**Journal of Electronics, Electromedical Engineering, and Medical Informatics**
**Multidisciplinary: Rapid Review: Open Access Journal** **Vol. 6, No. 4, October 2024, pp: 489-498; eISSN: 2656-8632**

**RESEARCH ARTICLE** **OPEN ACCESS**

Manuscript received July 20, 2024; revised August 13, 2024; accepted August 15, 2024; date of publication October 20, 2024
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v6i4.511

How to cite: Zendi Iklima, Trie Maya Kadarina, Rinto Priambodo, Riandini, Rika Novita Wardhani, Sulis Setiowati, "Dental Caries Segmentation using Deformable Dense Residual Half U-Net for Teledentistry System", Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 4, pp. 489-498–, October 2024.

# Dental Caries Segmentation using Deformable Dense Residual Half U-Net for Teledentistry System

**Zendi Iklima[1], Trie Maya Kadarina[1], Rinto Priambodo[2], Riandini[3], Rika Novita Wardhani[3], Sulis Setiowati[3]**

[1]Department of Electrical Engineering, Universitas Mercu Buana, Jakarta, Indonesia
[2]Department of Information System, Universitas Mercu Buana, Jakarta, Indonesia
[3]Department of Electrical Engineering, Politeknik Negeri Jakarta, Jakarta, Indonesia

Corresponding author: Trie Maya Kadarina (e-mail: trie.maya@mercubuana.ac.id).

The successful culmination of this research endeavor is attributed to the invaluable support and collaborative partnership established with the Department of Electrical Engineering and the Research Center of Universitas Mercu Buana, alongside the Department of Electrical Engineering at Politeknik Negeri Jakarta. The authors convey heartfelt gratitude to these esteemed institutions for their provision of essential resources, guidance, and motivation throughout the research endeavor. The unwavering dedication exhibited by these entities has served as the cornerstone in achieving the project's objectives.

**ABSTRACT** Clinical practitioners' workload and challenges are significantly reduced by classifying, predicting, and localizing lesions or dental caries. In recent research, a high-reliability diagnostic system within deep learning models has been implemented in a clinical teledentistry system. In order to construct an efficient, precise, and lightweight deep learning architecture, it is dynamically structured. In this paper, we present an efficient, accurate, and lightweight deep learning architecture for augmenting spatial locations and improving the transformation modeling abilities of fixed-structure CNNs. Deformable Dense Residual (DDR) enhances the efficacy of the residual convolution block by optimizing its structure, thereby mitigating model redundancy and ameliorating the challenge of vanishing gradients encountered during the training stages. DDR Half U-Net presents notable advancements to the simplified U-Net framework across three pivotal domains: the encoder, the decoder, and dense residual connection between them. Specifically, the encoder integrates deformable convolutions, thereby enhancing the model's capacity to discern features of diverse scales and configurations. In the decoder, a sophisticated arrangement of dense residual connections facilitates the fusion of low-level and high-level features, contributing to comprehensive feature extraction. For optimal efficiency in medical image segmentation, use Half U-Net ( $X^{3}_{De}$ ) due to its superior Dice Score of 0.8437, reduced FLOPs of 6.93e+09, and shorter training time of 0.16 hours, while DDR Half U-Net ( $X^{3}_{De}$ ) is recommended for balanced performance with a Dice Score of 0.8352 and moderate FLOPs of 1.02e+10, though it has a slightly longer training time of 0.18 hours. Moreover, the utilization of a weight-adaptive loss function ensures equitable consideration of both caries and non-caries samples, thereby promoting balanced optimization during training.

**INDEX TERMS** Dental Caries, Teledentistry System, Deformable Dense Residual (DDR), Deformable Dense Residual Half U-Net.

## I. INTRODUCTION

To enhance dental health services, it is crucial to emphasize the adoption of technological advancements within the telemedicine field. Encouraging the integration of digital tools, remote monitoring, and teleconsultations can significantly improve access to care. In particular, promoting the widespread adoption of teledentistry in rural areas can address many dental health challenges by providing remote consultations, screenings, and follow-up care[1]. By reaching underserved populations, teledentistry can contribute to

reducing dental disease prevalence [2]. Encourage the adoption of teledentistry as a cost-effective alternative. Highlight its ability to reduce expenses associated with traditional clinical approaches. By utilizing remote consultations, digital tools, and telemonitoring, practitioners can provide efficient and affordable dental care. Integrate teledentistry platforms to enable seamless exchange of diagnostic and treatment information among dental practitioners [3]. This facilitates collaborative decision-making and allows for real-time guidance and mentorship in addressing dental issues [4].

Integrating artificial intelligence, particularly deep learning techniques such as Convolutional Neural Networks (CNN) into medical imaging analysis can be used to improve caries detection and classification [5], [6]. AI systems can learn to reliably detect and identify objects using cameras by training on large datasets, improving recognition accuracy and reducing the workload on human operators [7], [8]. Given the significance of medical segmentation and diagnosis in healthcare, spend funding to research and development activities that focus on CNN-based techniques [9]. Healthcare businesses may open up new options for enhancing diagnostic accuracy, treatment efficacy, and overall patient care by investing in advanced CNN techniques suited for medical purposes [10], [11].

This advancement considerably improves the dental caries detection system's ability to recognize abnormal circumstances while requiring little radiation exposure. However, the system has a flaw in that it is unable to create high-quality detection, which leads to erroneous diagnosis. For example, an accurate and precise caries detection system is critical for dental caries treatment since teeth and bone regions seem similar in photographs, particularly along the boundaries of dental caries images [12], [13].

The deep learning model designed to classify dental caries was trained on a dataset comprising 45 images of caries and 15 images of non-caries. The architecture was methodically built, with a sequence of 2 convolutional layers for feature extraction, 2 pooling layers for dimensionality reduction, 3 dropout layers to prevent overfitting, 1 flatten layer to convert the output into a one-dimensional array, and 2 dense classification layers. This architectural style was meticulously refined to successfully distinguish between carious and non-carious states in dental images. Following thorough training and validation, the model displayed impressive performance, obtaining an accuracy rate of 71.43%. This degree of accuracy demonstrates the model's usefulness in automated dental caries identification, while more refining and validation may be required to improve its diagnostic skills and resilience for practical implementation [14].

Zhang Y constructed a Convolutional Neural Network (CNN) model that was tuned to increase detection efficiency by fourfold. This model uses MobileNetV2 and Depth-wise Separable Convolution (DSC) architectures to identify dental decay in pediatric patients. The use of MobileNetV2 with DSC indicates a strategic decision aimed at improving accuracy and processing efficiency. By combining these structures, Zhang Y's model can efficiently detect dental caries in pediatric patients while greatly lowering the computing resources required for detection. This improvement is especially useful in clinical situations where prompt and precise diagnosis is required, allowing for immediate intervention and treatment planning [15].

