segmentations on the *DRIVE* dataset (Table 5; Fig. 9C, F). These gains—particularly in sensitivity and false-positive reduction—can be attributed to its noise-adaptive decoding. The ablation study further reveals the complementary strengths of the model's architectural components (Tables 3 and S3; Fig. 3). Residual blocks enhance gradient flow and support the detection of densely packed filaments. Meanwhile, the ASE attention mechanism proves particularly beneficial in peripheral regions, where microtubule extremities appear with low intensity and are more prone to misclassification (Table S4).

By segmenting microtubules in noisy images, we addressed two main challenges; class imbalance and annotation quality. Microtubules often occupy a small fraction of an image, leading to a dominance of background pixels. While ASE_Res_UNet trained on the *MicSim_FluoMT complex* dataset achieved high scores on conventional metrics (accuracy = 0.9842 ± 0.0006, specificity = 0.9942 ± 0.0005, ROC AUC = 0.9911 ± 0.0011),

we observed that these metrics fail to capture segmentation quality at filament extremities. Therefore, we recommend using Dice, sensitivity, precision, MCC, and PR AUC, which better align with visual inspection. We also found that weighted cross-entropy (WCE), which gives more weight to underrepresented classes, performs better than other loss functions (Table 1). Although effective, WCE remains underused in the literature [49, 62, 75, 77]. The second challenge was the limited availability of high-quality annotations for microtubules in noisy fluorescence images. We addressed this by generating synthetic training data, which offers two advantages: (i) controlled variation of imaging parameters, and (ii) perfect ground truth masks for reliable metric evaluation. (i) Comparing performance on the *MicSim_FluoMT* easy and complex synthetic datasets reveals how the model adapts to different imaging challenges (Tables 2–3; Fig. 2). (ii) Performance of models trained on real datasets can be degraded by inaccuracies/noise in the annotations as we observed in the *MicReal_FluoMT* and *CORN-1* datasets. Although synthetic data often raise concerns about generalisation, our results demonstrate that ASE_Res_UNet developed on synthetic datasets generalises well to four distinct types of real images (Fig. 6, 7, 9 & 10). It even achieves promising results on live-embryo images without retraining (Fig. 8). Future work will explore transfer learning and domain-specific adaptation to further improve performance on more diverse real datasets.

Compared to previous tools for curvilinear structure segmentation—such as SIFNE [28] or SOAX [30], which often require multiple processing steps and manual parameter tuning—ASE_Res_UNet is an end-to-end deep learning solution requiring no handcrafted features or denoising. Unlike many models, it is not limited to a specific structure or imaging modality and remains performant even in noisy conditions. For instance, in contrast to GAN-Based denoising approaches [92, 93], ASE_Res_UNet directly integrates noise adaptation into the attention mechanism, avoiding the need for separate preprocessing steps. Additionally, its lightweight architecture enables deployment in resource-constrained settings and allows training with relatively few annotated images, as shown on the *MicReal_FluoMT* and *DRIVE* datasets. Importantly, ASE_Res_UNet outperforms various models: models with alternative attention mechanisms (CAR-UNet, AG_Res_UNet, SE_Res_UNet), models with distinct architectures (Pix2pix, TransUNet), and a recent microtubule-segmentation framework (SynSeg). Notably, ASE_Res_UNet outperforms CAR-UNet on the *DRIVE* dataset, despite having half the parameters and not being specifically designed for retinal vessel segmentation (Table 5; Figures 9C–D).

5 Conclusion

ASE_Res_UNet is a compact, generalisable, and noise-robust framework for curvilinear structure segmentation across diverse imaging contexts. Its contributions span both basic and translational research. In basic research, ASE_Res_UNet can advance cytoskeletal investigations by enabling reliable segmentation of microtubules, actin, and intermediate filaments, particularly in live-cell microscopy, where noise has previously limited quantitative analysis. This could open new avenues in understanding various cell processes and developing therapies targeting cytoskeletal dysfunction in diseases like cancer or neurodegenerative disorders [6-10]. For instance, ASE_Res_UNet could accelerate high-throughput drug screening by automating cytoskeletal phenotype quantification in noisy live-cell images. In clinical applications, ASE_Res_UNet holds promise for disease diagnosis. For instance, retinal vessel tortuosity is a biomarker for conditions like hypertension, diabetic retinopathy, and cardiovascular disease [94-97], while abnormal cerebral vasculature patterns may assist in early detection of brain tumours [98]. By enabling accurate segmentation in challenging conditions, ASE_Res_UNet bridges foundational research and clinical applications, offering tools for biomedical discovery and diagnostic innovation. Finally, our two novel microtubule segmentation datasets—*MicSim_FluoMT* (synthetic) and *MicReal_FluoMT* (real)—are publicly available to enable community-wide benchmarking. These datasets complement recent efforts—the SynthMT dataset [31] and the Higaki 2024 reannotated dataset [53]—by expanding the range of imaging conditions. These resources provide a comprehensive benchmarking framework to advance noise-robust segmentation methods in microscopy.

