ADAPTIVE ATTENTION RESIDUAL U-NET FOR FILAMENT SEGMENTATION IN BIOMEDICAL MICROSCOPY IMAGES

Achraf Ait Laydi, Mewen Crespo, Yousef El Mourabit, Hélène Bouvrais

This article was accepted at ISBI 2025 (International Symposium on Biomedical Imaging, April 14–17, 2025, Houston, USA). However, due to the authors' inability to obtain a visa, they were unable to attend the conference. As a result, the article was not published in the ISBI 2025 conference proceedings.

ADAPTIVE ATTENTION RESIDUAL U-NET FOR FILAMENT SEGMENTATION IN BIOMEDICAL MICROSCOPY IMAGES

Achraf Ait Laydi[1,2], Mewen Crespo[1,◊], Yousef El Mourabit[2], Hélène Bouvrais[1]

[1] CNRS, IGDR (Institute Genetics and Development of Rennes) – UMR 6290, Rennes, France.
[2] TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay Slimane University, Morocco.
[◊] Current affiliation: IRMAR, University of Rennes, Rennes, France.

ABSTRACT

Segmenting filamentous structures in biomedical images is challenging, particularly with noisy images and dense filament networks such as those found *in vivo*. To address this, we created two extensive datasets of synthetic images of fluorescently labeled microtubules, precisely annotated and mimicking real images in terms of noise. The fluorescence intensities varied along the filaments in the second dataset complicating segmentation further. Next, we developed an advanced architecture, the Adaptive Squeeze-and-Excitation Residual U-Net (ASE_Res_Unet), incorporating residual blocks and adaptive SE attention mechanisms and compared it with U-Net. Both visual and quantitative evaluations showed that ASE_Res_Unet outperformed U-Net and two state-of-the art models, especially in segmenting the most challenging dataset. This indicates improved noise handling and the ability to capture finer structural details. Our solution effectively segments noisy or low-contrast biomedical images, demonstrated by our success in segmenting retinal blood vessels, offering promising applications in disease diagnosis and treatment.

Index Terms— Filament segmentation, Noisy or low contrast images, U-Net based architecture, Residual blocks, Attention module.

1. INTRODUCTION

The analysis of filamentous structures in biomedical images is essential for disease diagnosis and clinical studies. Filaments, such as blood vessels, neurons, and fibrils, play critical roles in various biological processes and provide important insights into health status [1-3]. For instance, retinal blood vessel tortuosity can indicate cardiovascular diseases, while brain vessel patterns may aid early tumor detection [4,5]. Understanding cytoskeletal networks like microtubules and actin filaments is vital for elucidating their function in cells and tissues [6], with microtubules being a key target in cancer therapy [7]. Despite their importance, extracting geometric features—such as the number, length, or curvature of filaments—from noisy images is challenging. Biomedical images often suffer from low contrast and noise, obscuring the details needed for precise analysis. Filaments may also display uneven fluorescent intensity or appear sparsely visible.

Classical methods like Gabor filters and wavelet transforms have been used to enhance filamentous structures [8], and shape-based algorithms for segmentation [9]. However, these techniques struggle with noise and may introduce inaccuracies in geometric feature extraction, such as misidentified boundaries affecting filament curvature or length. Deep learning, particularly U-Net and its variants with residual and attention mechanisms, has shown great promise in segmenting filamentous structures [10-12]. These networks can handle fluorescence intensity and filament morphology variations but still face challenges with noisy, low-contrast images. This is especially true for live microscopic images of cytoskeletal filaments like microtubules, where fast acquisition often leads to low and noisy fluorescent signals, complicating accurate detection and segmentation.

To overcome these challenges, we developed a pipeline for generating synthetic images of fluorescently labeled microtubules that closely mimic real microscopy data, incorporating variations in fluorescence intensity, noise, and filament morphology. We also presented the Adaptive Squeeze-and-Excitation Residual U-Net (ASE_Res_Unet) specifically designed for segmenting filaments in these synthetic images and we assessed its effectiveness in isolating and detecting filaments in noisy image, where traditional methods might fail.