Indeed, one of the shortcomings of conventional CNN classification is its inability to correctly detect regions with dental caries. Classification models often produce binary labels of caries or non-caries, without providing comprehensive information on the size or location of carious lesions. As a result, there is a need to create segmentation models that can properly identify carious zones in dental pictures. Segmentation models, such as U-Net or Mask R-CNN, offer the capability to precisely delineate object boundaries within images. By incorporating such models into the dental caries detection framework, it becomes possible to not only classify the presence of caries but also to precisely delineate the affected regions within the image. This level of detail is invaluable for clinicians in treatment planning and monitoring disease progression. The development of segmentation models for dental caries detection represents a promising direction in advancing the accuracy and clinical utility of automated dental diagnostic systems. By combining classification and segmentation approaches, these systems can provide comprehensive and precise assessments of dental health, facilitating more effective patient care and management [16].

The segmentation model is implemented on near-infrared transillumination (TI) images, aiming to improve caries detection precision. Leveraging the differences in near-infrared light scattering and absorption, the model utilizes convolutional neural networks (CNNs) for semantic segmentation, achieving an overall average intersection-over-union (IOU) score of 72.7% and an AUC of ~50% in a 5-class segmentation task. This study utilizes a dataset comprising 217 grayscale images, addressing issues of training data scarcity and class imbalance through strategies such as data augmentation, batch normalization, and weighted loss functions. To expedite model training, the resolution of each image is reduced from 480 × 640 pixels to 256 × 320 pixels. Subsequently, a reference segmentation map is generated from our images through clinical expert interpretation using DIAGNOcam. Five segmentation classes are considered: background (B), enamel (E), dentin (D), proximal caries (PC), and occlusal caries (OC). Despite challenges in accurately delineating lesion boundaries, this model surpasses previous efforts, underscoring the potential of deep learning for accurate caries detection. Future work may involve increasing dataset sizes, reducing label artifacts, and integrating diverse inputs. In this study, a deep encoder model with residual connections and dilated convolution is employed to produce more accurate multiscale predictions [17].

Fariza A, et al., propose a novel model for segmenting dental X-ray images to distinguish teeth from their

background. The U-Net model is employed to perform tooth and background segmentation in dental X-ray images. The training process involves data augmentation, preprocessing with Contrast Limited Adaptive Histogram Equalization (CLAHE), and gamma adjustment. The U-Net convolutional network, comprising encoding and decoding units, is utilized for automatic segmentation, achieving an average accuracy of 97.60%. The methodology includes preprocessing of dental X-ray images, dataset augmentation, U-Net model training, and segmentation prediction on the test dataset. The dataset utilized consists of 119 dental X-ray images with ground truth, which are augmented to 5000 images with a resolution of 224 x 224. This study underscores the importance of adequate ground truth data for training and validation purposes. The evaluation of U-Net segmentation results demonstrates superior performance with an average accuracy of 97.60%. However, some segmentation errors are identified, encompassing 14.58% of the testing dataset [18].

Previous research has shown that computer-based diagnostic systems significantly enhance the effectiveness of high-reliability diagnostic equipment. In medical imaging, AI has been used for classifying, predicting, and localizing lesions or dental caries, which alleviates clinician workload and challenges. Recent studies have focused on simplifying U-Net architectures by introducing the Half-UNet model, which features a block-based structure. This simplified model demonstrated improved efficiency and achieved a 0.83% higher dice coefficient, indicating better segmentation accuracy compared to traditional U-Net models. Deformable Dense Residual Half UNet, this adaptability is refined through an improved deformation modeling mechanism and feature mimicking techniques, leading to better focus on relevant image regions and significantly enhanced performance in tasks like object detection and instance segmentation [19], [20]. Although the Half U-Net model has simplified the architecture and improved segmentation accuracy, there remains a need for further enhancement in adaptability to diverse image features and improved performance in detailed segmentation tasks.

This study presents a novel method which enhances both DDR and Half U-Net architecture. This study presents a novel method which enhances both DDR and Half U-Net architecture. It adds deformable convolutions in the encoder and Residual Multilevel Dense (MLDRC) connections in the decoder. The aim is to enhance the Half U-Net model's performance, particularly in teledentistry systems. The contributions of this paper are:

- DDR Half U-Net architecture: we introduce the DDR Half U-Net, a simplified and more efficient model that is better suited for use in teledentistry systems.
- Improved feature extraction: by adding deformable convolutions to the encoder, the model can better handle different shapes, sizes, and scales, leading to more accurate image segmentation.
- Enhanced segmentation performance: we improve the model's ability to focus on important image areas by adding Residual Multilevel Dense connections in the decoder, resulting in better performance in detecting and segmenting medical images.

## II. METHODS

### A. DATASET AND PREPROCESSING

The dataset distribution is structured with 67% allocated to the training set, 23% to the validation set, and 10% to the evaluation set in order to achieve stable model performance, avoiding overfitting or underfitting of the U-Net model. A total of 200 Dental Caries instances were compiled using an annotation tool for inclusion in the dataset. Segregation of datasets into discrete entities encompassed a training set, a validation set, and a test set. The training set served the purpose of weight adjustment, whereas the validation set facilitated the determination of optimal weights. Finally, the test set enabled thorough performance evaluation [21]. Data preprocessing assumes a pivotal role in the segmentation pipeline, as it holds the potential to amplify the efficiency and resilience of image segmentation algorithms. This article advocates for the adoption of data preprocessing techniques, specifically image augmentation. Image augmentation entails the application of various image transformations, including rotation, flipping, scaling, and cropping, aimed at expanding the dataset. Such augmentation endeavors serve to enrich the diversity and variability of the images, thereby serving as a mitigative measure against overfitting. The research methodologies employed alongside the dental caries dataset and underlying Ground Truth. The augmentation approach involves resizing the input images from their original resolution to 128x128 pixels. Subsequently, 'tf.reduce_mean' is utilized, which is a function employed to compute the mean of tensor elements along specified axes. This operation is fundamental in various neural network tasks, including loss computation and performance evaluation.

### B. PROPOSED METHOD

In the field of medical image segmentation, the U-Net architecture is a widely used approach. This model, distinguished by its unusual U-shaped architecture, incorporates skip connections, allowing the decoder to merge high-level semantic feature maps with low-level detailed feature maps generated by the encoder. Nonetheless, a prevailing misconception suggests that the efficacy of U-Net hinges solely on its U-shaped architecture, resulting in the proliferation of numerous U-Net-based models. Within the architecture denoted by the term "U," two distinct pathways emerge: the encoder, situated on the left and referred to as the contracting layer, and the decoder, positioned on the right and termed the expanding layer. The encoder undertakes downsampling operations to diminish the dimensions of the input matrix while concurrently increasing the number of feature maps. In contrast, the decoder pathway executes operations in an inverse manner, striving to restore the matrix to its original dimensions by diminishing the number of feature maps. Henceforth, the feasibility of conducting a

meticulous pixel-level comparison between segmentation results and the ground truth is realized. U-Net facilitates the coherent propagation of feature maps from each tier of the contracting trajectory to their corresponding echelons within the expanding trajectory, thereby endowing the classifier with the capacity to discern attributes of varying magnitudes and complexities.

