Grenze International Journal of Engineering and Technology, January Issue

Development of Oral Cancer Segmentation Framework with Histopathological Images and Dilated Trans-Mobile-Unet++

Ann Mathew[1] and Dr. A Maria Jossy[2]
[1-2]Department of ECE, SRM Institute of Science and Technology, Kattankulathur-603203, TamilNadu, India
Email: am1230@srmist.edu.in

***Abstract*—In order to detect Oral Squamous Cell Carcinoma (OSCC), a pathologist must segment the tumor regions on H&E-stained slides. Since labeling histological images is a highly skilled, intricate, and time-consuming operation, the availability of labeled training data affects histopathological image segmentation. Accurate segmentation of microvessels and nerves is needed for an effective oral cancer diagnosis process. Thus, an efficient oral cancer segmentation framework is designed with deep learning techniques to tackle the limitations of the existing mechanism. Several phases presented in the developed oral cancer segmentation framework are (a) Image Collection, (b) Image Denoising, and (c) Image Segmentation. At first, histopathological images needed for the experiment are accumulated from the benchmark sites. Next, the attained images are provided for the image denoising phase. Here, median filtering and Contrast Limited Adaptive Histogram Equalization (CLAHE) mechanisms are employed to perform denoising in the collected images. Next, the denoised images are given to the image segmentation region. In this stage, effective segmentation is performed using the Dilated Trans-Mobile-Unet++ (DTM-Unet++) model. Later, different experimental analyses were carried out to validate the effectualness rate of the developed system.**

***Index Terms*— Oral Cancer Segmentation; Histopathological Images; Image Denoising; Dilated Trans-Mobile-Unet.**

I. INTRODUCTION

One of the most prevalent types of cancer is oral cancer. High fatality rates and increased morbidity can result from delayed discovery of oral cancer. Smoking and drinking are the two most common causes of oral cancer. Early detection and categorization are possible by deep learning algorithms and advanced technologies [1]. Early cancer detection through the analysis of X-ray, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) scans are available to start treatments. These methods are useful to recognize the normal region from the tumor-prone regions and facilitate the anatomical examination of oral cancer [2]. A biopsy combined with histopathology is thought to be the most reliable method in recent days for diagnosing Oral Squamous Cell Carcinoma (OSCC). But, the biopsy specimen must be fixed, embedded, and stained before the results are ready, which takes several days. Although it can identify cellular and molecular alterations with the use of specialized procedures, its interpretation is upon the pathologist [3].
The pathologist usually evaluates the changes in the distribution of the cell throughout the tissue being

Grenze ID: 01.GIJET.11.1.8_1

examined, as well as any aberrations in the cell architecture. These assessments are completely qualitative and frequently result in significant variations [4] [5]. Potentially Malignant Lesions (PML), including leukoplakia and erythroplakia may occur prior to OSCCs. Erythroplakia is a velvety plaque and it is very rare. The most cancerous oral lesion is leukoplakia [6] [7]. Neural networks are employed in most of the segmentation process. But, it needs more training and extensive supervision for an effective texture-based segmentation. Moreover, the processing time of some existing methods was high [8] [9] [10]. Therefore, a deep learning technique for oral cancer segmentation is proposed.

Some of the salient contributions of the developed oral cancer segmentation framework using deep learning are given below:

- ❖ To implement oral cancer segmentation method to localize the affected region of cancer and further stop the progression of cancer at an early stage by taking appropriate treatments.
- ❖ To denoise the histopathological images for improving the segmentation process by eliminating the noise using median filtering and CLAHE techniques.
- ❖ To develop a DTM-Unet++ network to perform oral cancer segmentation in histopathological images. It is a combination of mobile UNet++, transformer, and dilation techniques. It helps to attain accurate segmentation results by learning specific features from the denoised images.
- ❖ To validate the capability of the designed model, different deep learning techniques are considered. The final results of the suggested model and the considered techniques are compared with one another.

The organization of the recommended deep learning technique to segment oral cancer is given in this part. Part II provides the literature survey of the existing works suggested for oral tumor segmentation. The challenges and the benefits of those models are given in the form of a table in this part. Part III presents the image collection process and describes the denoising techniques in detail. The oral cancer segmentation using the developed deep learning technique is explained and the diagrammatic representation is given in Part IV. The numerical outcomes and the conclusion are provided in Part V and Part VI.