2. METHODS

2.1 Dataset

To tackle the challenges of segmenting microtubules in environments with high levels of background noise, we needed a dataset of hundreds of noisy images to train the neural network. However, generating a real dataset of noisy images is difficult for two main reasons. (1) Collecting and annotating hundreds of images is both costly and time-consuming. (2) The high noise levels make accurate annotations unfeasible or imprecise. As a solution, we opted to create a dataset of synthetic biological images of

fluorescently labeled microtubules along with their corresponding binary images indicating microtubule positions (ground truth). This approach allowed us to generate a large, fully annotated dataset.

We developed a two-step pipeline for generating synthetic images and their corresponding ground-truths. First, we simulated the microtubule mitotic network in *Caenorhabditis elegans* zygote, a well-established model for cell division [13], using the cytoskeletal simulator *Cytosim* [14]. Next, we produced synthetic images based on the simulations using the *ConfocalGN* image generator [15] and created binary masks to accurately annotate the microtubules in each image. These annotations were automatically derived from the simulations, ensuring precise alignment between the simulated microtubules and their corresponding segmentation masks. The masks served as ground truth data for training supervised deep learning models and evaluating the performance of the segmentation algorithms. We conducted 500 simulation runs and selected a few time points from each run to construct two datasets, each comprising 1192 synthetic 2D images, each sized 666x666 pixels. The images varied in complexity, ranging from isolated microtubules in low-noise environments to more challenging scenarios with high noise and overlapping filaments. Figure 1 provides an overview of the dataset, showcasing examples of synthetic images along with their corresponding annotations. Both datasets are accessible on Zenodo (DOI 10.5281/zenodo.14696280). These examples highlighted the diversity in fluorescence intensity, microtubule density, and noise level, emphasizing the challenges presented by the dataset and its value in training and validating deep learning models for microtubule segmentation.

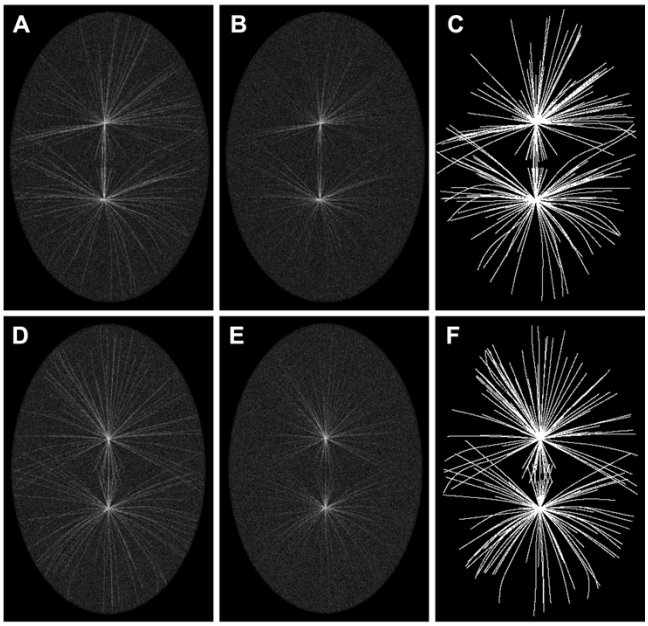

Fig. 1. Example images from two synthetic datasets with cropping of the surrounding dark area to improve filament visualization: (A, D) simple images with uniform fluorescence along the filaments and an SNR of 1.7 dB; (B, E) complex images featuring uneven fluorescence and an SNR of 1.2 dB; (C, F) the corresponding ground truths.

2.2 Proposed method

To enhance the performance of the U-Net architecture for segmenting filamentous structures in noisy images, we implemented two key modifications as illustrated in Figure 2: residual blocks in the encoder [16] and Adaptative Squeeze-and-Excitation (ASE) attention modules in the decoder [17]. Additionally, we used transposed convolution layers for reconstruction in the decoder [18]. Our modified U-Net consisted of four down-sampling (encoder) blocks and four up-sampling (decoder) blocks. Each encoder block comprised max-pooling followed by two distinct convolutional layers. Each of these layers was accompanied by batch normalization and ReLU activation. This was then succeeded by a residual block that incorporates a skip connection, enhancing gradient flow and preserving essential features as the spatial resolution decreases [16]. Transposed convolution layers were used for up-sampling to ensure precise spatial resolution recovery while maintaining learned patterns from the encoding phase [18]. After each up-sampling step, the output was concatenated with the corresponding feature map from the encoder. This was followed by two convolutional layers with batch normalization and ReLU activation and an ASE attention module in each decoder block. To address the challenges posed by variations in noise levels, we decided to modify the traditional SE attention mechanism [17]. We proposed an adaptive one that dynamically adjusts its attention weights based on the specific noise characteristics of the input images. The ASE attention module works by analyzing the noise level in the images and adapting the attention strengths accordingly. This adaptation involves noise level assessment, dynamic weight adjustment and channel-wise modulation. We called the proposed architecture: ASE_Res_Unet.