The U-Net architecture has become a cornerstone due to its innovative U-shaped design, which cleverly combines high-level semantic information with low-level details through skip connections [22]. However, there is a common misconception that the effectiveness of U-Net is solely dependent on this distinctive architecture, leading to a proliferation of U-Net-based models that may not fully address existing challenges. The U-Net's encoder-decoder structure, while powerful, involves complex training processes and often requires significant computational resources. Despite its strengths, U-Net and similar models face limitations in accuracy and practical application [23], [24], especially when it comes to detailed segmentation tasks like dental caries detection.

Traditional models struggle with precisely identifying and delineating carious lesions, and their performance can be hindered by issues like data scarcity and model complexity. To bridge these gaps, our research introduces a novel architecture specifically designed to address the limitations of existing models. Our approach integrates advanced techniques, such as deformable convolutions [25], into a simplified version of Half U-Net. This tailored architecture enhances the computational efficiency of caries detection while effectively reducing loss during model training. By developing this innovative model, we aim to overcome the existing challenges in dental caries segmentation, offering a more effective and practical solution for precise diagnostics.

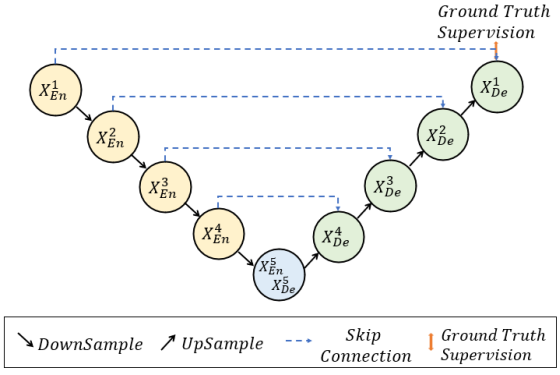


**FIGURE 1. UNet Architecture**

This advancement not only improves detection capabilities but also makes advanced segmentation technology more accessible, particularly in resource-constrained environments where traditional models may be impractical. Additionally, the UNet architecture is acknowledged for its intricacies in training, demanding considerable temporal investments for training execution [26], [27].

FIGURE 1. illustrates the feature map construction of UNet encoder and decoder each denotes $X^i_{De}$ as $X^i_{En}$ and $X^i_{De}$, let $i$ as indexes in down-sampling layer along the encoder, $X^i_{De}$ can be computes as Eq. (1) [28]:

$$X^i_{De} = \{^{X^i_{En}, i=N}_{H([C(D(X^k_{En}))^{i-1}_{k=1}, C(X^i_{En}), C(U(X^k_{De}))^{N}_{k=i+1}]), i=i,\ldots N-1} \quad (1)$$

where $H(.)$ is the feature aggregation mechanism which consist of a batch normalization and the activation function. The function $C(.)$ denotes feature map operation containing up-sampling $U(.)$ and down-sampling $D(.)$ operations.

The number parameters in the $i^{th}$ UNet decoder can be computed as in Eq. (2) [28]:

$$P^i_{U-De} = D_F \times D_F \times [d(X^{i+1}_{De}) \times d(X^i_{De}) + d(X^i_{De})^2 + d(X^i_{En} + X^i_{De}) \times d(X^i_{De})] \quad (2)$$

where $D_F$ is the kernel size of the convolution, $d(.)$ denotes as the node depth. A set of inter-encode-decode skip connections transmits the low-level tensor from $X^i_{En}$ mini-encoder into $X^{i+1}_{En}$, where a non-overlapping max pooling operation is executed. In UNet, the decoder skip connection uses bilinear interpolation to transport high-level tensors from decoder $X^i_{De}$ to decoder $X^{i+1}_{De}$. The UNet decoder has a more detailed feature map operation than the Half-UNet decoder, which displays a symmetric decoder. FIGURE 2 represents the Half UNet architecture given by decoder in $X^3_{De}$.

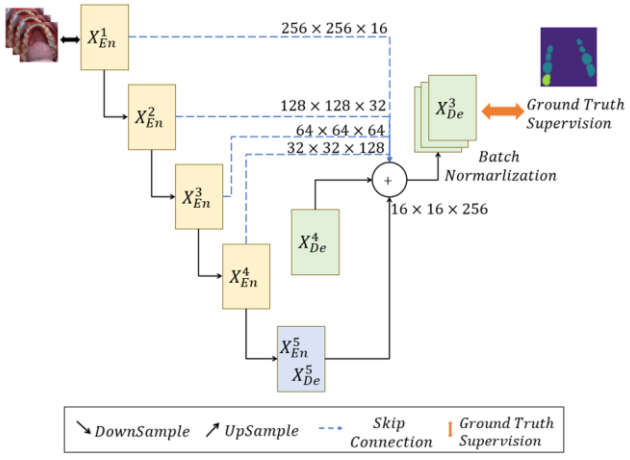


**FIGURE 2. Half U-Net (in $X^3_{De}$) Architecture**

The first step involves the consolidation of channel numbers within Half U-Net, a measure intended to streamline the network structure and augment feature fusion within the decoder. Following this, in order to mitigate the proliferation of additional parameters and Floating-Point Operations (FLOPs) associated with full-scale feature aggregation, a novel technique known as full-scale feature fusion is introduced, serving to replace the four same-scale feature

fusions present in U-Net [29]. Deformable Dense Residual U-Net (DDR U-Net) presents a novel method developed for high-accuracy and computationally efficient ore image segmentation. This method aims to address challenges in segmenting ore particles from images, crucial for measuring ore particle size distribution. DDR-UNet enhances the traditional UNet architecture by integrating deformable convolutions into the U-Net decoder. Performance evaluation of DDR U-Net on ore datasets compared with ten other segmentation methods demonstrates superior segmentation accuracy and efficiency. Moreover, DDR U-Net accurately estimates feature size distribution. Deformable convolution enables DDR U-Net to handle diverse shapes, sizes, and scales of feature extraction originating from irregular and overlapping feature extraction distributions. Additionally, DDR U-Net exhibits a smaller model size and faster computational time compared to evaluated network-in-network architectures [30]. FIGURE 3 shows the feature extraction of conventional convolution and Deformable convolution.

