A Deep Learning Approach to Vessel Segementation:UNET

Siddh Sanghvi

Sahil Shah

Vansh Shah

In this research endeavor, we address the crucial challenge of retinal vessel segmentation which is of great use when diagnosing for retinal vascular dieaseas. Our primary goal is to develop an effective and adaptable segmentation system capable of achieving precision, even when confronted with limited annotated data.Our project is focused on establishing a comprehensive pipeline for the training and validation of a U-Net-based segmentation system. In this pursuit, we have utilized an existing model as a foundational template and fine-tuned it to create distinct, referenced models. Each of these models is meticulously tailored with specific hyperparameters and trained on domain-specific datasets, purposefully designed to excel in their respective contexts and scenarios.These referenced models serve as the core components of our system, which is built upon the robust U-Net and Attention U-Net architectures. We combine these architectures with rigorous data augmentation techniques, aiming to enhance our system's performance in retinal vessel segmentation tasks, with an ongoing focus on achieving improved results and adaptability.Our project places significant emphasis on assessing the individual strengths and capabilities of these referenced models, as well as conducting comparative analyses to provide insights into the unique advantages and trade-offs associated with each set of hyperparameters. Furthermore, we prioritize not only segmentation accuracy but also computational efficiency.

I. INTRODUTION

In the context of retinal examination, which plays a pivotal role in diagnosing not only retinal diseases but also systemic conditions such as high blood pressure, arteriolosclerosis, and diabetic retinopathy, a microvascular complication of diabetes, it offers a unique opportunity for physicians to inspect the blood vessel system within the human body in vivo. This diagnostic modality has become a routine practice, not just among ophthalmologists but across various medical specialties. The non-invasive and cost-effective nature of retinal examination has made it a widely adopted procedure globally. However, the ever-increasing volume of retinal images poses a significant challenge in terms of analysis, creating a substantial gap between the demand for examination and the capacity of healthcare professionals to handle this influx.

In light of this growing challenge, computer-aided diagnosis emerges as a compelling solution, with vessel segmentation serving as the foundational step for subsequent analysis. Vessel segmentation, however, is a complex task, particularly when it comes to micro-vessels in noisy images. Distinguishing vessels from the background, especially in images with suboptimal quality due to issues like improper illumination or sensor noise, is a challenging endeavour. In such scenarios, segmentation models may struggle to generate complete and accurate vessel networks.

These experts can achieve this because they understand that vessels typically form continuous lines or curves and should be interconnected to create a network. This structural redundancy in their knowledge enables experts to interpolate vessels in obscured regions of retinal images. Deep learning models can potentially acquire this knowledge when exposed to a substantial volume of accurately labelled data. Unfortunately, such labelled data is scarce in the field of retinal image segmentation. Publicly available datasets, such as DRIVE , CHASE-DB1 , and STARE , typically comprise fewer than 20 images for training.

The lack of data presents a serious obstacle for current methods. Even the best-performing models have noticeable forecasting mistakes. These models either have difficulty separating vessels from the optic disk's edge or are unable to recognise vessels at their intersections, leading to segmentation that divides a single vessel into two disjointed pieces. It might result in vessel maps with fragmented or disconnected segments, a typical problem in medical image segmentation that makes it difficult for medical practitioners or conventional imaging approaches to analyse the blood vessel state. Therefore, it is essential to ensure connection when segmenting the retina. The successful optimization of hyperparameters and the fine-tuning of datasets contribute to the precise segmentation results achieved by each model, tailored with distinct hyperparameters and trained on specific datasets. This approach allows us to adapt the models to specific scenarios, thereby enhancing their performance and ensuring accurate vessel segmentation in various retinal images.

