

Comparative Analysis of GAN-based Segmentation Models on Dental Panoramic Radiography Dataset

Hakan ÖCAL^{1*}, Gürdal ALTUNDAĞ²

¹ Computer Engineering, Faculty of Engineering Architecture and Design, Bartın University, Bartın, Türkiye

² Intelligent Systems Engineering, Graduate School, Bartın University, Bartın, Türkiye

*¹ hocal@bartin.edu.tr, ² 23015248009@ogrenci.bartın.edu.tr

(Geliş/Received: 18/02/2025;

Kabul/Accepted: 15/03/2025)

Abstract: Especially in criminal investigations, the identification of the victim is essential. The branch of forensic medicine that uses the method of identification from the teeth of the victims is called forensic odontology. In forensic odontology, physical information about the individual can be obtained from the bone and enamel structure of the teeth. Panoramic, periapical, and cephalometric imaging techniques are the most commonly used in the odontological identification of the individual. Forensic odontology is increasingly recognized for its essential role in personal identification during mass disasters, sexual assault cases, and child abuse investigations. Deep learning algorithms have recently successfully detected dental disorders such as caries, periodontal bone loss, and apical lesions. Generative adversarial networks (GAN) models have mainly achieved high segmentation performance in medical images. In this study, GAN models were designed and comparatively analyzed using U-Net, Volumetric convolutional neural network (V-Net), spatial and channel Squeeze-Excitation-based U-Net (scSEU-Net), Transformer-based U-Net (TransU-Net), and U-Net like pure Transformer (SwinU-Net) segmentation architectures which are widely used in the literature as generators. As a result of the comparative analyses, scSEU-Net-based GAN achieved the highest performance values with 0.8826 Thresholded Dice (DSC), 0.7901 Thresholded Intersection over Union (Thresh-IoU), 0.9805 Accuracy (ACC), 0.9268 Precision (PREC), and 0.9001 Recall (REC).

Key words: Deep Learning, Dental panoramic radiography segmentation, Generative adversarial networks, U-shaped Segmentation models, V-Net.

Dental Panoramik Radyografi Veri Setinde GAN Tabanlı Segmentasyon Modellerinin Karşılaştırmalı Analizi

Öz: Özellikle ceza soruşturmalarında mağdurun kimliği esastır. Mağdurların dişlerinden kimlik tespiti yöntemini kullanan adli tıp dalına adli odontoloji denir. Adli odontolojide, bireye ait fiziksel bilgiler dişlerin kemik ve mine yapısından elde edilebilir. Bireyin odontolojik kimlik tespitinde en sık kullanılanlar panoramik, periapikal ve sefalometrik görüntüleme teknikleridir. Adli odontoloji, kitlesel felaketler, cinsel saldırı vakaları ve çocuk istismarı soruşturmaları sırasında kişisel kimlik tespitindeki temel rolüyle giderek daha fazla tanınmaktadır. Derin öğrenme algoritmaları son zamanlarda çürük, periodontal kemik kaybı ve apikal lezyonlar gibi diş bozukluklarını başarıyla tespit etmiştir. Üretken Çekişmeli Ağlar (GAN) modelleri çoğunlukla medikal görüntülerde yüksek segmentasyon performansı elde etmiştir. Bu çalışmada, literatürde üreticiler olarak yaygın olarak kullanılan U-Net, Volumetrik Evrişimli Sinir Ağı (V-Net), Uzaysal ve Kanal Sıkıştırma-Uyartım Tabanlı U-Net (scSEU-Net), Transformatör Tabanlı U-Net (TransU-Net) ve U-Net benzeri Saf Transformatör (SwinU-Net) segmentasyon mimarileri kullanılarak GAN modelleri tasarlanmış ve karşılaştırmalı olarak analiz edilmiştir. Karşılaştırmalı analizler sonucunda scSEU-Net tabanlı GAN, 0,8826 Eşikli Zar (DSC), 0,7901 Eşikli Birleşim Kesişimi (Thresh-IoU), 0,9805 Doğruluk (ACC), 0,9268 Hassasiyet (PREC) ve 0,9001 Geri Çağırma (REC) değerleriyle en yüksek performans değerlerine ulaşmıştır.

Anahtar kelimeler: Derin Öğrenme, Dental panoramik radyografi segmentasyonu, Üretken çekişmeli ağlar, U şekilli segmentasyon modelleri, V-Net.

1. Introduction

Machine Learning (ML) based algorithms, especially Deep Learning models based on Convolutional Neural Networks (CNN), are one of the common areas of study in complex segmentation processes in dentistry [1-4]. However, imbalances in real-world data and the fact that standard data is more than pathological data cause deep learning models to perceive pathological data as noise and fail to learn. Due to the scarcity of pathological data in the dataset, deep learning architectures classify the pathological images they see as non-pathological images. As a solution, methods of augmenting pathological data by obtaining synthetic data from scarce pathological data are

* Sorumlu yazar: hocal@bartin.edu.tr. Yazarların ORCID Numarası: ¹ 0000-0002-8061-8059, ² 0009-0001-9935-5893

widely used [5, 6]. Many deep learning-based segmentation models have tried to cope with the difficulties in segmenting dental images [7]. Especially CNN-based architectures with spatial pyramid pooling (R-CNN) and architectures such as Resnet with residual blocks and Generative Adversarial Networks (GAN) have also shown high segmentation performances [8-13].