II. LITERATURE SURVEY

A. Problem statement

Oral cancer is the deadliest cancer type, which badly affects individuals. Classical oral cancer detection techniques suffer from precision issues. So, automatic oral cancer detection techniques are designed to identify oral cancer cells among individuals. EM algorithm [11] effectively enhances the accuracy and gives a better visibility rate in segmentation. Yet, it requires minimizing the system processing time. Unet [12] collects the fine and coarse characteristics efficiently and also it improves the segmentation performance rate. However, it needs more memory space to store the information and also it implementation procedure is complicated. Xception residual block [13] effectively encodes the contextual information and also improves the multi-scale features. But, it slows down the entire system in the training phase and also it needs to enhance the performance rate. RCAug [14] implementation process is simple and also effectively speeds up the system performance rate. Yet, its implementation process is expensive and also it requires enormous labeled data for observation. CNN [15] effectively resolves the overfitting issues and also it effectively minimizes the processing time. However, it needs to overcome the interpretability issues. Thus, it is required to design an efficient oral cancer segmentation model by considering the above with deep learning techniques.

III. INTRODUCTION TO PROPOSED ORAL CANCER SEGMENTATION FRAMEWORK: REPRESENTATION OF HISTOPATHOLOGICAL IMAGES AND DENOISING TECHNIQUES

A. Proposed Oral Cancer Segmentation Framework: Architectural View

The pictorial view of the developed oral cancer segmentation framework is shown in Fig. 1.

An oral cancer segmentation framework via deep learning is designed to locate the tumor regions for an earlier diagnosis process. At the beginning, the histopathological images are collected from online sites. Then, the image denoising process is done with the help of median filtering and CLAHE approaches to remove the noise in images. The denoised images are passed to the segmentation stage. The segmentation procedure is done using the proposed DTM-Unet++ technique. The denoising process is useful to perform segmentation more easily and accurately in a short amount of time. Also, the developed deep learning method is more efficient in segmenting oral cancer without any manual help. The capacity of the suggested model in segmenting oral cancer is validated by comparing the simulation outcomes with various heuristic algorithms and deep learning mechanisms.

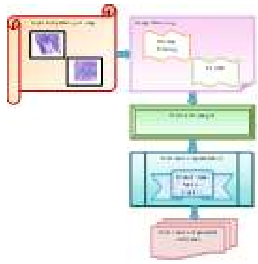

Fig 1. Pictorial view of the developed oral cancer segmentation framework

B. Histopathological Image Collection

The (**VGG16-histopathology-oral cancer detection**) dataset is utilized in this work. This database is accessed using the link https://www.kaggle.com/code/vuppalaadithyasairam/vgg16-histopathology-oral-cancer-detection/input on 2024-03-19. It consists of two set of directories named normal and OSCC. Totally, 95 files are in OSCC folder and 31 files are in normal folder. The total size of the database is 3.19GB and all the files are in jpg format.

The sample histopathological images are specified as OC_S^{HP}. The total count of accumulated images is stated as s

C. Image Denoising Techniques

The input images OC_S^{HP} are given to the image denoising phase. The noise in input histopathological images is eliminated for precise segmentation of oral cancer using median filtering and CLAHE techniques.

Median filtering [16]: Median filtering is the process of sorting gray values of the image pixels in the sliding window. The median is computed via Eq. (1).

$$D = \begin{cases} y_{(f+1)/2} & f \text{ is odd} \\ \frac{1}{2}\left[y_{f/2} + y_{(f+2)+1}\right] & f \text{ is even} \end{cases} \tag{1}$$

In the above equation, the median is denoted as D. The adjacent pixel values and the gray values of noise points are different from one another. The noise point is the median after sorting the gray values. The noise point is eliminated by replacing the median value with the adjacent pixel's value. The median filtering approach is very useful for smoothing noise. This technique protects the image's edge information by reducing salt-and-pepper noise.