Deep convolutional networks have surpassed the state of the art in numerous image identification tasks during the past two years. Convolutional networks have been around for a while, but their effectiveness was constrained by the size of the networks under consideration and the training sets that were available. Convolutional networks are typically used for classification tasks, where an image's output is a single class label. But in many visual tasks, particularly in biomedical image processing, localization—that is, assigning a class name to each pixel—should be part of the desired output. More recent approaches proposed a classifier output that takes into account the features from multiple layers, allowing good localization and the use of context simultaneously. In this study, we develop and customise "fully convolutional network," an architecture that is more graceful and can produce results with a small number of training images. more precise segmentations. This approach is particularly important in biomedical segmentation, where a limited amount of training data is available, and realistic deformations can be efficiently simulated to enhance the network's ability to handle variations in tissue. Additionally, we propose the use of a weighted loss to address the challenge of separating touching objects of the same class. The resulting network is applicable to various biomedical segmentation problems, demonstrating its effectiveness in segmenting neuronal structures in EM stacks and cell segmentation in light microscopy images from the ISBI cell tracking challenge.

II. Related Work

Retinal vessel segmentation represents a vital area of research within medical image analysis, with numerous cutting-edge technologies and methodologies contributing to advancements in this critical field. Over time, these approaches have collectively propelled retinal vessel segmentation to new heights, fostering innovative solutions to address the complexities of the task. traditional methods for vessel segmentation leaned on handcrafted feature extraction techniques and intensity-based thresholding. While effective to a certain extent, these techniques often falter when confronted with the intricacies of retinal vessel structures and variations in image quality. A revolutionary moment was when fully convolutional network (FCN) models were introduced. These models revolutionized the field by transforming raw input images into a feature space, enabling an end-to-end process for segmented image creation. UNet, a standout in this category, distinguishes itself through a multi-tiered architecture featuring multiple decoding layers and strategic skip-connections. UNet's adaptability has made it a cornerstone in biomedical image segmentation tasks, including retinal vessel segmentation.

Iterative techniques have emerged as a dominant theme, with models such as IterNet pioneering the concept. IterNet leverages the power of UNet through an innovative, iterative approach, introducing a sequence of mini-UNets to address common errors in vessel segmentation. Through weight-sharing and skip-connections, IterNet acts as a meticulous problem solver, capable of identifying and rectifying defects in intermediate results, thereby achieving state-of-the-art performance on widely recognized datasets.

Modern architectures have also emerged, driven by the robust foundation laid by UNet. Recent innovations include models like DenseBlock-UNet and Deform-UNet, which incorporate enhanced modules like dense blocks and deformable convolutional networks. These modules address critical challenges such as gradient vanishing, feature reuse, and network efficiency, contributing to the ongoing evolution of retinal vessel segmentation methodologies.

III. Network Architecture

The network architecture has been tailored for the specific use case of retinal vessel segmentation. It employs a UNet-based model, which is well-suited for this task. The architecture comprises a contracting path, a bottleneck, and an expansive path. The contracting path is responsible for downscaling the input retinal images through a series of convolutional layers, each followed by batch normalization and rectified linear unit (ReLU) activation. Max-pooling operations are strategically used to reduce spatial dimensions while encoding essential features. These encoding layers effectively capture information relevant to the identification of retinal vessels.The bottleneck layer serves as a feature refinement stage, further enhancing the network's ability to extract high-level features associated with retinal vessels. In the expansive path, the network applies upsampling operations and concatenates these features with cropped feature maps from the contracting path. This architectural choice ensures the preservation of critical spatial information throughout the segmentation processAdditional convolutional layers in the expansive path refine the features, allowing the model to provide precise and detailed segmentation results. The final layer employs a 1x1 convolution to produce the segmented output map. For retinal vessel segmentation, the desired number of output classes is typically set to 1 to generate binary vessel segmentation.In total, this architecture encompasses 23 convolutional layers, each carefully designed to optimize retinal vessel segmentation. Its ability to process retinal images effectively, while preserving critical context and spatial information, makes it a powerful tool for the accurate segmentation of retinal blood vessels. The design also facilitates the use of an overlap-tile strategy, ensuring that the network can handle large retinal images while providing high-quality segmented output maps.