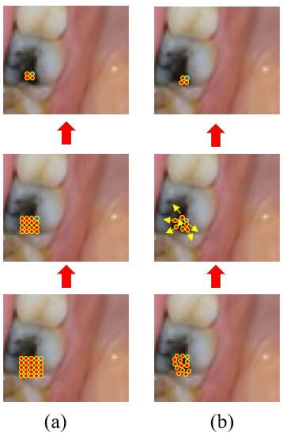


**FIGURE 3. Feature Extraction of (a) Conventional Convolution and (b) Deformable Convolution**

U-Net disregards object shape particulars in images by relying on pre-established geometrically shaped convolution kernels. Furthermore, the inclusion of pooling and convolution operations with significant strides induces the loss of intricate spatial information. To combat the information degradation stemming from pre-defined geometric structures, a residual framework is introduced to produce feature maps and avert information loss. This initiative is geared towards enhancing the adaptability of deformable convolution kernels to diverse shapes and positions within the dataset. Bilinear interpolation is employed to compute pixel values at the final sampling location, as offset values often assume fractional values, rendering the deformable convolution sampling locations irregular [31], [32], [33]. FIGURE 4 represents the DDR Half UNet architecture given by decoder in $X_{De}^{4}$.

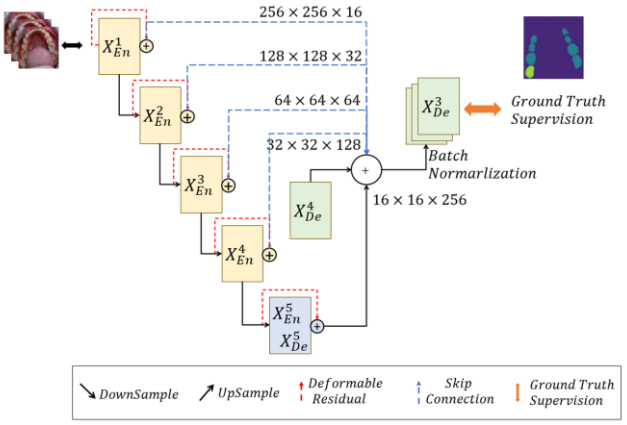


**FIGURE 4. DDR Half U-Net (in $X_{De}^{4}$) Architecture**

### C. PERFORMANCE METRICS

In assessing segmentation performance, the Dice coefficient serves as a valuable metric, quantifying the extent of overlap between the model's predictions and the corresponding ground truth regions. It is derived by dividing the intersection area of the two regions by the sum of their individual areas. A higher Dice coefficient signifies a closer alignment between the model's predictions and the ground truth, thereby leading to enhanced picture segmentation results. The Dice coefficient as shown in Eq. (3) is computed utilizing the following formula, wherein the model's prediction result area is represented as $P$ and the ground truth of breast lumps is denoted as $M$ [27].

$$Dice = \frac{2 \mid P \cup M \mid}{\mid P \mid + \mid M \mid} \tag{3}$$

## III. RESULT

This study introduces several enhancements to the classic UNet architecture, namely the DDR Half UNet architecture, which aims to improve efficiency and performance in semantic image segmentation. This architecture comprises encoder and decoder modules interconnected via skip connections. This design has garnered acclaim for its efficacy in tasks such as image segmentation, leveraging the encoder to extract features and the decoder to fine-tune the segmentation map. This structure is well-known for its effectiveness in image segmentation tasks, where the encoder extracts features and the decoder refines the segmentation map. The inclusion of deformable convolutions within DDR Half UNet optimizes the model's ability to extract features efficiently. By dynamically adjusting sampling points, these convolutions reduce redundancy in feature maps, thereby enhancing computational efficiency while maintaining or even improving segmentation accuracy compared to traditional convolutions. Compared to standard UNet and other variants, DDR Half UNet demonstrates lower FLOPs due to its streamlined architecture and the strategic use of deformable convolutions. This reduction in computational complexity

translates to faster inference times and lower resource demands, making it suitable for real-time applications or scenarios with limited computational resources.

The models were trained for 75 epochs using NVIDIA-SMI T4 GPUs, which recognized that Tensor Core provided efficient deep learning performance. TABLE 1 shows the models architectures of U-Net, Half U-Net, Half U-Net ( $X_{De}^{3}$ ), DDR U-Net, DDR Half U-Net, and DDR Half U-Net ( $X_{De}^{3}$ ).

**TABLE 1**
**Model Architectures**

| Model | Block ( $X_{En}^{i}$ ) | Block ( $X_{De}^{i}$ ) | FLOPs |
|---|---|---|---|
| U-Net | 5 | 5 | 1.69e+10 |
| Half U-Net ( $X_{De}^{4}$ ) | 5 | 2 | 6.93e+09 |
| Half U-Net ( $X_{De}^{3}$ ) | 5 | 1 | 4.95e+10 |
| DDR U-Net | 5 | 5 | 9.56e+09 |
| DDR Half U-Net ( $X_{De}^{4}$ ) | 5 | 2 | 5.56e+09 |
| DDR Half U-Net ( $X_{De}^{3}$ ) | 5 | 1 | 3.91e+09 |

FLOPs (Floating Point Operations per Second) is a metric that measures the number of floating point operations performed by a deep learning model during one inference or training cycle. Comparing FLOPs across all mentioned UNet models U-Net, Half U-Net ( $X_{De}^{4}$ ), Half U-Net ( $X_{De}^{3}$ ), DDR U-Net, DDR Half U-Net ( $X_{De}^{4}$ ), and DDR Half U-Net ( $X_{De}^{3}$ ), it is evident that U-Net exhibits the highest FLOPs, followed by DDR U-Net. Despite U-Net having the highest FLOPs, the DDR Half U-Net ( $X_{De}^{3}$ ), models show potential for reducing FLOPs with more efficient structures, such as the use of deformable convolutions. DDR Half U-Net ( $X_{De}^{3}$ ) stands out for its low FLOPs (6.56e+09) and outstanding segmentation performance, demonstrating that architecture optimization may greatly accelerate model training and inference with the purpose of balancing computational speed, memory requirements, and desired segmentation performance. TABLE 2 and FIGURE 5 represents the model performance of U-Net, Half U-Net ( $X_{De}^{4}$ ), Half U-Net ( $X_{De}^{3}$ ), DDR U-Net, DDR Half U-Net ( $X_{De}^{4}$ ), and DDR Half U-Net ( $X_{De}^{3}$ ).