CLAHE [17]: In order to enhance the contrast of the images, the distribution of the image histogram is distributed in the CLAHE method. Enlarging the gray values of the pixels is the actual process of a histogram. The d^{th} gray level's probability is calculated using Eq. (2).

$$B_x(x_d) = \frac{m_d}{m} \tag{2}$$

Here, the gray level probability is represented as $B_x(x)$. The d^{th} gray scale is denoted as x_d. Below Eq. (3) provides the computation for Cumulative Distribution Function (CDF).

$$h = F(x_d) = \sum_{b=0}^{d} B_x(x_d) = \sum_{b=0}^{d} \frac{m_d}{m} \tag{3}$$

The interpolation of the gray level mapping is done via Eq. (4).

$$\begin{aligned} l' &= (1-\alpha)(1-\beta)hB_1(l) + \beta hB_2(l) + \\ &\alpha((1-\beta)hB_3(l) + \beta hB_4(l)) \end{aligned} \tag{4}$$

Here, the gray level is indicated by the term h. The normalized distance is mentioned as α and β, respectively. The denoised images are denoted as OC_S^{PRE}.

IV. DEVELOPED HISTOPATHOLOGICAL IMAGE SEGMENTATION MECHANISM USING ADVANCED DEEP STRUCTURE WITH DILATION NETWORK

A. Implemented DTM-Unet++ for Segmentation

In order to perform oral cancer segmentation in histopathological images, a DTM-Unet++ approach is developed. This technique consists of a Mobile Unet++, a transformer, and a dilation block. This network receives the denoised images OC_S^{PRE} as input. As the noise in the input histopathological images is already removed, the segmentation process becomes more effective and accurate. The MobileNet has the ability to generate new relevant features that are useful for the segmentation process. Similarly, the UNet++ architecture has the ability to capture fine and coarse feature information that is beneficial for precise segmentation. The long-term dependencies are captured by the transformer block that helps to process the histopathological images without taking much time. Then, the dilation block used in the suggested DTM-Unet++ effectively manages the correlation between different regions. Therefore, the proposed DTM-Unet++ is highly capable of segmenting oral cancer in histopathological images. The structural view of the designed DTM-Unet++ for oral cancer segmentation is displayed in Fig. 2.

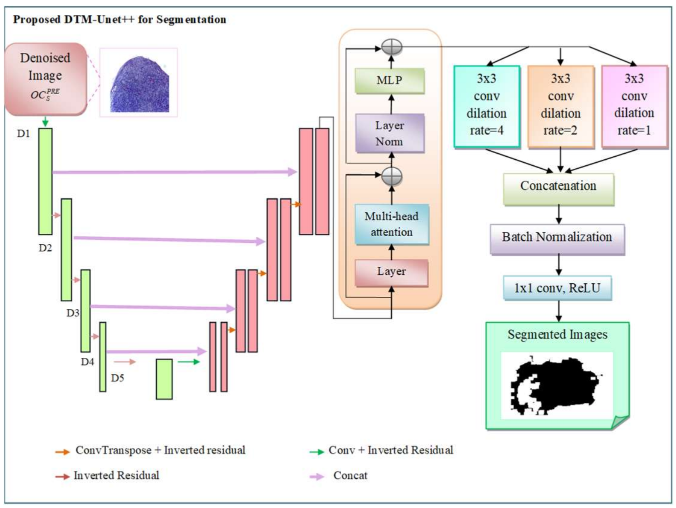

Fig 2. Structural view of the designed DTM-Unet++ for oral cancer segmentation

B. Overview of Mobile Unet++

The denoised images are given to the Mobile UNet++ [18] network. The input size images are decreased at each step. The shortcut connections in the up-sampling part of the UNet++ network are utilized to capture low-level and high-level features. MobileNet is present in the down-sampling part of the UNet++ architecture. Five deconvolutional layers are presented in the up-sampling portion of UNet++. Output received from the previous convolutional layer is given to the next layer.

UNet++ network processes the input as given in Eq. (5).

$$m^{p,q} = \begin{cases} C\left(m^{p-1,q}\right), q = 0 \\ C\left(\left[m^{p,q}\right], F\left(f^{p+1,q-1}\right)\right), q > 0 \end{cases} \tag{5}$$

Here, the term $F(.)$ and $C(.)$ indicates the activation function and convolution operation. The feature maps are indicated as $m^{p,q}$. The MobileNet approach filters the input without generating new features using the depthwise convolution process. Then, this network creates the features using pointwise convolution. After, both of these convolution process, a Rectified Linear Unit (ReLU) and Batch Normalization (BN) are used.