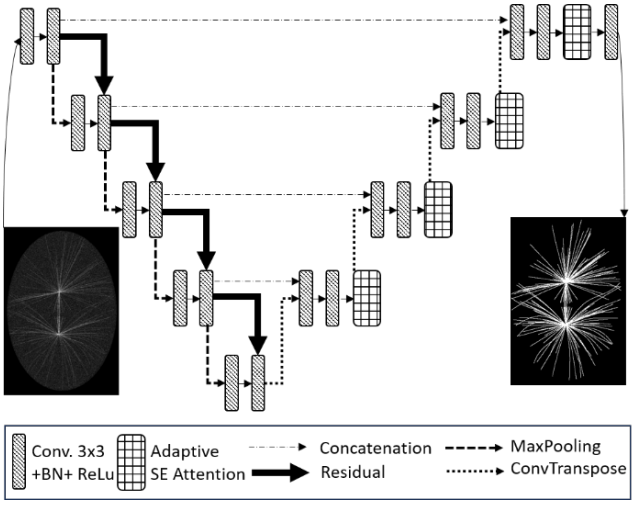

Fig. 2. Proposed architecture of ASE_Res_Unet.

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

For weight initialization of ASE_Res_Unet, we utilized pretrained weights from a U-Net model trained on our specific microtubule datasets (simple or complex). This approach ensured consistency in the learned features and facilitated efficient fine-tuning of the proposed architecture.

We tested several loss functions including binary cross-entropy (BCE), Focal Loss, Dice Loss—to improve the model's performance, particularly in handling noisy and imbalanced data. Based on the results, we chose the Weighted binary Cross-Entropy (WCE), setting the weight for the microtubules (weight_one) to 1 and the background one (weight_zero) to 0.25 [19]. This choice reflected the unbalanced nature of our dataset, where the background is significantly more prevalent than the microtubules.

$$\boldsymbol{WCE}(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \text{weight\_one} \cdot \log(\hat{y}_i) + (1-y_i) \cdot \text{weight\_zero} \cdot \log(1-\hat{y}_i)]$$

Besides, to assess model performance, we utilized three metrics [20-21]: the Dice coefficient, intersection-over-union (IoU) and accuracy. During training, we tuned several hyperparameters, including learning rate, batch size, filter size, and weighting factors for WCE loss. The learning rate started at 0.001 and was dynamically reduced based on validation performance. We used a batch size of 2 to balance memory efficiency with convergence speed. The Adam optimizer was configured with the parameters β_1 and β_2 equal to 0.9 and 0.999, ensuring stable and efficient gradient updates. The model training was terminated when the validation loss stabilized over 10 consecutive epochs, indicating convergence. A filter size of 3x3 was chosen for its reduced sensitivity to noise and its effectiveness in capturing fine details, such as edges and corners.

3. EXPIREMENTS AND RESULTS

3.1. Performance of the model

Model	Loss (WCE)	Dice	IoU	Accuracy
U-Net	0.028	0.921	0.979	99.4%
ASE_Res_Unet	0.038	0.924	0.984	99.4%

Table. 1. Quantitative performance on the simple dataset.

We evaluated the performance of two architectures, a classic U-Net and the proposed model ASE_Res_Unet, using our newly generated synthetic microtubule dataset. Both models were trained and tested on the simple and complex datasets under identical conditions without any image preprocessing step. As seen in Table 1, the U-Net architecture already achieved high performance across all metrics on the simple dataset. The differences between the U-Net and our ASE_Res_Unet model in this case were relatively small, indicating that the simpler dataset did not fully test the capabilities of the advanced ASE_Res_Unet architecture. The U-Net effectively captured most of the structural information from the filaments, leaving limited opportunities for improvement even with an advanced model. We foresaw that the complex dataset, with varying fluorescence intensity along the microtubules, would more effectively challenge the capabilities of the advanced architecture. Interestingly, our ASE_Res_Unet architecture outperformed U-Net, especially in terms of the IoU and Dice coefficient (Table 2). This improvement suggested that the integration of residual blocks and ASE attention mechanisms enhanced the model's ability to capture the intricate details of microtubule structures, even under challenging conditions such as noise and intensity variations.