Supplementary material

Supplementary material will be available online.

Acknowledgments

We thank Drs Guangshuo Ou and Zhengyang Guo for generously sharing the Higaki 2024 reannotated dataset [53], which was instrumental in evaluating our model's performance. We thank S. Dutertre and X. Pinson of the Microscopy Rennes Imaging Center (MRic, BIOSIT, Biogenouest) for their assistance. MRic is part of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04). We thank Drs. Jacques Pécréaux, Christophe Heligon, Sidi M. Sid'El Moctar, Thierry Pécot, Charles Kervrann, and Florent Autrusseau for helpful discussions about the project. We thank Pécréaux lab and TIAD lab for their support.

Author contributions

Achraf Ait Laydi: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. Louis Cueff: Validation, Software, Methodology, Investigation, Data curation**.** Mewen Crespo: Software, Methodology, Data curation**.** Yousef El Mourabit: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisitions, Data curation, Conceptualization. Hélène Bouvrais: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisitions, Formal analysis, Data curation, Conceptualization.

Funding

This work was supported by PHC Toubkal 2024 (n° 49945RE). H.B. was supported by the Agence Nationale de la Recherche (JCJC project MICENN, ANR-22-CE45-001601) and the University of Rennes (Défis scientifiques, 2020 ; Soutien aux collaborations internationales, 2024). L.C. was supported by a PhD fellowship from "La Ligue Nationale Contre le Cancer" (2021-2024). The servers used for the computations were funded by the Brittany region (AAP PME 2018-2019 - Roboscope) and by the Agence Nationale de la Recherche" (PRCE project SAMIC, ANR-19-CE45-0011). This project was provided with computing AI and storage resources by GENCI at IDRIS thanks to the grants 2024-AD010315542 and 2025-AD010315542R1 on the supercomputer Jean Zay's V100 partition.

Data availability.

Original microtubule datasets (synthetic and real noisy images) are available on Zenodo (DOIs: 10.5281/zenodo.14696279 and 10.5281/zenodo.15852660). ASE_Res_UNet model is also shared on Zenodo (DOI: 10.5281/zenodo.17456868).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Authors declare no conflict of interest.