C. Description of Transformer Block

Transformer network [19] is an encoder-decoder framework. It solves the long dependencies problem. From the input, the contextual information is retrieved by the transformer block. Self-attention blocks are presented in the decoder. The attention layer is described in Eq. (6).

$$A(Y, L, N) = SF\left(\frac{YL^K}{\sqrt{q_l}}\right)N \quad (6)$$

Here, the keys and their dimensions are indicated as q_l and L, correspondingly. The values and queries are specified as N and Y, respectively. The terms SF and A represent the softmax and attention. The weights are used to determine the output and it is estimated by the softmax function. In order to obtain stable gradients, the dimensions of the keys are used to divide the key and query values.

D. Dilation Block

The dilation mechanism [20] has the ability to learn distant spatial dependency and it solves the correlation issues. The dilated convolution block is arithmetically expressed as in below Eq. (7).

$$J_K^{b+1} = n\left(Y_{K\times hb}^b J_K^b + c_K^b\right) \quad (7)$$

In the above equation, the activation function is specified as $n(\ )$. The bias vector is mentioned as c_K^b. The dilation factor is represented as hb. The weight matrix, output, and input of the dilation block is referred as Y_K^b, J_K^{b+1} and J_K^b, respectively. With the help of various dilation factors, the same filters are applied in different ranges by the hb^{th} dilated convolution. In addition, the correlation between two regions is captured by this dilated convolution hb.

V. Numerical Findings and Discussions

A. Experimental Configuration

The implementation process of the developed deep learning-assisted oral cancer segmentation method was done in Python. The efficacy of the established deep learning framework was verified by contrasting the outcomes with other techniques. The techniques considered for comparing the results with the proposed model were Unet [21], Unet3+ [22], Res-Unet [23] and Trans-Unet [24].

B. Resultant image outcomes

The resultant segmented images are visualized in the below Fig. 3

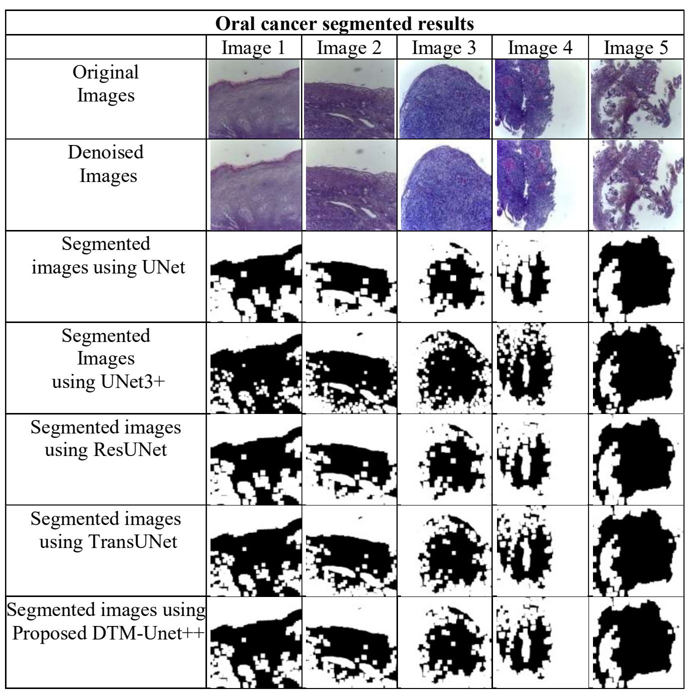

Fig 3. Resultant segmented images of oral cancer in histopathological images

C. Performance measures

The computations for the performance metrics utilized to validate the presented oral cancer segmentation are given in the following link (https://en.wikipedia.org/wiki/Sensitivity_and_specificity).

D. Performance analysis of the proposed method

Fig. 4 shows the oral cancer segmentation results of the proposed model.