**TABLE 2**
**Model Performances**

| Model | Train Loss (%) | Dice (%) | Time Exec. (ours) |
|---|---|---|---|
| U-Net | 0.8201 | 0.7963 | 0.43 |
| Half U-Net ( $X_{De}^{4}$ ) | 0.3419 | 0.8437 | 0.16 |
| Half U-Net ( $X_{De}^{3}$ ) | 0.3079 | 0.8352 | 0.11 |
| DDR U-Net | 0.3414 | 0.8348 | 0.14 |
| DDR Half U-Net ( $X_{De}^{4}$ ) | 0.4039 | 0.8332 | 0.14 |
| DDR Half U-Net ( $X_{De}^{3}$ ) | 0.3104 | 0.8491 | 0.11 |

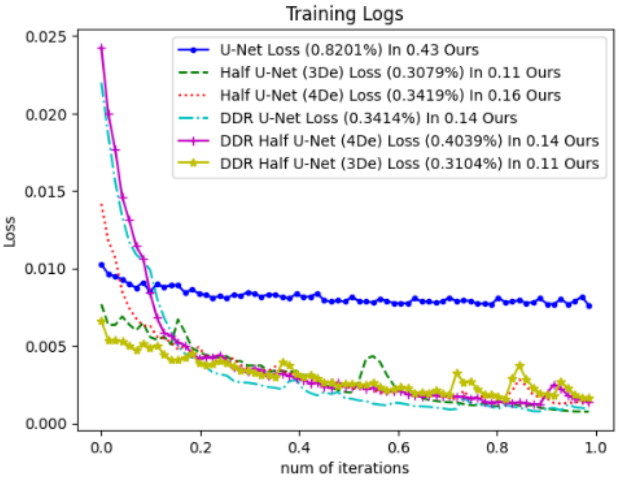


**FIGURE 5. Model Performance of (a) U-Net, (b) Half U-Net, (c) Half U-Net ( $X_{De}^{4}$ ), (d) Half U-Net ( $X_{De}^{3}$ ), (e) DDR U-Net, (f) DDR Half U-Net, (g) DDR Half U-Net ( $X_{De}^{4}$ ), and (h) DDR Half U-Net ( $X_{De}^{3}$ )**

Based on TABLE 2, U-Net has a Dice score of 0.7963, indicating a lower similarity between the model's predictions and the ground truth compared to the Half U-Net ( $X_{De}^{4}$ ) and Half U-Net ( $X_{De}^{3}$ ) architectures. The train loss of 0.8201% indicates how well the model is trained to minimize errors in its training data. With FLOPs of around 1.69e+10, U-Net requires a significant amount of computation, which may affect the training time of 0.43 hours. Half U-Net ( $X_{De}^{4}$ ), with a Dice score of 0.8437, shows a significant improvement in the similarity of predictions to the ground truth compared to U-Net. The lower train loss of 0.3419% indicates that this model may be more effective in minimizing the error in its training data. With lower FLOPs of around 6.93e+09 and a shorter training time of 0.16 hours, Half U-Net ( $X_{De}^{4}$ ) offers a more efficient solution without sacrificing performance. Half U-Net ( $X_{De}^{3}$ ) has a Dice score of 0.8352, which is slightly lower than Half U-Net ( $X_{De}^{4}$ ), but still higher than U-Net. The train loss of 0.3079% shows that this model is effective in the training process. However, with higher FLOPs, around 4.95e+10, this model requires more computation, although its training time is only 0.11 hours, which is shorter than U-Net. The difference in performance between these three architectures can be explained by a combination of factors such as architectural complexity, the use of decoder layers in ( $X_{De}^{3}$ and $X_{De}^{4}$ ), the number of filters, and other hyperparameter settings.

## IV. DISCUSSION

DDR U-Net has a Dice score of 0.8348, indicating a good level of similarity between the prediction results and the ground truth. The train loss of 0.3414% shows the efficiency of the model in minimizing errors during the training process. With FLOPs of around 9.56e+09, DDR U-Net requires quite a large computation, but its training time is relatively efficient, which is 0.14 hours.

This architecture consists of two main parts, namely the encoder block and the decoder block. The encoder block uses two conventional convolution layers followed by an offset convolution layer, which are then combined through deformable convolutions. This process helps the model adjust the spatial pattern of features based on local conditions around it, which can improve segmentation capabilities. Each encoder block is followed by max pooling to reduce the spatial dimension of the extracted features. The decoder block implements an up-convolution operation using Conv2DTranspose to restore the spatial resolution, followed by combining the results with skip connections from the corresponding encoder block. This process allows the model to rebuild the spatial details lost during the pooling process in the encoder of 0.8332, indicating competitive performance, although slightly lower than DDR U-Net.

The train loss of 0.4039% may indicate challenges in the training process that affect the model's accuracy. With FLOPs of around 5.56e+09, this model is more computationally efficient compared to DDR U-Net, with the same training time of 0.14 hours. DDR Half U-Net ( $X_{De}^{3}$ ) has a Dice score of 0.8491, indicating a high degree of similarity between predictions and ground truth. The train loss of 0.3104% indicates the efficiency of the model in the training process to minimize errors.

This architecture has FLOPs of around 3.91e+09, indicating lower computation compared to U-Net but higher compared to Half U-Net ( $X_{De}^{4}$ ), with a training time of around 0.11 hours, indicating good time efficiency. The encoder block in DDR Half U-Net ( $X_{De}^{3}$ ) consists of two conventional convolutional layers followed by an offset convolutional layer using deformable convolutions. This process helps in capturing better spatial patterns of objects in the image, which enriches the feature representation. Each encoder block ends with max pooling to reduce the spatial dimensionality of the extracted features. The decoder block implements an up-convolution operation with Conv2DTranspose to restore the spatial resolution, followed by concatenating the result with skip connections from the corresponding encoder block. This helps in reconstructing the spatial details lost during the pooling process in the encoder block. The use of batch normalization at several key points in this architecture helps in accelerating convergence during training, while the use of the kernel initializer 'he_normal' helps in proper initialization of the model weights.

The difference in performance between these three DDR architectures with U-Net, Half U-Net ( $X_{De}^{4}$ ), and Half U-Net ( $X_{De}^{3}$ ) can be caused using deformable convolution and other configurations such as the number of filters, kernel size, and hyperparameter settings. DDR Half U-Net ( $X_{De}^{3}$ ) stands out with the highest Dice score, indicating that the use of deformable convolution and other configurations can improve the model's ability to segment images. FIGURE 6 shows the segmentation predictions of the U-Net, Half U-Net ( $X_{De}^{4}$ ), Half U-Net ( $X_{De}^{3}$ ), DDR U-Net, DDR Half U-Net ( $X_{De}^{4}$ ), and DDR Half U-Net ( $X_{De}^{3}$ ) models. variations of DDR Half U-Net typically outperform U-Net and Half U-Net variations in segmentation performance parameters while using less computing power. Deformable convolutions are included in segmentation generated by the models, DDR Half U-Net to improve segmentation outcomes by improving the network's capacity to gather precise spatial information. The noted reductions in training times and FLOPs for all DDR Half U-Net configurations further highlight their usefulness in resource-constrained settings or for applications that need quick inference. Due to its ability to maximize both performance and efficiency in semantic segmentation tasks, DDR Half U-Net emerges as a viable design, utilizing recent advances in convolutional approaches.

