# Towards Long-Range Pixels Connection for Context-Aware Semantic Segmentation

Muhammad Zubair Khan[1,*], Yugyung Lee[2], Muazzam A. Khan[3,*] and Arslan Munir[4]

***Abstract*—Semantic segmentation is one of the challenging tasks in computer vision. Before the advent of deep learning, hand-crafted features were used to semantically extract the region-of-interest (ROI). Deep learning has recently achieved enormous response in semantic image segmentation. The previously developed U-Net inspired architectures operate with continuous stride and pooling operations, leading to spatial data loss. Also, the methods lack establishing long-term pixels connection to preserve context knowledge and reduce spatial loss in prediction. This article developed encoder-decoder architecture with a sequential block embedded in long skip-connections and densely connected convolution blocks. The network non-linearly combines the feature maps across encoder-decoder paths for finding dependency and correlation between image pixels. Additionally, the densely connected convolutional blocks are kept in the final encoding layer to reuse features and prevent redundant data sharing. The method applied batch-normalization to reduce internal covariate shift in data distributions. We have used LUNA, ISIC2018, and DRIVE datasets to reflect three different segmentation problems (lung nodules, skin lesions, vessels) and claim the effectiveness of the proposed architecture. The network is also compared with other techniques designed to highlight similar problems. It is found through empirical evidence that our method shows promising results when compared with other segmentation techniques.**

***Index Terms*—deep learning, semantic segmentation, image analysis, pixels connection, convolution neural network.**

## I. Introduction

In computer vision, deep learning architectures have achieved promising results by outperforming most shallow learning techniques designed for image segmentation and have widely fascinated the research community. Deep architectures are often used in classification problems that yield a single value to an input image. The models require massive data and structural features such as activation and dropout function, optimization algorithms for training and inference. It also plays an integral role in medical diagnosis and treatment by improving system efficiency and providing a better opportunity to rapidly treat a large number of patients. Besides, automatic processing is cost-effective and reduces human intervention in analyzing imaging data. The challenge in image segmentation is the availability of extensive annotated data [1]–[3]. It needs pixel-level labeling instead of image-level annotation. A fully convolutional network was the initial effort made to apply deep learning for image segmentation [4].

The article [5] has extended FCN to U-Net architecture with better segmentation results on scarce input data. It mainly involved encoding and decoding operations. The dimensions of input samples are reduced in the encoding path and obtained back with the same size in the decoding path. Multiple U-Net inspired models are discussed in [6]. The most prominent transformation is about skip-connections. The extracted features are initially fed to processing before concatenation operation in some variants, as mentioned in [7], [8]. The major issue is the individual processing step conducted for each set of feature maps which later concatenates together. This article applied a sequence model in long-term skip-connection to reuse features maps with dense convolutions. The feature maps obtained from the encoding layer have higher dimensions, while the maps extracted from the transpose convolutional layer contain essential semantic details. Therefore, we have combined feature maps with non-linear functions to obtain precise segmentation output.

Our paper has developed a U-Net inspired encoder-decoder architecture and added a bi-directional LSTM module within the skip-connections to join feature maps together. Our approach has used the concept of dense connections to improve information flow across the network. The convolutional blocks in the final layer of the encoder path have channel-wise concatenation such that the features learned in each block are advanced to the subsequent block for using collective knowledge acquired by each previous layer and prevent a network from learning redundant features. Besides, we applied batch-normalization to avoid the internal covariate shift and accelerate the network convergence rate. Our method is evaluated on three distinct datasets. The results revealed that our network demonstrates a promising response for multiple segmentation applications.

### A. Objectives and Contributions

The main objectives of our research were to avoid the appearance of redundant features in successive convolutions, enhance the prediction response by bringing a two-way temporal stance in decision-making and reduce the internal covariate shift for model convergence. The main contributions include:

1) In our method, we established a dependency between image intensities for preserving context information using bi-directional LSTM.