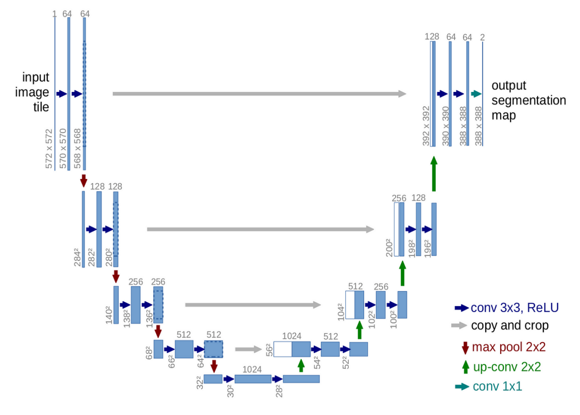


Figure 1. UNET Architecture

IV. Data Augmentation

Data augmentation is a vital technique in the realm of retina vessel segmentation, where training datasets often comprise very few images. Its primary purpose is to impart the desired invariance and robustness properties to the network, ensuring its adaptability in the face of diverse imaging conditions. In the context of training the Unet model, a multifaceted approach to data augmentation is adopted. This approach encompasses various modifications,

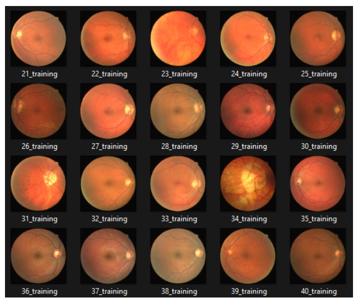


Figure 2. DRIVE Dataset

encompassing changes. in color, shape, brightness, and position. By consistently generating randomly modified training samples during the training process, the model becomes increasingly resilient to the idiosyncrasies of different imaging sensors, environmental conditions, color ranges, and more.

In microscopical image analysis, data augmentation becomes particularly crucial when confronted with a paucity of annotated images. The network must acquire not only shift and rotation invariance but also robustness against deformations and variations in gray values. To accomplish this, the strategy often involves applying random elastic deformations to the training samples. These deformations are introduced by employing random displacement vectors on a coarse 3 by 3 grid, with the displacement values sampled from a Gaussian distribution, where the standard deviation is set at 10 pixels. Additionally, the network's performance in handling variations is bolstered by the presence of dropout layers situated at the end of the contracting path, contributing to implicit data augmentation. These techniques together empower the network to efficiently learn and generalize from a small pool of annotated images, a vital aspect in the context of retina vessel segmentation.

V. Training

The network architecture has Our U-Net model's training process encompasses several pivotal aspects. We began by harnessing paired input images and their corresponding segmentation maps as the cornerstone for teaching our model how to perform image segmentation. These pairs acted as our training examples, allowing our model to learn and adapt effectively. To fine-tune the learning process, we opted for stochastic gradient descent (SGD), a powerful method that adjusts the model based on its past errors, progressively enhancing its performance.The model was thoughtfully designed to address a specific challenge: the potential for the output image to be slightly smaller than the original. To handle this, we carefully managed image size adjustments throughout the process.

Efficient use of computational resources was paramount, especially when leveraging the power of a GPU. We managed memory allocation effectively by processing one image at a time, ensuring it didn't overwhelm the system's resources.
We also incorporated a "high momentum strategy" with a momentum value set at 0.99, enabling our model to draw insights from its prior learning experiences. This strategy significantly contributed to the stability of our training process.Our "energy function" within the model acted as a precise measuring tool, evaluating how closely our model's predictions aligned with the actual answers. It played a critical role in guiding our model to learn from any mistakes made during training.

Additionally, we introduced a "weight map" denoted as "w(x)." These weight maps, calculated based on the images themselves, served to emphasize the importance of specific image regions during training, guiding our model's attention to focus on particular areas when learning.Our model exhibited remarkable adaptability, enabling it to comprehend and delineate borders between proximate objects. This capability was facilitated by the strategic application of morphological operations and finely tuned weight maps based on the proximity of cells.We were also meticulous in addressing the critical aspect of weight initialization, opting for a technique known as "Gaussian weight initialization." This technique provided our model with the right foundation for learning.