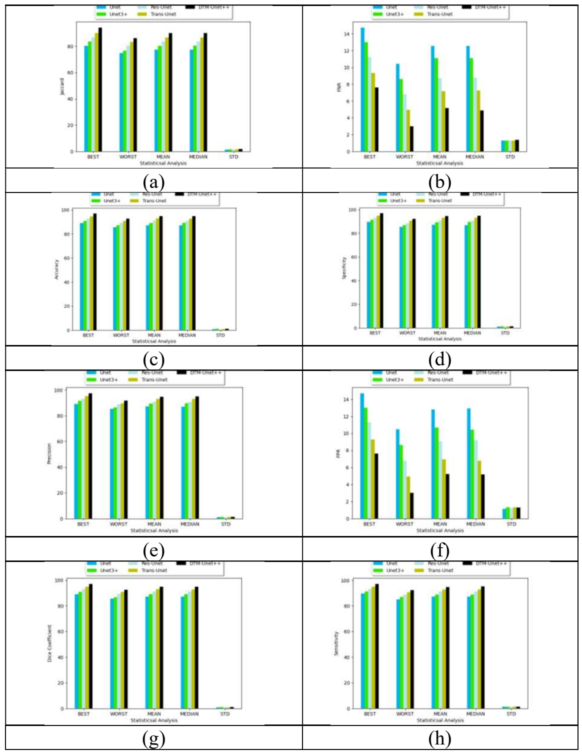

Fig 4. Performance evaluation of the implemented deep learning-based oral cancer segmentation with respect to a) Jaccard, b) FNR, c) Accuracy, d) Specificity, e) Precision, f) FPR, g) Dice coefficient and h) Sensitivity

VI. Conclusion

An oral cancer segmentation framework via deep learning was designed to locate the tumor regions for an earlier diagnosis process. Initially, the histopathological images were collected from standard sites. Then, the noise in the input histopathological images was eliminated by techniques including median filtering and CLAHE. The denoised images were passed to the segmentation stage. The segmentation procedure was done using the proposed DTM-Unet++. This technique consisted of a mobile UNet++, transformer, and dilation techniques. The overall efficacy of the executed model was evaluated with other deep learning mechanisms. In Fig. 6, the accuracy of the oral cancer segmentation using DTM-Unet++ was greater than UNet, UNet3+, Res-UNet, and Trans-UNet with 10.86%, 8.69%, 6.52% and 4.24% at mean analysis. Thus, the performance of the oral cancer segmentation via the suggested DTM-Unet++ is better than other conventional techniques.

References

[1] N. Wang, Y. Liu, and H. Li, "An Efficient and Fast, Noninvasive, Auto-Fluorescence Detection Method for Early-Stage Oral Cancer," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-11, 2022.

[2] D. Verma, S. K. Yadav, A. Kalkal, R. Pradhan and G. Packirisamy, "An Ultrasensitive Electrochemical Immunosensor Comprising Green Synthesized α-Fe2O3NPs_rGO Nanocomposite for Determination of Oral Cancer," *IEEE Sensors Letters*, vol. 7, no. 12, pp. 1-4, Dec. 2023.

[3] HadjouniMyriam; Abdelaziz A. Abdelhamid; El-Sayed M. El-Kenawy; Abdelhameed Ibrahim; MarwaMetwallyEid; Mona M. Jamjoom; Doaa Sami Khafaga, "Advanced Meta-Heuristic Algorithm Based on Particle Swarm and Al-Biruni Earth Radius Optimization Methods for Oral Cancer Detection," *IEEE Access*, vol. 11, pp. 23681-23700, 2023.

[4] Andrew EmonHeidari; Sumsum P. Sunny; Bonney L. James; Tracie M. Lam; Anne V. Tran; Junxiao Yu; Ravindra D. Ramanjinappa; Uma K.; Praveen Birur; Amritha Suresh; Moni A. Kuriakose; Zhongping Chen; Petra Wilder-Smith "Optical Coherence Tomography as an Oral Cancer Screening Adjunct in a Low Resource Settings," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, pp. 1-8, Jan.-Feb. 2019.

[5] C. -H. Chan, T. -T. Huang, C. -Y. Chen, C. -C. Lee, M. -Y. Chan and P. -C. Chung, "Texture-Map-Based Branch-Collaborative Network for Oral Cancer Detection," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 4, pp. 766-780, Aug. 2019.

[6] Mohammed Zubair M Shamim; Sadatullah Syed; Mohammad Shiblee; Mohammed Usman; Syed Jaffar Ali; Hany S Hussein; Mohammed Farrag "Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer," *in The Computer Journal*, vol. 65, no. 1, pp. 91-104, Jan. 2020.

[7] Neha Sharma; DebaleenaNawn; SawonPratiher; SayaniShome; Ritam Chatterjee; Karabi Biswas; Mousumi Pal; RanjanRashmi Paul; Srimonti Dutta; Jyotirmoy Chatterjee, "Multifractal Texture Analysis of Salivary Fern Pattern for Oral Pre-Cancers and Cancer Assessment," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9333-9340, 1 April1, 2021.