The efficiency improvements observed in this study for Half U-Net variants resonate with the findings of H. Lu et al. [29], who explored the Half-UNet architecture and highlighted its simplified structure for medical image segmentation. Their research emphasizes reduced computational complexity and training time while maintaining competitive performance metrics, aligning with the reduced FLOPs and training times seen in our Half U-Net variants. Additionally, Sonavane et al. investigated convolutional neural networks (CNNs) for dental cavity classification [13]. Although their work is not directly related to U-Net architectures, it demonstrates the broader application of CNNs and underscores the significance of model efficiency, a goal shared by the improvements observed in this study. Furthermore, F. Li et al. introduced DDR-Unet for ore image segmentation, emphasizing high accuracy and efficiency [30]. Their results align with our findings, where DDR U-Net maintains high performance with moderate FLOPs, supporting the efficacy of deformable convolutions in enhancing model accuracy and managing computational load effectively.

## V. CONCLUSION

In this study, we conducted a detailed comparison performance across several variants of the UNet architecture: U-Net, Half U-Net ( $X_{De}^{4}$ ), Half U-Net ( $X_{De}^{3}$ ), DDR U-Net, DDR Half U-Net ( $X_{De}^{4}$ ), and DDR Half U-Net ( $X_{De}^{3}$ ). FLOPs serve as a critical metric for evaluating the computational complexity of these models, which is crucial for understanding their efficiency in medical image segmentation tasks.

Comparisons between different U-Net architectures using deformable convolutions provide deep insights into the impact of this innovation on network performance and efficiency. Conventional U-Net offers a solid foundation in medical image segmentation, with a Dice Score of around 0.7963 and a training time of around 0.43 hours. However, this architecture has high FLOPs, reaching 1.69e+10, indicating a high computational burden. Then, half-U-Net with deformable convolutions shows significant improvements in performance. For example, Half U-Net ( $X_{De}^{4}$ ) achieves a Dice Score of 0.8437 with lower FLOPs of 6.93e+09 and a training time of only 0.16 hours. Likewise, Half U-Net ( $X_{De}^{3}$ ) achieves a Dice Score of 0.8352 with a shorter training time of 0.11 hours, despite having slightly higher FLOPs than Half U-Net ( $X_{De}^{4}$ ).

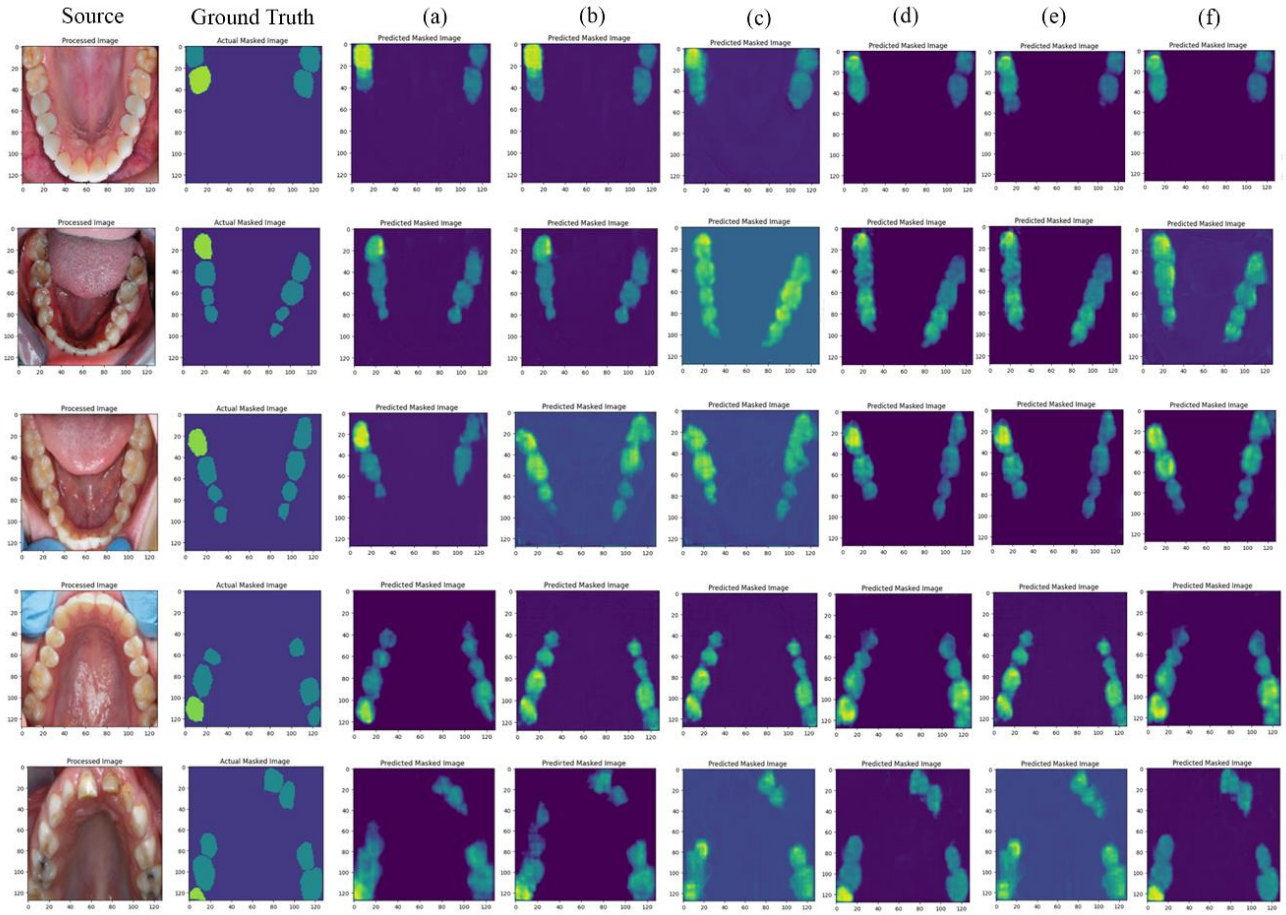


**FIGURE 6.** **Dental Caries Segmentation of (a) U-Net, (b) Half U-Net, (c) Half U-Net ( $X_{De}^{4}$ ), (d) Half U-Net ( $X_{De}^{3}$ ), (e) DDR U-Net, (f) DDR Half U-Net, (g) DDR Half U-Net ( $X_{De}^{4}$ ), and (h) DDR Half U-Net ( $X_{De}^{3}$ )**

DDR U-Net variant of this architecture introduce deformable convolutions to improve the model's ability to capture complex spatial information. DDR U-Net, for example, with a Dice Score of 0.8348, maintains good performance with 9.56e+09 FLOPs. Although the training time is slightly longer than Half U-Net ( $X_{De}^{4}$ ), the use of deformable convolutions helps in improving the segmentation ability.