Model	Loss (WCE)	Dice	IoU	Accuracy
U-Net	0.044	0.707	0.905	95.4%
ASE_Res_Unet	0.082	0.783	0.966	98.3%

Table. 2. Quantitative performance on the complex dataset

3.2. Visual quality assessment

To complement the metric-based comparison of the two architectures, we performed a qualitative analysis by examining segmentation results alongside the microtubule ground-truth across several test images from both the simple and complex datasets. While visual inspection of the simple image predictions showed no significant difference, the analysis of the complex image predictions corroborated our metric findings: the advanced architecture ASE_Res_Unet outperformed U-Net (Figure 3).

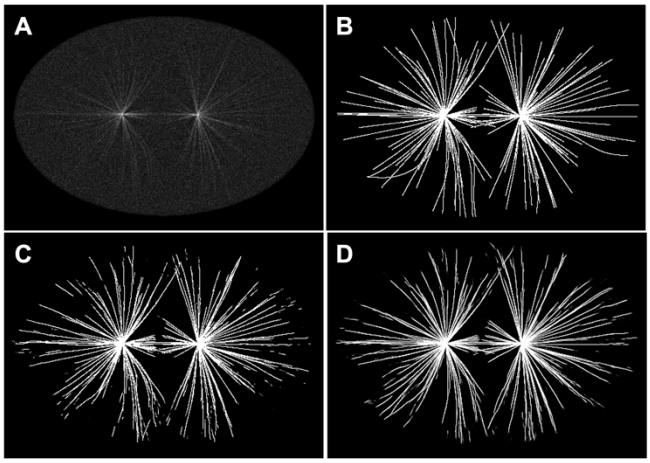

Fig. 3. (A) Input complex image; (B) ground truth; (C) U-Net prediction; (D) ASE_Res_Unet prediction.

The U-Net model produced a significantly higher number of false predictions (pixel count across 120 test images: U-Net: 404209, ASE_Res_Unet: 379910) and

recovered fewer true microtubules (pixel count across 120 test images: U-Net: 1475509, ASE_Res_Unet: 1514037). It suggested that the standard U-Net struggled more with noise, resulting in incorrect segmentation in certain regions, especially at the edges where filament intensity is reduced. This visual analysis demonstrated that our proposed ASE_Res_Unet architecture significantly improved segmentation accuracy and noise robustness compared to the standard U-Net. It also identified areas for improvement, especially in segmenting the filament extremities with lower intensities in the complex dataset. We proposed that the integration of the ASE attention module and residual blocks enhanced feature extraction, enabling the model to better differentiate between relevant structures and background noise.

3.3 Comparison with other advanced architectures

To further benchmark our approach, we compared our architecture with existing state-of-the-art methods using the complex dataset. We selected two deep learning architectures, CAR U-Net and Res-Net, for comparison due to their strong performance in filament segmentation tasks [22-23]. We trained these two architectures on the complex dataset and evaluated their performance. Our model ASE_Res_Unet demonstrated superior results, surpassing both Res-UNet and CAR U-Net in effectively segmenting dense networks of filamentous structures in noisy images (Table 3).

Model	Loss (WCE)	Dice	IoU	Accuracy
Res-Unet [22]	0.059	0.751	0.939	97.2%
CAR U-Net [23]	0.073	0.766	0.954	98.2%
ASE_Res_Unet	0.082	0.783	0.966	98.3%

Table. 3. Quantitative performance comparison of state-of-the-art models on the complex dataset.

3.4 Use of ASE_Res_Unet to segment retinal blood vessels in biomedical images

To evaluate the applicability of our ASE_Res_Unet architecture for segmenting other filamentous structures, we used the DRIVE (Digital Retinal Images for Vessel Extraction) dataset [24]. The dataset comprises 40 high-resolution retinal vessel images, complete with their manual annotations serving as ground truth, and is widely recognized for benchmarking segmentation methods. Besides, it presents challenges such as illumination variations, complex vessel branching, and varying vessel widths.