[1]Muhammad Zubair Khan, School of Computing and Engineering, University of Missouri-Kansas City, Kansas City, USA (mkzb3@mail.umkc.edu)

[2]Yugyung Lee, School of Computing and Engineering, University of Missouri-Kansas City, Kansas City, USA (leeyu@umkc.edu)

[3]Muazzam A. Khan, Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan (muazzam.khattak@qau.edu.pk)

[4]Arslan Munir, Department of Computer Science, Kansas State University, Manhattan, USA (amunir@ksu.edu)

*Corresponding authors: [1]Muhammad Zubair Khan, [3]Muazzam A. Khan

2) We reused features and prevented redundant data sharing using dense convolutional blocks.
3) Our method has non-linearly combined the feature maps across the encoder-decoder path to find the correlation between image pixels.
4) We applied batch-normalization to improve the information flow for improving model convergence rate.

## II. Related Work

In the past few years, scientists have witnessed improved results in image segmentation using deep learning techniques. In [9], Kleesiek et al. proposed 3-dimensional CNN for image segmentation. The network used multiple channels of FLAIR contrasts, non-enhanced and contrast-enhanced T1w, and T2w to process input data. [10] developed multi-level deep CNN for extracting pancreas from abdominal CT scans using a probabilistic bottom-up approach. It is observed that CNN causes spatial loss when convolutional features are provided to the fully connected layers during the segmentation task. To resolve this issue, Long et al. [4] came up with the concept of a fully connected convolutional neural network. The network substitutes fully connected layers with convolutional and deconvolutional layers to preserve the image dimensions and obtain the feature maps similar to the raw input sample. In [11], the authors performed end-to-end FCN training for extracting anatomical structures from 3D scans. They have conducted voxel-wise multi-class classification to assign a label for each voxel.

[12] used short and long skip-connections to build deep FCN with improved response. Ronneberger et al. [5] has developed U-Net model for medical image segmentation. It has used image augmentation to operate with limited data efficiently. The main contributions include data augmentation, separation of joint objects, and overlapping-tile approach. The V-Net is proposed in [13] by Milletari et al. for volumetric image analysis. The network is evaluated on 3D MRI volumes for finding the prostate condition. It also introduced a unique objective function, applied histogram matching, and used random non-linear transformations for image augmentation. The authors in [14] proposed a dense volumetric 3D U-Net. The analysis path in a network contains 3D convolutions followed by ReLU activation and max-pool layers, whereas the synthesis path contains 3D up-sampling with ReLU activation function. An end-to-end feed-forward recurrent neural network is proposed in [15]. It preserves long-term dependency between intensities in a scene while limiting the model capacity. Zhao et al. in [16] have proposed PSPNet with global pyramid pooling and an efficient optimization approach for embedding critical scenery context features. Two segmentation models, recurrent U-Net and recurrent residual U-Net are built by Alom et al. [6]. The encoder-decoder structure with residual and recurrent blocks assured the improved feature representation of resultant maps.

Wang et al. [17] Proposed an end-to-end hard-attention network with multiple decoder networks to dynamically identify hard and easy regions in an image to segment. The deeplab architecture [18] replaced fully connected layers with convolutions and increased feature dimension through atrous convolutional layers. [19] has combined a fully convolutional network with RNN to incorporate temporal information and spatial details. Laibacher et al. [20] developed M2U-Net for medical image segmentation. It has added pre-trained components in the encoder part and contractive bottleneck blocks in the decoder part. Li et al. [21] have developed an IterNet model that contains multiple iterations of a mini-UNet and adopted skip-connections with weight-sharing features. Hervella et al. [22] proposed a self-supervised pre-training method to reduce the need for extensive data and provided a multi-modal solution for medical image segmentation. Akbar et al. [23] have developed a bi-modular clinical decision system. The first module used a support vector machine with an RBF kernel for classification, and the second module performed analysis to detect the symptoms of papilledema.