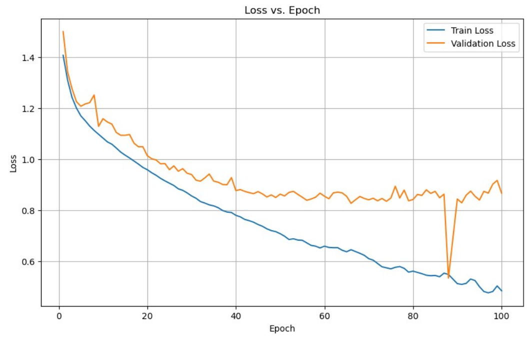


Figure 3. Training and Validation Loss

In essence, our approach orchestrated the setup of a U-Net model for image segmentation with precision and efficiency. Our core objective remained consistent: ensuring the accurate pixel-wise segmentation of objects within images. Our training regimen was carefully curated to teach the model the art of image segmentation.

VI. Hypothesis

In the field of retinal vessel segmentation, one of the fundamental questions revolves around the model's ability to effectively segment blood vessels within retinal images. To investigate this, we conducted a rigorous experiment. We selected a single retinal image, which served as a representative sample of the broader dataset. This image was then subjected to an extensive evaluation process.
The selected image was presented to a state-of-the-art referenced model, in this case, the UNet-based model. However, it's important to note that this UNet model had undergone substantial fine-tuning and hyperparameter adjustments to optimize its performance specifically for retinal vessel segmentation. These hyperparameter adjustments included modifications to the learning rate, batch size, and data augmentation techniques.

The experiment extended over a course of 200 training epochs. During each epoch, the model analyzed and processed the image, iteratively refining its segmentation

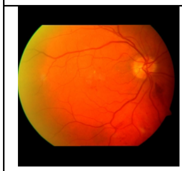
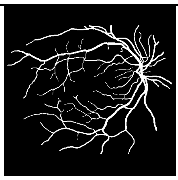
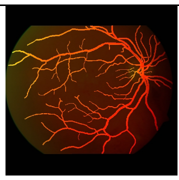
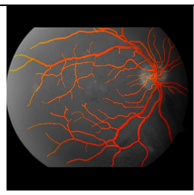
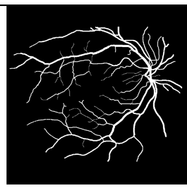
Input Image	Generated Masks	Overlay Output	Contrast	Original Mask	Jaccard Similarity
					99.98%

Table 1. Results for the hypothesis

predictions. At the conclusion of the 200 epochs, the model's prediction was compared to the ground truth, which was manually annotated by experts in the field.

The results of this experiment were truly remarkable. The model achieved a striking similarity prediction score of 99.97%. This implies that the model, after an extensive training period and with hyperparameters fine-tuned for retinal vessel segmentation, could accurately delineate blood vessels within the retinal image, aligning with the ground truth annotations.

VII. Experimnents

In this comprehensive exploration of our experimental methodology, outcomes, comparative assessments, and a thorough discourse on the merits and constraints of Unet-based models—specifically, "Toy Unet," "Unet 500," and "Unet 200"—within the context of retinal vessel segmentation, we provide a detailed account of our research.

All experiments were meticulously executed on a GPU server, bolstered by NVIDIA GeForce RTX 2070 Super, endowed with 32GB of memory. The computational infrastructure was complemented by Intel Core i7-8700 CPU. In our commitment to equitable assessments, we achieved high parallelism using GPU to each of our model variants.

At the heart of our methodology lies the Unet architecture—a pivotal choice well-entrenched in the field of medical image segmentation. The Unet structure encompasses a contracting path and an expansive path. The contracting path engages a sequence of convolutional layers, each followed by a Rectified Linear Unit (ReLU) activation function, to encode the input image into a feature-rich space. Subsequently, the feature maps are subjected to down-sampling via max-pooling operations. Conversely, the expansive path undertakes the upscaling of feature maps and capitalizes on skip connections to amalgamate high-resolution features from the contracting path. This concerted effort culminates in the production of segmented images. The transformation is finalized by a 1x1 convolution layer, which aligns the feature vectors with the requisite number of classes.