References

1. Bershadsky AD, Vasiliev JM: Cytoskeleton. 2012.
2. Huber F, Boire A, López MP, Koenderink GH: Cytoskeletal crosstalk: when three different personalities team up. *Current Opinion in Cell Biology* 2015, 32:39-47.
3. Garcin C, Straube A: Microtubules in cell migration. *Essays in Biochemistry* 2019, 63(5):509-520.
4. Matis M: The Mechanical Role of Microtubules in Tissue Remodeling. *BioEssays* 2020, 42(5):1900244.
5. Röper K: Microtubules enter centre stage for morphogenesis. *Philosophical Transactions of the Royal Society B* 2020, 375(1809):20190557.
6. Lafanechère L: The microtubule cytoskeleton: An old validated target for novel therapeutic drugs. *Frontiers in Pharmacology* 2022, 13.
7. Čermák V, Dostál V, Jelínek M, Libusová L, Kovář J, Rösel D, Brábek J: Microtubule-targeting agents and their impact on cancer treatment. *European Journal of Cell Biology* 2020, 99(4):151075.
8. Penna LS, Henriques JAP, Bonatto D: Anti-mitotic agents: Are they emerging molecules for cancer treatment? *Pharmacology & Therapeutics* 2017, 173:67-82.
9. Trojanowski JQ, Smith AB, Huryn D, Lee VM: Microtubule-stabilising drugs for therapy of Alzheimer's disease and other neurodegenerative disorders with axonal transport impairments. 2005.
10. Brunden KR, Lee VM, Smith III AB, Trojanowski JQ, Ballatore C: Altered microtubule dynamics in neurodegenerative disease: Therapeutic potential of microtubule-stabilizing drugs. *Neurobiology of disease* 2017, 105:328-335.
11. Wade RH: On and Around Microtubules: An Overview. *Molecular Biotechnology* 2009, 43(2):177-191.
12. Brouhard GJ, Rice LM: Microtubule dynamics: an interplay of biochemistry and mechanics. *Nature Reviews Molecular Cell Biology* 2018, 19(7):451-463.
13. Waterman-Storer CM: Microtubules and Microscopes: How the Development of Light Microscopic Imaging Technologies Has Contributed to Discoveries about Microtubule Dynamics in Living Cells. *Molecular biology of the cell* 1998, 9(12):3263-3271.
14. Applegate KT, Besson S, Matov A, Bagonis MH, Jaqaman K, Danuser G: plusTipTracker: Quantitative image analysis software for the measurement of microtubule dynamics. *Journal of structural biology* 2011, 176(2):168-184.
15. Matov A, Applegate K, Kumar P, Thoma C, Krek W, Danuser G, Wittmann T: Analysis of microtubule dynamic instability using a plus-end growth marker. *Nature methods* 2010, 7(9):761-768.
16. Akhmanova A, Steinmetz MO: Microtubule+ TIPs at a glance. *Journal of cell science* 2010, 123(20):3415-3419.
17. Lansbergen G, Akhmanova A: Microtubule plus end: a hub of cellular activities. *Traffic* 2006, 7(5):499-507.
18. Théry M, Blanchoin L: Microtubule self-repair. *Current Opinion in Cell Biology* 2021, 68:144-154.
19. Cross RA: Microtubule lattice plasticity. *Current Opinion in Cell Biology* 2019, 56:88-93.
20. Romeiro Motta M, Biswas S, Schaedel L: Beyond uniformity: Exploring the heterogeneous and dynamic nature of the microtubule lattice. *European Journal of Cell Biology* 2023, 102(4):151370.
21. Janke C, Magiera MM: The tubulin code and its role in controlling microtubule properties and functions. *Nature Reviews Molecular Cell Biology* 2020.
22. Icha J, Weber M, Waters JC, Norden C: Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* 2017, 39(8):1700003.
23. Skylaki S, Hilsenbeck O, Schroeder T: Challenges in long-term imaging and quantification of single-cell dynamics. *Nature Biotechnology* 2016, 34(11):1137-1144.
24. Özdemir B, Reski R: Automated and semi-automated enhancement, segmentation and tracing of cytoskeletal networks in microscopic images: A review. *Comput Struct Biotechnol J* 2021, 19:2106-2120.
25. Sargin M, Altinok A, Kiris E, Feinstein S, Wilson L, Rose K, Manjunath B: Tracing microtubules in live cell images. In: *Biomedical Imaging: From Nano to Macro, 2007 ISBI 2007 4th IEEE International Symposium on: 2007*. IEEE: 296-299.
26. Kamath K, Oroudjev E, Jordan MA: Determination of microtubule dynamic instability in living cells. *Methods in cell biology* 2010, 97:1-14.
27. Pallavicini C, Levi V, Wetzler DE, Angiolini JF, Benseñor L, Despósito MA, Bruno L: Lateral Motion and Bending of Microtubules Studied with a New Single-Filament Tracking Routine in Living Cells. *Biophysical journal* 2014, 106(12):2625-2635.
28. Zhang Z, Nishimura Y, Kanchanawong P: Extracting microtubule networks from superresolution single-molecule localization microscopy data. *Molecular biology of the cell* 2017, 28(2):333-345.
29. Smith MB, Li H, Shen T, Huang X, Yusuf E, Vavylonis D: Segmentation and tracking of cytoskeletal filaments using open active contours. *Cytoskeleton* 2010, 67(11):693-705.
30. Xu T, Vavylonis D, Tsai F-C, Koenderink GH, Nie W, Yusuf E, Lee IJ, Wu J-Q, Huang X: SOAX: A software for quantification of 3D biopolymer networks. *Scientific Reports* 2015, 5(1):9081.