In [24], the authors defined a bi-directional attention block for capturing the correlation between image intensities. It helps incorporate multi-scale information and adaptively learn a rich set of features. A LadderNet architecture proposed in [25] contains multiple U-Nets and residual blocks with a weight-sharing feature. The multi-path architecture supports data throughput, and the use of residual block limits the number of parameters. A context-encoder network developed for 2D image segmentation was proposed to conserve spatial information and avoid data loss caused by max-pool operation [26]. The network encompassed a modular approach and performed dense atrous convolutions and residual multi-kernel pooling. [27] have applied a multi-scale architecture with a custom-designed feature extractor for data-driven analysis. The authors in [28] have proposed a multi-modal transfer learning approach for semantic image segmentation by defining a self-supervised method that can effectively handle data scarcity. [29] has used the spatial-attention module to conduct adaptive feature refinement. The process discussed in [30] highlighted a graphical vascular structure to show a strong tie between neighboring pixels. The research modified the neural graphs for effective local and global representations.

## III. Proposed Methodology

The proposed model shown in the Fig. 1 has used the strength of bi-lstm, long skip-connections, and dense convolutions. The encoding path consists of four convolutional blocks. Each block includes two convolutional layers followed by max-pool operation with a $2\times2$ kernel. The count of filters increases twice at each subsequent block. The encoding path progressively extracts spatial representations layer to layer, and eventually, the last layer generates high dimensional representation with semantic information. We have applied densely connected convolutions as the final step in the encoding path to launch an idea of collective knowledge by reusing the feature maps through the network. We introduced a sequence of N-blocks with two successive convolutional layers. The feature maps obtained from all the previous layers are combined with maps learned in the current layer and then delivered to the subsequent layer. It avoids the appearance of redundant

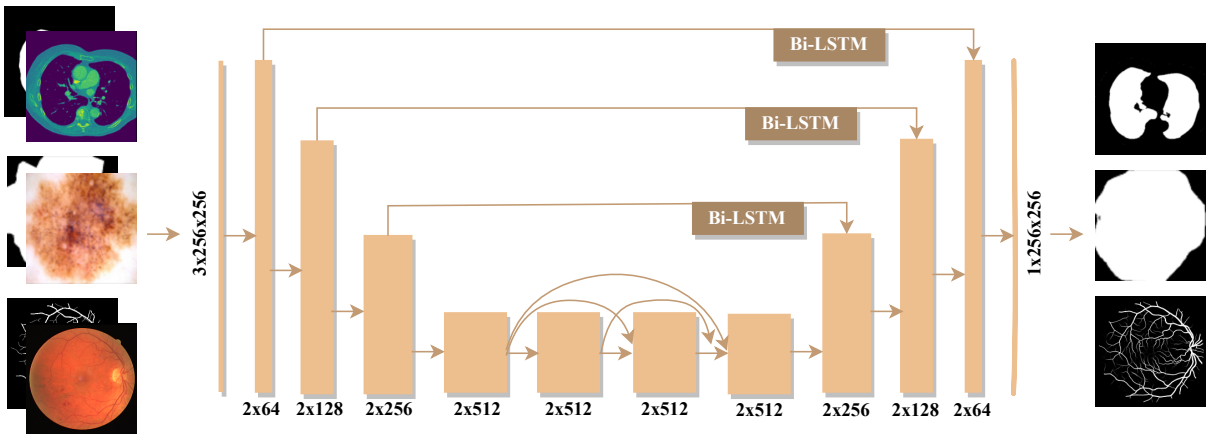


Fig. 1. The figure depicts the proposed architecture with bi-directional LSTM and dense blocks. Number of filters are mentioned under each highlighted bar.

features in successive convolutions and mitigate the risk of vanishing gradient by benefiting from all previous features.