The findings from our comparison provide valuable insights into optimizing computational efficiency without compromising segmentation accuracy in medical imaging tasks. Models like DDR Half U-Net ( $X_{De}^{3}$ ), with significantly lower FLOPs, represent a promising direction for accelerating training and inference processes. These models leverage advanced techniques in conventional convolution to enhance performance while managing computational resources effectively. This study underscores the benefits of adopting simplified and deformable convolution-based U-Net architectures, which not only enhance performance but also improve computational efficiency, making them more practical for medical imaging applications. The careful consideration of FLOPs in UNet-based architectures is pivotal for optimizing model design and deployment in medical image analysis. By properly harnessing computing resources, researchers and practitioners can increase deep learning models' capacity to handle complex healthcare issues, ultimately boosting diagnostic accuracy and patient care outcomes.

The study's findings are constrained by dataset-specific performance; the results are tied to the particular medical imaging dataset used, which means the observed performance improvements might not generalize to other datasets or medical imaging tasks, thus limiting the applicability of the findings. Additionally, the study does not address scalability concerns; while it discusses efficiency for the specific tasks and dataset at hand, it does not explore how well these models perform with larger or more complex datasets, which could affect their practical deployment in varied clinical settings. The evaluation is also somewhat narrow, focusing primarily on Dice Score and FLOPs, while other crucial aspects such as model robustness, real-time performance, interpretability, and

clinical relevance are not considered, potentially overlooking factors essential for practical application. Furthermore, although training times are discussed, inference times are not covered, which is a critical factor in real-world scenarios, particularly in clinical settings requiring real-time analysis.

Future research could explore hybrid architectures that combine the strengths of different UNet variants with additional techniques such as attention mechanisms or self-attention layers. These enhancements aim to further refine segmentation accuracy and computational efficiency in challenging medical imaging scenarios, such as multi-modal image fusion or real-time clinical decision support systems. Further advancements could address current limitations and increase model versatility, ultimately leading to faster and more efficient medical image analysis, improved clinical decision-making, and better patient outcomes. Additionally, the critical importance of efficient resource management in deep learning models suggests that future studies should continue to innovate in model design and computing efficiency to meet the evolving demands of clinical practice.

## ACKNOWLEDGMENT

The generous support and cooperation of the Department of Electrical Engineering and the Research Center of Universitas Mercu Buana, together with the Department of Electrical Engineering at Politeknik Negeri Jakarta, were indispensable in successfully completing this research. The authors sincerely thank these esteemed institutions for their invaluable resources, guidance, and encouragement throughout the research process. Their steadfast dedication has played a crucial role in achieving the objective.

## REFERENCES

[1] M. S. Alauddin, A. S. Baharuddin, and M. I. M. Ghazali, "The Modern and Digital Transformation of Oral Health Care: A Mini Review," *Healthcare 2021, Vol. 9, Page 118*, vol. 9, no. 2, p. 118, Jan. 2021, doi: 10.3390/HEALTHCARE9020118.

[2] R. Thakkar, J. V Pimpale, A. Kaur, P. Thakkar, M. Sheth, and S. Karre, "Teledentistry for Underserved Populations: An Evidence-Based Exploration of Access, Outcomes, and Implications," 2023. [Online]. Available: www.jrmds.in

[3] M. R. R. Islam *et al.*, "Teledentistry as an Effective Tool for the Communication Improvement between Dentists and Patients: An Overview," *Healthcare 2022, Vol. 10, Page 1586*, vol. 10, no. 8, p. 1586, Aug. 2022, doi: 10.3390/HEALTHCARE10081586.

[4] M. Irving, R. Stewart, H. Spallek, and A. Blinkhorn, "Using teledentistry in clinical practice as an enabler to improve access to clinical care: A qualitative systematic review," *J Telemed Telecare*, vol. 24, no. 3, pp. 129–146, 2018, doi: 10.1177/1357633X16686776.

[5] I. S. Bayrakdar *et al.*, "Deep-learning approach for caries detection and segmentation on dental bitewing radiographs," *Oral Radiol*, vol. 38, no. 4, pp. 468–479, Oct. 2022, doi: 10.1007/S11282-021-00577-9/METRICS.

[6] Y. Liu, K. Xia, Y. Cen, S. Ying, and Z. Zhao, "Artificial intelligence for caries detection: a novel diagnostic tool using deep learning algorithms," *Oral Radiol*, vol. 40, no. 3, pp. 375–384, Jul. 2024, doi: 10.1007/S11282-024-00741-X/METRICS.

[7] R. Muwardi, J. Mada, R. Permana, H. Gao, and M. Yunita, "Human Object Detection for Real-Time Camera using Mobilenet-SSD," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 3, no. 2, pp. 141–150, 2023, doi: 10.5162/jiae.v3i2.108.

[8] R. Muwardi, M. Rhozaly, M. Yunita, and G. Erica Yehezkiel, "Fast Human Recognition System on Real-Time Camera," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 4, pp. 895–903, 2023, doi: 10.26555/jiteki.v9i4.27009.

[9] W. Yao, J. Bai, W. Liao, Y. Chen, M. Liu, and Y. Xie, "From CNN to Transformer: A Review of Medical Image Segmentation Models," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 4, pp. 1529–1547, Mar. 2024, doi: 10.1007/S10278-024-00981-7/METRICS.

[10] J. H. Lee, D. H. Kim, S. N. Jeong, and S. H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *J Dent*, vol. 77, pp. 106–111, Oct. 2018, doi: 10.1016/j.jdent.2018.07.015.

[11] H. Zhu, Z. Cao, L. Lian, G. Ye, H. Gao, and J. Wu, "CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image," *Neural Comput Appl*, vol. 2, 2022, doi: 10.1007/s00521-021-06684-2.

[12] V. Geetha, K. S. Aprameya, and D. M. Hinduja, "Dental caries diagnosis in digital radiographs using back-propagation neural network," *Health Inf Sci Syst*, vol. 8, no. 1, 2020, doi: 10.1007/s13755-019-0096-y.

[13] M. Prados-Privado, J. G. Villalón, C. H. Martínez-Martínez, C. Ivorra, and J. C. Prados-Frutos, "Dental caries diagnosis and detection using neural networks: A systematic review," *J Clin Med*, vol. 9, no. 11, pp. 1–13, 2020, doi: 10.3390/jcm9113579.

[14] A. Sonavane, R. Yadav, and A. Khamparia, "Dental cavity classification of using convolutional neural network," *IOP Conf Ser Mater Sci Eng*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012116.

[15] Y. Zhang, H. Liao, J. Xiao, N. Al Jallad, O. Ly-Mapes, and J. Luo, "A Smartphone-Based System for Real-Time Early Childhood Caries Diagnosis," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12437 LNCS, pp. 233–242, 2020, doi: 10.1007/978-3-030-60334-2_23.

[16] V. Majanga and S. Viriri, "A Survey of Dental Caries Segmentation and Detection Techniques," *Scientific World Journal*, vol. 2022, 2022, doi: 10.1155/2022/8415705.