31. Koddenbrock M, Westerhoff J, Fachet D, Reber S, Gers F, Rodner E: Synthetic data enables human-grade microtubule analysis with foundation models for segmentation. *bioRxiv* 2026:2026.2001.2009.698597.
32. Frangi AF, Niessen WJ, Vincken KL, Viergever MA: Multiscale vessel enhancement filtering. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1: 1998*. Springer: 130-137.
33. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ: Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on medical Imaging* 2006, 25(9):1214-1222.
34. Vicas C, Nedevschi S: Detecting curvilinear features using structure tensors. *IEEE Transactions on Image Processing* 2015, 24(11):3874-3887.
35. Cetin S, Unal G: A higher-order tensor vessel tractography for segmentation of vascular structures. *IEEE transactions on medical imaging* 2015, 34(10):2172-2185.
36. Moreno R, Smedby Ö: Gradient-based enhancement of tubular structures in medical images. *Medical Image Analysis* 2015, 26(1):19-29.
37. Alharbi SS, Sazak Ç, Nelson CJ, Alhasson HF, Obara B: The multiscale top-hat tensor enables specific enhancement of curvilinear structures in 2D and 3D images. *Methods* 2020, 173:3-15.
38. Al-Diri B, Hunter A, Steel D: An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical imaging* 2009, 28(9):1488-1497.
39. Zhao Y, Liu Y, Wu X, Harding SP, Zheng Y: Retinal vessel segmentation: An efficient graph cut approach with retinex and local phase. *PloS one* 2015, 10(4):e0122332.
40. Chen Y, Zhang Y, Yang J, Cao Q, Yang G, Chen J, Shu H, Luo L, Coatrieux J-L, Feng Q: Curve-like structure extraction using minimal path propagation with backtracking. *IEEE Transactions on image processing* 2015, 25(2):988-1003.
41. Liu L, Chen D, Shu M, Li B, Shu H, Paques M, Cohen LD: Trajectory grouping with curvature regularization for tubular structure tracking. *IEEE Transactions on Image Processing* 2021, 31:405-418.
42. Xu T, Langouras C, Koudehi MA, Vos BE, Wang N, Koenderink GH, Huang X, Vavylonis D: Automated Tracking of Biopolymer Growth and Network Deformation with TSOAX. *Scientific Reports* 2019, 9(1):1717.
43. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. In*: 2015; Cham*. Springer International Publishing: 234-241.
44. Cortinovis D, Srl O: Retina blood vessel segmentation with a convolution neural network (u-net). In.; 2016.
45. Siddique N, Sidike P, Elkin C, Devabhaktuni V: U-Net and its variants for medical image segmentation: theory and applications. *arXiv preprint arXiv:201101118* 2020.
46. Li D, Dharmawan DA, Ng BP, Rahardja S: Residual U-Net for Retinal Vessel Segmentation. In: *2019 IEEE International Conference on Image Processing (ICIP): 22-25 Sept. 2019 2019*. 1425-1429.
47. Huang K-W, Yang Y-R, Huang Z-H, Liu Y-Y, Lee S-H: Retinal Vascular Image Segmentation Using Improved UNet Based on Residual Module. *Bioengineering* 2023, 10(6):722.
48. Si Z, Fu D, Li J: U-Net with attention mechanism for retinal vessel segmentation. In: *International Conference on Image and Graphics: 2019*. Springer: 668-677.
49. Rong Y, Xiong Y, Li C, Chen Y, Wei P, Wei C, Fan Z: Segmentation of retinal vessels in fundus images based on U-Net with self-calibrated convolutions and spatial attention modules. *Medical & Biological Engineering & Computing* 2023, 61(7):1745-1755.
50. Li D, Peng L, Peng S, Xiao H, Zhang Y: Retinal vessel segmentation by using AFNet. *The Visual Computer* 2023, 39(5):1929-1941.
51. Lee H-C, Cherng ST, Miotto R, Dudley JT: Enhancing high-content imaging for studying microtubule networks at large-scale. *arXiv preprint arXiv:191000662* 2019.
52. Horiuchi R, Kamimura A, Hanaki Y, Matsumoto H, Ueda M, Higaki T: Deep learning-based cytoskeleton segmentation for accurate high-throughput measurement of cytoskeleton density. *Protoplasma* 2025, 262(3):739-751.
53. Guo Z, Wang Z, Chen Z, Xu K, Chai Y, Ke J, Huang J, Ye Y, Wang H, Zhang J *et al*: SynSeg: A synthetic data-driven approach for robust subcellular structure segmentation. *Journal of Cell Biology* 2025, 225(3).
54. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B: Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging* 2004, 23(4):501-509.
55. iMed: CORN: Corneal nerve fiber dataset. In.: Zenodo; 2024.
56. Nedelec F, Foethke D: Collective Langevin dynamics of flexible cytoskeletal fibers. *New Journal of Physics* 2007, 9(11):427.
57. Dmitrieff S, Nédélec F: ConfocalGN: A minimalistic confocal image generator. *SoftwareX* 2017, 6:243-247.
58. Bouvrais H, Crespo M: MicSim_FluoMT: Two synthetic datasets of images of fluorescent microtubules (Ait Laydi et al., 2025). In.: Zenodo; 2025.
59. Cueff L, Huet E, Schmitt L, Pastezeur S, Coquil M, Savary T, Pécréaux J, Bouvrais H: Microtubule stiffening by doublecortin-domain protein ZYG-8 contributes to spindle orientation during C. elegans zygote division. *bioRxiv* 2025:2024.2011.2029.624795.