The decoding path performs a de-convolutional function over the previous layer output. In the conventional U-Net approach, the model crops the feature maps in the encoding path, copy them to the decoder, and concatenates with the output of the de-convolutional function. However, we have applied bi-directional lstm with long-term skip connections to process these maps in our approach. The number of filters increases as we move through the decoding path to get the final map equal in size to the original input image. After up-sampling, we performed batch-normalization to reduce the internal covariate shift. Each hidden layer needs to adaptively learn a new distribution during the training step. This variable activation distribution results in a slow convergence rate. Therefore, we have used a batch-normalization function to accelerate the training speed and bring stability to our network with the regularization effect. The output of batch-normalization is provided to the bi-directional convolution LSTM layer.

The cell-gate structure helps exploit the convolution operation in input-to-state and state-to-state transitions. The method applies convolutional LSTM for each forward and backward path and creates a long-range pixels connection to preserve dependency between image intensities in both directions. It enhances the prediction response by bringing a two-way temporal stance in decision-making. We have used the ReLU activation function for all the hidden layers and a sigmoid function for the output layer. The dropout function is used to avoid overfitting. The error is measured with binary cross-entropy, and the model is tuned with adam optimizer, keeping the initial learning rate of $\lambda = 10^{-4}$.

## IV. Experimentation Results and Discussion

The empirical evaluation shows that our method outperformed all the compared techniques designed against LUNA, ISIC2018, and DRIVE datasets. We considered multiple evaluation metrics to assess our method, including sensitivity (SE), specificity (SP), F-Measure, Jaccard score (JS), accuracy (AC), and area under the curve (AUC). We used the callback function and monitored a validation loss for ten consecutive epochs to halt the training process on plateau response. The DRIVE dataset provides 40 fundus images with a size of 565x584 [39]. These images are divided into train/test splits, such as each split containing 20 images. The STARE dataset contains 20 fundus images in PPM format with 700x605 dimension and captured at 35° angle [40]. The number of images is not sufficient to train our network. Therefore, we took help from previous literature to perform patch-based training. Each image is divided into a number of patches. One thousand patches are randomly selected from each image for model training. We have applied the batch-size of 32x32.

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH OTHER TECHNIQUES DESIGNED FOR VASCULAR TREE EXTRACTION.

| **Dataset** | **DRIVE** | | **STARE** | |
|---|---|---|---|---|
| **Method** | **ACC** | **AUC** | **ACC** | **AUC** |
| Azzopardi [31] | 0.9442 | 0.9614 | 0.9497 | 0.9563 |
| Zhuo [32] | – | 0.9754 | – | – |
| Alom [33] | – | 0.9784 | – | 0.9815 |
| Shin [30] | – | 0.9801 | – | – |
| Laibacher [20] | – | 0.9714 | – | – |
| Ronneberger [5] | 0.9531 | 0.9755 | 0.9690 | 0.9898 |
| Orlando [34] | – | 0.9507 | – | – |
| Zhuang [25] | 0.9561 | 0.9793 | – | – |
| Yan [35] | 0.9542 | 0.9752 | 0.9612 | 0.9901 |
| Zhang [36] | 0.9476 | 0.9636 | 0.9554 | 0.9748 |
| Liu [37] | – | 0.9798 | – | – |
| Mou [38] | – | 0.9796 | – | – |
| Proposed Architecture | 0.9708 | 0.9891 | 0.9724 | 0.9915 |

The ISIC18 dataset is taken from the lesion segmentation challenge [41]. It contains 2594 color images with 700x900 dimensions. Our experiment used 1815 samples for training, 259 samples kept for validation, and the remaining 520 images were used for testing. We applied the essential pre-processing by reducing the image size to 256x256. The data comes in