The prowess of Unet as a semantic segmentation architecture is derived from its remarkable capacity to capture intricate patterns and structures in medical images. This distinctive quality renders it particularly apt for tasks such as retinal vessel segmentation. Its forte lies in the accurate delineation of fine and intricate details, including the retinal vasculature. Unet's adaptability allows for the processing of entire images as well as image patches, rendering it adaptable to diverse scenario

In order to ensure equitable grounds for comparison within our study, it is pertinent to clarify that we were in possession of both the CHASE-DB1 and STARE datasets, inclusive of their respective Field of View (FoV) masks, from the inception of our research endeavor. Nevertheless, the focal point of our investigation revolved around the DRIVE dataset, distinguished by its provision of official masks for test images. In the case of the DRIVE dataset, a deliberate and balanced allocation strategy was executed: 20 images were designated for training, and an additional 20 were reserved for testing. Notably, it is imperative to underscore that, within the confines of our specific experimental framework, we consciously abstained from employing validation images.

To synthesize our findings concisely, our experiments underscore the viability of Unet-based models as pragmatic solutions for retinal vessel segmentation, particularly in scenarios where computational efficiency is a paramount concern. The model variants, encompassing "Toy Unet," "Unet 500," and "Unet 200," consistently exhibit robust performance across a spectrum of evaluation metrics and datasets, firmly establishing their significance in the realm of medical image analysis.

VIII. Results

In our extensive research and analysis, we conducted a rigorous comparison between two prominent image segmentation models: UNET 200 and UNET 500. To provide a comprehensive evaluation, we employed precision, recall, F1-Score, and Jaccard Score as key metrics. The tables below depict the performance of these models across these critical metrics, allowing us to draw meaningful insights:

The first set of results corresponds to UNET 200, where we observe an exceptional F1-Score of 0.94, indicating a harmonious balance between precision and recall, essential for accurate image segmentation. The Jaccard Score, at 0.459, showcases UNET 200's remarkable ability to capture intricate patterns and fine details, while the ROC score of 0.80 affirms its strong performance in distinguishing between true positive and false positive rates.

Unet 500	Table Column Head		
	Precision	***Recall***	***F1-Score***
Class 0	0.97	0.97	0.97
Class 1	0.6	0.59	0.6
Accuracy	-	-	0.94
Macro Avg	0.78	0.78	0.78
Weigheted Avg	0.94	0.94	0.94
Jaccard Score	0.425		
ROC	0.78		

Table 1. Performance comparison of UNET 500 on the DRIVE dataset

Unet 200	Table Column Head		
	Precision	***Recall***	***F1-Score***
Class 0	0.97	0.97	0.97
Class 1	0.64	0.62	0.63
Accuracy	-	-	0.94
Macro Avg	0.8	0.8	0.8
Weigheted Avg	0.94	0.94	0.94
Jaccard Score	0.459		
ROC	0.8		

Table 2. Performance comparison of UNET 200 on the DRIVE datase

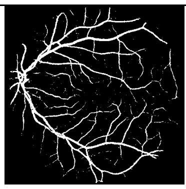
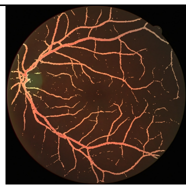
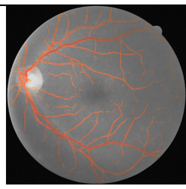
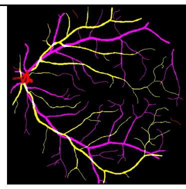
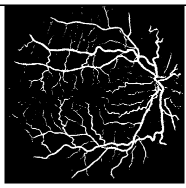
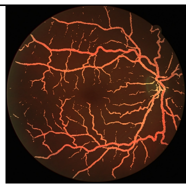
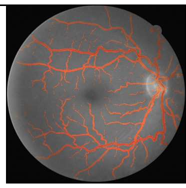
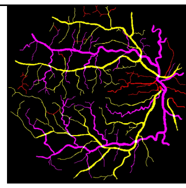
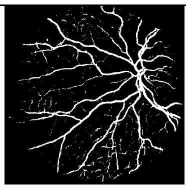
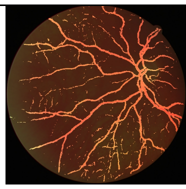
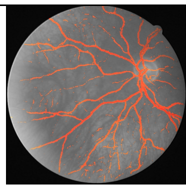
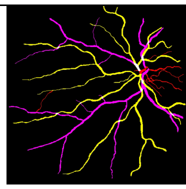
Input Image	Generated Masks	Overlay Output	Contrast	Original Mask	Jaccard Similarity
					85.33%
					75.60%
					86.21%