We trained our architecture on the DRIVE dataset images to learn how to segment intricate vascular structures. Testing the segmentation of vessels confirmed the robustness and generalization capabilities of ASE_Res_Unet, which effectively captured complex vessel patterns under varying conditions, as illustrated in Figure 4 (WCE loss = 0.082, Dice = 0.795, IoU = 0.85). Additionally, the metrics in Table 4 further validated our model's performance in comparison to two previous models specifically developed for retinal blood vessel segmentation.

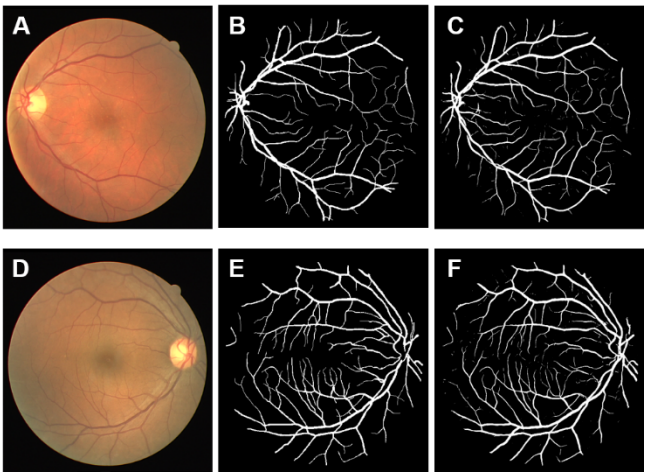

Fig. 4. (A, D) Test images from DRIVE dataset; (B, E) the corresponding ground truths; (C, F) ASE_Res_Unet predictions.

Model	U-Net [10]	DDNet [25]	ASE_Res_Unet
Accuracy	95.3%	95.9%	96.7%

Table. 4. Quantitative results on DRIVE dataset.

4. CONCLUSION

The proposed ASE_Res_Unet architecture demonstrated significant improvements in segmentation performance, particularly on the complex dataset, compared to other U-Net-based models, making it a robust tool for filament segmentation in challenging noisy images. The use of residual blocks preserved features from earlier layers, enhancing the model's ability to capture details. Additionally, the incorporation of ASE attention modules dynamically recalibrated feature importance, helping to reduce noise and improve differentiation between background noise and relevant structures. Pre-training the model with U-Net weights contributed to faster convergence and enhanced generalization. Future works will focus on (1) refining the architecture for segmenting microtubules in peripheral regions with minimal fluorescence, (2) applying the model to real biological images through transfer learning, and (3) expanding its application to other filamentous structures. This architecture offers an effective solution for biomedical image segmentation tasks in cases of low contrast or high background noise with potential applications in diverse clinical and diagnostic areas, e.g. disease detection and analysis through retinal blood vessel segmentation as we illustrated [3,5].

5. ACKNOWLEDGMENTS

This work was supported by PHC Toubkal 2024 (n° 49945RE). H.B. was supported by ANR grant (MICENN) and the University of Rennes (Défis scientifiques, 2020; Soutien aux collaborations internationales, 2024). We thank Pécréaux lab and TIAD lab for their support and helpful discussions.

6. COMPLIANCE WITH ETHICAL STANDARDS

This works includes numerical simulations and deep-learning studies for which no ethical approval was required.

7. REFERENCES

[1] E. Bullitt *et al.*, « Vessel Tortuosity and Brain Tumor Malignancy: A Blinded Study1 », *Acad. Radiol.*, vol. 12, nº 10, p. 1232-1240, oct. 2005.

[2] B. S. Shaheen, M. Bakir, et S. Jain, « Corneal nerves in health and disease », *Surv. Ophthalmol.*, vol. 59, nº 3, p. 263-285, mai 2014.

[3] F. Kahe *et al.*, « Coronary artery tortuosity: a narrative review », *Coron. Artery Dis.*, vol. 31, nº 2, p. 187, mars 2020.

[4] E. A. Grace, C. A. Rabiner, et J. Busciglio, « Characterization of neuronal dystrophy induced by fibrillar amyloid β: implications for Alzheimer's disease », *Neuroscience*, vol. 114, nº 1, p. 265-273, sept. 2002.