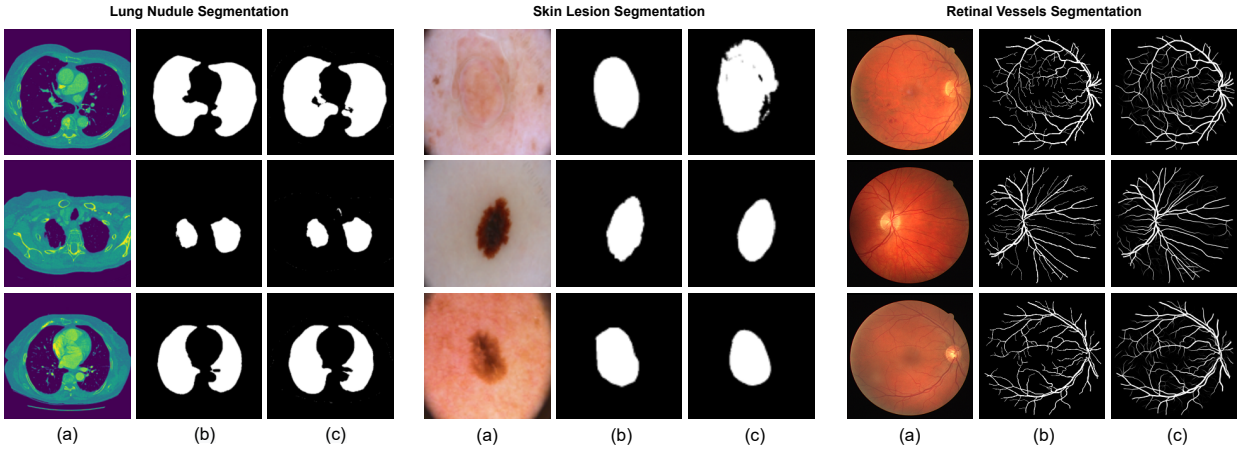


Fig. 2. The figure demonstrates the prediction results on LUNA, ISIC18 and DRIVE. Here, a) color input image, b) groundtruth image, c) output prediction.

RGB images with corresponding ground-truth images. The lung images are collected from the LUNA competition [42]. The data contains CT scans of 512x512 pixel resolution. We have used $70\%$ of data for training and validation, whereas $30\%$ is used for testing. Since the scans also include unwanted regions of interest like bones, we extracted new masks with standard pre-processing steps for model training. We have obtained promising results shown in Fig. 2. It is found that the network with dense units and bi-directional LSTM has outperformed other alternatives mentioned in Table I and Table II. It is observed during experimentation that our model started converging rapidly after $25$ epochs on DRIVE and STARE datasets.

Initially, our model was overly fitted on fundus data. Later, the successive attempts with distinct dropout values helped us overcome this issue. We also observed a fast convergence rate for ISIC18 and LUNA datasets, similar to retinal datasets. The segmentation results were not impressive initially but gradually advanced after $50$ epochs. In Table I and Table II, it is visible that the bi-directional LSTM and dense connectivity have improved the results for all the modalities. We combined encoding-decoding features to capture local spatial and fine semantic information. The batch-normalization helped control the mean and variance distribution to reduce generalization cost. Moreover, the dense connections enabled the network to share feature maps between convolution blocks. It produces diverse features and expands the efficacy of deeper architectures. The dense block supported backward supervision to propagate the error signals directly to the earlier network layers and reduce the chances of vanishing gradient.

## V. Conclusion and Future Works

In conclusion, we have proposed a deep neural architecture to create a long-range pixels connection for context-aware semantic segmentation. We found that inserting dense convolutions and embedding bi-directional convolutional LSTM in long-term skip connections reduces the redundant feature representations and enables the network to capture essential discriminative features. It is also found that the batch-normalization function helped in rapid convergence and regularized our training process. The results obtained on three benchmark dataset has shown that our method has a better response than other compared techniques. In the future, the method would be evaluated against other modalities. Also, the impact of generative adversarial networks would be analyzed by making architectural modifications.