[17] A. E. Rad, M. S. M. Rahim, H. Kolivand, and A. Norouzi, "Automatic computer-aided caries detection from dental x-ray images using intelligent level set," *Multimed Tools Appl*, vol. 77, no. 21, pp. 28843–28862, 2018, doi: 10.1007/s11042-018-6035-0.

[18] F. Casalegno *et al.*, "Caries Detection with Near-Infrared Transillumination Using Deep Learning," *J Dent Res*, vol. 98, no. 11, pp. 1227–1233, 2019, doi: 10.1177/0022034519871884.

[19] T. M. Kadarina *et al.*, "A simplified dental caries segmentation using Half U-Net for a teledentistry system," *SINERGI*, vol. 28, no. 2, pp. 251–258, Apr. 2024, doi: 10.22441/SINERGI.2024.2.005.

[20] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018, doi: 10.1109/LGRS.2018.2830403.

[21] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl Based Syst*, vol. 178, pp. 149–162, 2019, doi: 10.1016/j.knosys.2019.04.025.

[22] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S...*, vol. 11045, pp. 3–11, 2018, doi: 10.1007/978-3-030-00889-5_1.

[23] X. Yang, Z. Li, Y. Guo, and D. Zhou, "DCU-net: a deformable convolutional neural network based on cascade U-net for retinal vessel segmentation," *Multimed Tools Appl*, vol. 81, no. 11, pp. 15593–15607, May 2022, doi: 10.1007/S11042-022-12418-W/METRICS.

[24] D. Huang, H. Guo, and Y. Zhang, "ADD-Net:Attention U-Net with Dilated Skip Connection and Dense Connected Decoder for Retinal Vessel Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13002 LNCS, pp. 327–338, 2021, doi: 10.1007/978-3-030-89029-2_26.

[25] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets V2: More Deformable, Better Results," 2019.

[26] A. A. Pravitasari *et al.*, "UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1310–1318, 2020, doi: 10.12928/TELKOMNIKA.v18i3.14753.
[27] S. Sukegawa *et al.*, "Deep Neural Networks for Dental Implant System Classification," *Biomolecules 2020, Vol. 10, Page 984*, vol. 10, no. 7, p. 984, Jul. 2020, doi: 10.3390/BIOM10070984.
[28] H. Huang *et al.*, "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2020-May, no. iii, pp. 1055–1059, 2020, doi: 10.1109/ICASSP40776.2020.9053405.
[29] H. Lu, Y. She, J. Tie, and S. Xu, "Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation," *Front Neuroinform*, vol. 16, no. June, pp. 1–10, 2022, doi: 10.3389/fninf.2022.911679.
[30] F. Li, X. Liu, Y. Yin, and Z. Li, "DDR-Unet: A High-Accuracy and Efficient Ore Image Segmentation Method," *IEEE Trans Instrum Meas*, vol. 72, 2023, doi: 10.1109/TIM.2023.3317480.
[31] S. Saumiya and S. W. Franklin, "Residual Deformable Split Channel and Spatial U-Net for Automated Liver and Liver Tumour Segmentation," *J Digit Imaging*, vol. 36, no. 5, pp. 2164–2178, 2023, doi: 10.1007/s10278-023-00874-1.
[32] H. Zhou, H. Leung, and B. Balaji, "AR-UNet: A Deformable Image Registration Network with Cyclic Training," *IEEE/ACM Trans Comput Biol Bioinform*, vol. PP, pp. 1–10, 2023, doi: 10.1109/TCBB.2023.3284215.
[33] W. Yu, B. Liu, H. Liu, and G. Gou, "Recurrent Residual Deformable Conv Unit and Multi-Head with Channel Self-Attention Based on U-Net for Building Extraction from Remote Sensing Images," *Remote Sens (Basel)*, vol. 15, no. 20, 2023, doi: 10.3390/rs15205048.

**ZENDI IKLIMA** received the M.Sc. degree in Software Engineering from Beijing Institute of Technology in 2018. His research interests include robotics, deep learning, and cloud computing. He has an experience as the Chief Technology Officer of Diaspora Connect Indonesia. Indonesian Diaspora Connect is one of the controversial start-up companies in 2018. The idea of connecting Indonesia is to help the Indonesian government find out about citizens who are studying or living abroad which integrates with Artificial Intelligence. Based on the idea, he achieved 1st place in ALE Hackathon Competition in 2018 which held by Alcatel-Lucent (Jakarta). He can be contacted at email: zendi.iklima@mercubuana.ac.id.

**TRIE MAYA KADARINA** received her bachelor's degree in Electrical Engineering from Institut Teknologi Nasional in 2001. She received her master's degree in Biomedical Engineering from the Department of Electrical Engineering, Institut Teknologi Bandung, in 2005. Currently, she is a lecturer in the Department of Electrical Engineering at Universitas Mercu Buana. Her research interests include biomedical instrumentation, electronics and control systems, machine learning, and Internet of Things. She can be contacted at email: trie.maya@mercubuana.ac.id.

**RINTO PRIAMBODO** received his bachelor degree in electrical engineering from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2004 and his master degree in information technology from Universitas Indonesia in 2011. He is currently pursuing the doctoral degree in the faculty of computer science in Universitas Indonesia. His research interest includes information system, business intelligence and Internet of Things in healthcare. He can be contacted at email: rinto.priambodo@mercubuana.ac.id.

**RIANDINI** is active as a Lecturer at the Department of Electrical Engineering, Politeknik Negeri Jakarta, Indonesia. She received her MSc in Biomedical Engineering with Full Funded the Netherlands Govt. Scholarship (STUNED) from Rijksuniversiteit Groningen (RuG), The Netherlands 2007. Her research interests include Image Processing, Deep Learning, Biomedical Engineering among many other fields. She can be contacted at email: riandini@elektro.pnj.ac.id.

**RIKA NOVITA WARDHANI** is a Senior Lecturer and a former Head of Program of Instrumentation and Industrial Control, Department of Electrical Engineering, Politeknik Negeri Jakarta, Indonesia. She received her Master of Engineering in System and Control from the Department of Electrical Engineering, ITS, Surabaya, Indonesia in 1999 fully funded by the YPTS organization. Her research interests cover Smart Sensor and Monitoring System, Modelling System, Neuro-Fuzzy and Internet of Things. She can be contacted at email: rika.novitawardhani@elektro.pnj.ac.id .

**SULIS SETIOWATI** received her M. Eng in Electrical Engineering with Full Funded Scholarship from Indonesia Endowment Fund for Education (LPDP) 2018. She is a lecturer at the Department of Electrical Engineering, Program of Instrumentation and Industrial Control, Politeknik Negeri Jakarta, Indonesia. Her research interests include Internet of Things, Machine Learning and Recommendation Systems.