[8] Shipu Xu; Chang Liu; YongshuoZong; Sirui Chen; Yiwen Lu; Longzhi Yang; Eddie Y. K. Ng; Yongtong Wang; Yunsheng Wang; Yong Liu; Wenwen Hu; Chenxi Zhang, "An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 158603-158611, 2019.

[9] C. Huang, G. Zhang, S. Chen and V. H. C. de Albuquerque, "An Intelligent Multisampling Tensor Model for Oral Cancer Classification," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7853-7861, Nov. 2022.

[10] Mark Marsden; Brent W. Weyers; Julien Bec; Tianchen Sun; Regina F. Gandour-Edwards; Andrew C. Birkeland; Marianne Abouyared; Arnaud F. Bewley; D. Gregory Farwell; Laura Marcu, "Intraoperative Margin Assessment in Oral and Oropharyngeal Cancer Using Label-Free Fluorescence Lifetime Imaging and Machine Learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 857-868, March 2021.

[11] Abdulkader M. Albasri, Abdulrahman H. Ali, Ayesha A. Nathiha, " Segmentation of immunohistochemical staining of β-catenin expression of oral cancer using EM algorithm," Journal of Taibah University Medical Sciences, Vol. 10, no. 2, pp. 169-174, June 2015.

[12] Jonathan Folmsbee, Lei Zhang, Xulei Lu, Jawaria Rahman, John Gentry, Brendan Conn, MarilenaVered, Paromita Roy, Ruta Gupta, Diana Lin, ShabnamSamankan, Pooja Dhorajiva, Anu Peter, Minhua Wang, Anna Israel, Margaret Brandwein-Weber, Scott Doyle, " Histology segmentation using active learning on regions of interest in oral cavity squamous cell carcinoma, " *Journal of Pathology Informatics*, Vol. 13, pp. 100146, 2022.

[13] M. M. Fraz, S. A. Khurram, S. Graham, M. Shaban, M. Hassan, A. Loya& N. M. Rajpoot, " FABnet: feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer, " *Neural Comput&Applic*, no. 32, pp. 9915–9928, 2020.

[14] Dalí F. D. dos Santos, Paulo R. de Faria, Bruno A. N. Travençolo& Marcelo Z. do Nascimento, "Influence of Data Augmentation Strategies on the Segmentation of Oral Histological Images Using Fully Convolutional Neural Networks," *J Digit Imaging*, no. 36, pp. 1608–1623, 2023.

[15] Bishal Bhandari, AbeerAlsadoon, P. W. C. Prasad, Salma Abdullah & Sami Haddad, "Deep learning neural network for texture feature extraction in oral cancer: enhanced loss function," *Multimed Tools Appl*, vol. 79, pp. 27867–27890, 2020.

[16] Ying Qian, "Image denoising algorithm based on improved wavelet threshold function and median filter," *In 2018 IEEE 18th International Conference on Communication Technology (ICCT)*, pp. 1197-1202, 2018.

[17] Ruiqin Fan, Xiaoyun Li, Sanghyuk Lee, Tongliang Li, and Hao Lan Zhang, "Smart image enhancement using CLAHE based on an F-shift transformation during decompression," *Electronics* 9, no. 9, pp.1374, 2020.

[18] Junfeng Jing, Zhen Wang, Matthias Rätsch, and Huanhuan Zhang, "Mobile-Unet: An efficient convolutional neural network for fabric defect detection," *Textile Research Journal*, vol. 92, no. 1-2, pp.30-42, 2022.

[19] Youngsaeng Jin, David Han, and Hanseok Ko, "Trseg: Transformer for semantic segmentation," *Pattern recognition letters*, vol. 148, pp.29-35, 2021.

[20] Chujie Tian, Xinning Zhu, Zheng Hu, and Jian Ma, "Deep spatial-temporal networks for crowd flows prediction by dilated convolutions and region-shifting attention mechanism," *Applied Intelligence*, vol. 50, pp.3057-3070, 2020.

[21] E. Bousias Alexakis and C. Armenakis, "Evaluation of UNet and UNet++ architectures in high resolution image change detection applications." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp.1507-1514, 2020.

[22] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," *In ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 1055-1059, 2020.

[23] Hao Zhang, Xianggong Hong, Shifen Zhou, and Qingcai Wang, "Infrared image segmentation for photovoltaic panels based on Res-UNet," *In Chinese conference on pattern recognition and computer vision (PRCV)*, pp. 611-622, 2019.

[24] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv, *Computer Vision and Pattern Recognition*, 2021.