TABLE II
COMPARISON OF THE PROPOSED METHOD WITH OTHER TECHNIQUES DESIGNED FOR LUNG AND LESION SEGMENTATION.

| **Dataset** | **LUNA** | | **ISIC18** | |
|---|---|---|---|---|
| **Method** | **ACC** | **F-Measure** | **ACC** | **F-Measure** |
| Ronneberger [5] | 0.9872 | 0.9658 | 0.8906 | 0.6472 |
| RU-Net [33] | 0.9836 | 0.9638 | 0.9367 | 0.8799 |
| Oktay [7] | – | – | 0.8976 | 0.6658 |
| Zhang [43] | 0.9868 | – | – | – |
| Gu [26] | 0.9900 | – | – | – |
| R2U-Net [33] | 0.9918 | 0.9832 | 0.9372 | 0.8823 |
| Proposed Architecture | 0.9943 | 0.9853 | 0.9462 | 0.8868 |

## References


[1] M. Z. Khan and Y. Lee, "Localization of ocular vessels with context sensitive semantic segmentation," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2021, pp. 1–5.

[2] M. Z. Khan, M. K. Gajendran, Y. Lee, and M. A. Khan, "Deep neural architectures for medical image semantic segmentation," *IEEE Access*, vol. 9, pp. 83 002–83 024, 2021.

[3] M. Z. Khan and Y. Lee, "Dynamic inductive transfer learning with decision support feedback to optimize retina analysis," in *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*. IEEE, 2021, pp. 93–100.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[6] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.

[7] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[8] M. Z. Khan, Y. Lee, A. Munir, and M. A. Khan, "Multi-feature extraction with ensemble network for tracing chronic retinal disorders," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 926–932.

[9] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, "Deep mri brain extraction: A 3d convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, 2016.

[10] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564.

[11] X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Three-dimensional ct image segmentation by combining 2d fully convolutional network with 3d majority voting," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 111–120.

[12] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep learning and data labeling for medical applications*. Springer, 2016, pp. 179–187.

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[15] P. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *International conference on machine learning*. PMLR, 2014, pp. 82–90.

[16] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[17] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3384–3396, 2020.

[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[19] W. Bai, H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews, and D. Rueckert, "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 586–594.

[20] T. Laibacher, T. Weyde, and S. Jalali, "M2u-net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[21] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3656–3665.

[22] Á. S. Hervella, L. Ramos, J. Rouco, J. Novo, and M. Ortega, "Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 961–965.

[23] S. Akbar, M. U. Akram, M. Sharif, A. Tariq, and U. ullah Yasin, "Arteriovenous ratio and papilledema based hybrid decision support system for detection and grading of hypertensive retinopathy," *Computer methods and programs in biomedicine*, vol. 154, pp. 123–141, 2018.

[24] K. Li, X. Qi, Y. Luo, Z. Yao, X. Zhou, and M. Sun, "Accurate retinal vessel segmentation in color fundus images via fully attention-based networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2071–2081, 2020.

[25] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," *arXiv preprint arXiv:1810.07810*, 2018.

[26] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[27] R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, "A data-driven approach to referable diabetic retinopathy detection," *Artificial intelligence in medicine*, vol. 96, pp. 93–106, 2019.

[28] S. Hervella, L. Ramos, J. Rouco, J. Novo, and M. Ortega, "Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 961–965.

[29] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," *arXiv preprint arXiv:2004.03696*, 2020.

[30] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Medical image analysis*, vol. 58, p. 101556, 2019.

[31] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical image analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[32] Z. Zhuo, J. Huang, K. Lu, D. Pan, and S. Feng, "A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index," *Computer methods and programs in biomedicine*, vol. 196, p. 105508, 2020.

[33] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual u-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.

[34] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 16–27, 2016.

[35] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1427–1436, 2018.

[36] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2631–2644, 2016.

[37] N. Liu, L. Liu, and J. Wang, "Local adaptive u-net for medical image segmentation," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 670–674.

[38] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1392–1403, 2019.

[39] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE transactions on medical imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[40] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[41] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.

[42] "Lung Dataset, https://www.kaggle.com/kmader/finding-lungs-in-ct-data, Accessed: 2022-01-20."

[43] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "Et-net: A generic edge-attention guidance network for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 442–450.