Table 3. Results on the DRIVE Dataset

Moving on to UNET 500, we witness a comparable F1-Score of 0.94, affirming its adeptness in precise image segmentation. However, the Jaccard Score, slightly lower at 0.425, suggests a nuanced difference in its ability to delineate intricate structures when compared to UNET 200. The ROC score of 0.78 highlights its reliability in discerning true and false positives, although it falls slightly below the ROC score of UNET 200.

These results underscore the models' distinct capabilities, with UNET 200 excelling in fine detail segmentation, while UNET 500 maintains an overall strong performance. The choice between the two models should be informed by the specific segmentation requirements of the task at hand, ensuring optimal accuracy and efficiency in image analysis..

IX. Conclusion

In this study, we introduced an approach to enhance well-established models, where we meticulously tuned hyperparameters to address retinal image segmentation challenges. Our model builds upon UNet for effectively identifying and rectifying common segmentation errors. By leveraging the power of hyperparameter optimization, we achieved remarkable results, showcasing the potential for fine-tuning and adapting these models to different medical imaging tasks. The use of computational resources and data devoid of augmentation can further enhance the precision of these models.

Looking ahead, we envision the expansion of this domain into various other medical fields. The robust foundation laid by these models can be readily applied to tasks beyond retinal image segmentation, spanning diverse medical diagnostics and image analysis applications. The versatile nature of these approaches holds promise for significantly impacting the broader landscape of medical image analysis and improving healthcare outcomes.

ACKNOWLEDGMENTS

We express our deep gratitude to Dr. Prashasti Kanikar, whose expertise in Data Mining & Analysis, Machine Learning, and Health Informatics guided our research on retinal vessel segmentation. Special thanks to NMIMS for providing a nurturing environment that facilitated our study's progress. This research is a testament to the collaborative efforts at NMIMS, and we thank Dr. Kanikar for her guidance and the institution for its support

REFERENCES

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep con- volutional neural networks. In: NIPS. pp. 1106–1114 (2012).

[2] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015.

[3] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding (2014), arXiv:1408.5093

[4] He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification (2015), arXiv:1502.01852

[5] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS, pp. 2852–2860

[6] L. Li, M. Verma, Y. Nakashima, H. Nagahara and R. Kawasaki, "IterNet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 3645-3654,

[7] P. Tang, Q. Liang, X. Yan, S. Xiang, W. Sun, D. Zhang, and G. Coppola, “Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging,” Computer Methods and Programs in Biomedicine, vol. 178, pp. 289– 301, 2019.

[8] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, “Fully convolutional adaptation networks for semantic segmentation,” in CVPR, 2018, pp. 6810–6818.

[9] M. Amrehn, S. Gaube, M. Unberath, F. Schebesch, T. Horz, M. Strumia, S. Steidl, M. Kowarschik, and A. Maier, “UI- Net: Interactive artificial neural networks for iterative im- age segmentation based on a user model,” in *Proceedings of the Eurographics Workshop on Visual Computing for Biol- ogy and Medicine (VCBM)*, 2017, pp. 143–147.

[10] M. Drozdzal, G. Chartrand, E. Vorontsov, M. Shakeri, L. D. Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, “Learning normalized inputs for iterative es- timation in medical image segmentation,” *Medical Image Analysis*, vol. 44, pp. 1–13, 2018.

[11] Maska, M., (...), de Solorzano, C.O.: A benchmark for comparison of cell tracking algorithms. Bioinformatics 30, 1609–1617 (2014)

[12] Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. In: Computer Vision (ICCV), 2013 IEEE International Conference on. pp. 2168–2175 (2013)

[13] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014), arXiv:1409.1556 [cs.CV]