[5] C. Y. Cheung *et al.*, « Retinal Vascular Tortuosity, Blood Pressure, and Cardiovascular Risk Factors », *Ophthalmology*, vol. 118, nº 5, p. 812-818, mai 2011.

[6] A. D. Bershadsky et J. M. Vasiliev, *Cytoskeleton*. Springer Science & Business Media, 2012.

[7] M. A. Jordan et L. Wilson, « Microtubules as a target for anticancer drugs », *Nat. Rev. Cancer*, vol. 4, nº 4, p. 253-265, avr. 2004.

[8] C. Gottschlich, « Curved-Region-Based Ridge Frequency Estimation and Curved Gabor Filters for Fingerprint Image Enhancement », *IEEE Trans. Image Process.*, vol. 21, nº 4, p. 2220-2227, avr. 2012.

[9] « Segmentation and tracking of cytoskeletal filaments using open active contours - Smith - 2010 - Cytoskeleton - Wiley Online Library ».

[10] O. Ronneberger, P. Fischer, et T. Brox, « U-Net: Convolutional Networks for Biomedical Image Segmentation », in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi, Éd., Cham: Springer International Publishing, 2015.

[11] Z. Li, H. Zhang, Z. Li, et Z. Ren, « Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation », *Appl. Sci.*, vol. 12, nº 14, Art. nº 14, janv. 2022.

[12] Y. Wan *et al.*, « Retinal Blood Vessels Segmentation With Improved SE-UNet Model », *Int. J. Imaging Syst. Technol.*, vol. 34, nº 4, p. e23145, 2024.

[13] N. Hattersley *et al.*, « Employing the one-cell *C. elegans* embryo to study cell division processes », in *Methods in Cell Biology*, vol. 144, H. Maiato et M. Schuh, Éd., in Mitosis and Meiosis Part A, vol. 144. , Academic Press, 2018, p. 185-231.

[14] F. Nedelec et D. Foethke, « Collective Langevin dynamics of flexible cytoskeletal fibers », *New J. Phys.*, vol. 9, nº 11, p. 427, nov. 2007.

[15] S. Dmitrieff et F. Nédélec, « ConfocalGN: A minimalistic confocal image generator », *SoftwareX*, vol. 6, p. 243-247, janv. 2017.

[16] E. Goceri, « Analysis of Deep Networks with Residual Blocks and Different Activation Functions: Classification of Skin Diseases », in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, nov. 2019.

[17] J. Hu, L. Shen, S. Albanie, G. Sun, et E. Wu, « Squeeze-and-Excitation Networks », 16 mai 2019, *arXiv*: arXiv:1709.01507.

[18] F. Sahito, P. Zhiwen, F. Sahito, et J. Ahmed, « Transpose convolution based model for super-resolution image reconstruction », *Appl. Intell.*, vol. 53, nº 9, p. 10574-10584, mai 2023.

[19] « Addressing Imbalance in Multi-Label Classification Using Weighted Cross Entropy Loss Function | IEEE Conference Publication | IEEE Xplore ».

[20] K. H. Zou *et al.*, « Statistical validation of image segmentation quality based on a spatial overlap index1 », *Acad. Radiol.*, vol. 11, nº 2, p. 178-189, févr. 2004.

[21] W. R. Crum, O. Camara, et D. L. G. Hill, « Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis », *IEEE Trans. Med. Imaging*, vol. 25, nº 11, p. 1451-1461, nov. 2006.

[22] L. Huang, A. Miron, K. Hone, et Y. Li, « Segmenting Medical Images: From UNet to Res-UNet and nnUNet », in *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, juin 2024, p. 483-489.

[23] C. Guo, M. Szemenyei, Y. Hu, W. Wang, W. Zhou, et Y. Yi, « Channel Attention Residual U-Net for Retinal Vessel Segmentation », in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, juin 2021, p. 1185-1189.

[24] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, et B. van Ginneken, « Ridge-based vessel segmentation in color images of the retina », *IEEE Trans. Med. Imaging*, vol. 23, nº 4, p. 501-509, avr. 2004.

[25] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao and J. Liu, "DenseDilated Network With Probability Regularized Walk for Vessel Detection," in *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1392-1403.