60. Cueff L, Pecreaux J, Bouvrais H: MicReal_FluoMT: A dataset of microscopy images with stained microtubules (Ait Laydi et al., 2025). In.: Zenodo; 2025.
61. Mou L, Zhao Y, Fu H, Liu Y, Cheng J, Zheng Y, Su P, Yang J, Chen L, Frangi AF *et al*: CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical Image Analysis* 2021, 67:101874.
62. Zhu Z, An Q, Wang Z, Li Q, Fang H, Huang Z: ILU-Net: Inception-Like U-Net for retinal vessel segmentation. *Optik* 2022, 260:169012.
63. Wu C, Zou Y, Yang Z: U-GAN: Generative Adversarial Networks with U-Net for Retinal Vessel Segmentation. In: *2019 14th International Conference on Computer Science & Education (ICCSE): 19-21 Aug. 2019 2019*. 642-646.
64. Yıldız E, Arslan AT, Yıldız Taş A, Acer AF, Demir S, Şahin A, Erol Barkana D: Generative Adversarial Network Based Automatic Segmentation of Corneal Subbasal Nerves on In Vivo Confocal Microscopy Images. *Translational Vision Science & Technology* 2021, 10(6):33-33.
65. Hu J, Shen L, Sun G: Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2018*. 7132-7141.
66. Chowdary GJ, G S, M P, Yogarajah P: Nucleus segmentation and classification using residual SE-UNet and feature concatenation approach incervical cytopathology cell images. *Technology in Cancer Research & Treatment* 2023, 22:15330338221134833.
67. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2016*. 770-778.
68. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S: On evaluation metrics for medical applications of artificial intelligence. *Scientific reports* 2022, 12(1):5979.
69. Schlosser T, Friedrich M, Meyer T, Kowerko D: A consolidated overview of evaluation and performance metrics for machine learning and computer vision. *Tobias Schlosser, Michael Friedrich, Trixy Meyer, and Danny Kowerko–Junior Professorship of Media Computing, Chemnitz University of Technology* 2024, 9107.
70. Saito T, Rehmsmeier M: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 2015, 10(3):e0118432.
71. Chicco D, Jurman G: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* 2020, 21:1-13.
72. Jadon S: A survey of loss functions for semantic segmentation. In: *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB): 2020*. IEEE: 1-7.
73. Rezaei-Dastjerdehei MR, Mijani A, Fatemizadeh E: Addressing imbalance in multi-label classification using weighted cross entropy loss function. In: *2020 27th national and 5th international iranian conference on biomedical engineering (ICBME): 2020*. IEEE: 333-338.
74. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y: Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:210204306* 2021.
75. Guo C, Szemenyei M, Hu Y, Wang W, Zhou W, Yi Y: Channel Attention Residual U-Net for Retinal Vessel Segmentation. In: *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): 6-11 June 2021 2021*. 1185-1189.
76. Wang J, Li X, Lv P, Shi C: SERR-U-Net: Squeeze-and-Excitation Residual and Recurrent Block-Based U-Net for Automatic Vessel Segmentation in Retinal Image. *Computational and Mathematical Methods in Medicine* 2021, 2021(1):5976097.
77. Ma Z, Li X: An improved supervised and attention mechanism-based U-Net algorithm for retinal vessel segmentation. *Computers in Biology and Medicine* 2024, 168:107770.
78. Zhou Y, Zhong L, Wang Z, Ge Y: A semi-supervised fracture-attention model for segmenting tubular objects with improved topological connectivity. *Bioinformatics* 2025, 41(1).
79. Huang J, Luo Y, Guo Y, Li W, Wang Z, Liu G, Yang G: Accurate segmentation of intracellular organelle networks using low-level features and topological self-similarity. *Bioinformatics* 2024, 40(10).
80. Wu C, Zou Y, Yang Z: U-GAN: Generative Adversarial Networks with U-Net for Retinal Vessel Segmentation. *2019 14th International Conference on Computer Science & Education (ICCSE)* 2019:642-646.
81. Yue C, Ye M, Wang P, Huang D, Lu X: SRV-GAN: A generative adversarial network for segmenting retinal vessels. *Math Biosci Eng* 2022, 19(10):9948-9965.
82. Yang T, Wu T, Li L, Zhu C: SUD-GAN: Deep Convolution Generative Adversarial Network Combined with Short Connection and Dense Block for Retinal Vessel Segmentation. *Journal of Digital Imaging* 2020, 33(4):946-957.
83. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B: Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:180403999* 2018.
84. Isola P, Zhu J-Y, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2017*. 1125-1134.
85. Chen J, Mei J, Li X, Lu Y, Yu Q, Wei Q, Luo X, Xie Y, Adeli E, Wang Y *et al*: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis* 2024, 97:103280.

86. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M: Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In*: 2023; Cham*. Springer Nature Switzerland: 205-218.
87. Liu X, Gao P, Yu T, Wang F, Yuan R-Y: CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation. *Information Fusion* 2025, 113:102634.
88. Zhang J, Qin Q, Ye Q, Ruan T: ST-unet: Swin transformer boosted U-net with cross-layer feature enhancement for medical image segmentation. *Computers in Biology and Medicine* 2023, 153:106516.
89. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al*: Fiji: an open-source platform for biological-image analysis. *Nature methods* 2012, 9(7):676-682.
90. Vatani N: Gray level image edge detection using a hybrid model of cellular learning automata and stochastic cellular automata. *OALib* 2015.
91. Phansalkar N, More S, Sabale A, Joshi M: Adaptive local thresholding for detection of nuclei in diversity stained cytology images. In: *2011 International conference on communications and signal processing: 2011*. IEEE: 218-220.
92. Saidulu N, Muduli PR: Asymmetric Convolution-based GAN Framework for Low-Dose CT Image Denoising. *Computers in Biology and Medicine* 2025, 190:109965.
93. Huang Z, Zhang J, Zhang Y, Shan H: DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising. *IEEE Transactions on Instrumentation and Measurement* 2021, 71:1-12.
94. Cheung CY-l, Zheng Y, Hsu W, Lee ML, Lau QP, Mitchell P, Wang JJ, Klein R, Wong TY: Retinal Vascular Tortuosity, Blood Pressure, and Cardiovascular Risk Factors. *Ophthalmology* 2011, 118(5):812-818.
95. Sasongko MB, Wong TY, Nguyen TT, Cheung CY, Shaw JE, Wang JJ: Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. *Diabetologia* 2011, 54(9):2409-2416.
96. DuPont M, Lambert S, Rodriguez-Martin A, Hernandez O, Lagatuz M, Yilmaz T, Foderaro A, Baird GL, Parsons-Wingerter P, Lahm T: Retinal vessel changes in pulmonary arterial hypertension. *Pulmonary Circulation* 2022, 12(1):e12035.
97. Xue CC, Li C, Hu JF, Wei CC, Wang H, Ahemaitijiang K, Zhang Q, Chen DN, Zhang C, Li F *et al*: Retinal vessel caliber and tortuosity and prediction of 5-year incidence of hypertension. *Journal of Hypertension* 2023, 41(5):830-837.
98. Bullitt E, Zeng D, Gerig G, Aylward S, Joshi S, Smith JK, Lin W, Ewend MG: Vessel tortuosity and brain tumor malignancy: a blinded study1. *Academic radiology* 2005, 12(10):1232-1240.