AI-driven models, particularly those employing CNNs, have also advanced the segmentation and classification of dental images, addressing critical challenges in identifying oral diseases. In dentistry, the first models based on convolutional neural networks (CNNs) and 2D and 3D photography are emerging for the 3D design of dental prostheses, with very encouraging results [14,15]. These advancements enhance diagnostic accuracy and support personalized treatment planning, demonstrating the growing synergy between AI and dental healthcare. While the integration of AI in dentistry is still evolving, several studies underscore its potential to enhance diagnostic workflows. For instance, Mask R-CNNs, GANs, and diffusion models combined with U-Net or V-Net architectures have effectively segmented complex dental structures, addressing anatomical variations and image quality challenges [16].

Despite these advancements, challenges remain, particularly in ensuring the generalizability and robustness of AI models across diverse populations and imaging conditions. Limited datasets, imbalanced learning tasks, and the need for extensive expert annotations highlight the technical and practical hurdles to widespread AI adoption. Addressing these issues requires collaborative efforts to standardize data collection, enhance model training methodologies, and develop explainable AI systems that align with clinical requirements. Deep convolutional neural networks (CNNs) have shown promise in various medical imaging applications, including dental image analysis. For instance, Mask R-CNN has been employed to detect and segment oral diseases, demonstrating its effectiveness in identifying and localizing dental conditions from intraoral images [17,18].

The continuous evolution of deep learning technologies offers significant opportunities to bridge these gaps. The convolutional neural network, a substantial type of deep learning architecture, is best known for its vast visual imagery analysis capabilities. These networks comprise regularized versions of complicated multi-layer perceptrons, enabling comprehensive analysis in various medical applications. Furthermore, CNNs have demonstrated their applicability in identifying and localizing dental conditions from intraoral images. These advanced models allow dental practitioners to optimize workflows and enhance patient care. Various retention and restoration methods, which have been proposed and improved for treating dental caries, have been successfully developed over the past few decades [19,20].

Additionally, emerging technologies such as GANs and improved U- and V-shaped segmentation architectures offer innovative solutions to long-standing challenges in dentistry, such as segmentation of complex dental structures and accurate tooth identification on panoramic radiographs. These technologies are particularly effective in addressing the unique challenges presented by complex anatomical structures and differences in tooth morphology among adults. By comparing the estimated dental images obtained by the segmentation-based generators of GAN architectures with the images segmented by forensic experts with the help of a discriminator, segmentation architectures can be better optimized by distinguishing whether the forensic expert segments the images.

Among the studies on the segmentation of dental panoramic radiography images in the literature, Li et al. attempted to segment dental panoramic X-ray images using the fine-tuned Segment Anything Model 2 (SAM 2) [21]. Şahin et al. used the U-Net 3+ model for fine segmentation of a dataset of 2,693 dental panoramic radiographs of children aged 2 to 13 years [22]. Yang et al. segmented dental panoramic radiographs obtained from 100 patients using a ResU-Net-based CNN model. The proposed study used Principal Component Analysis (PCA) and linear regression techniques to feature extracted tooth regions in the image [23]. Özçelik et al. proposed the Squeeze and Excitation Inception Block-based Encoder-Decoder (SE-IB-ED) architecture to segment teeth from panoramic X-ray images. They adopted the polygonal semantic pixel labeling (PSPL) method in labeling teeth [24]. Zhou et al. created a dual-labeled dataset to find both the status of teeth and the number of teeth affected in panoramic radiographs. They segmented the dataset with a GAN-based architecture consisting of the proposed end-to-end fully automatic YOLOv9-e segmentation model and the EfficientNetV2-I discriminator [25]. Kong et al. used Super-Resolution Generative Adversarial Networks (SRGANs) architecture, which combines super-resolution and GAN architectures, to segment periodontal bone loss [26]. Altundağ and Öcal proposed a Semi-stage Hybrid Convolution U-Net-based GAN architecture that reduced the number of parameters from 22 million to 11 million by modifying the U-Net architecture in the U-Net-based GAN architecture used for segmentation of teeth from panoramic radiographs [27].

In this study, considering the high-performance values of GAN architectures in Segmentation, GAN models in which segmentation models that have achieved high-performance results in the literature are used as generators and fed to a discriminator are comparatively analyzed.

2. Materials and Methods

This section gives comprehensive information about the dataset used, the deep learning models on which comparative analyses were made, the loss function used, and the performance metrics.

2.1. Preparing the panoramic radiography dataset

The dataset for training and testing the proposed model consists of high-resolution digital X-ray and radiographic dental images obtained from forensic institutions. The publicly available dataset to be used includes a total of 7000 images covering various age groups and tooth types. The dataset to be used was published in the Scientific Data Journal of Nature and shared with scientists [9]. The images in the dataset were taken at different angles, using various lighting conditions and imaging techniques. The dataset was divided as follows for training and testing: In the training phase, 5000 images were used for model training, while 1000 images were separated to monitor the model's performance during training and to prevent overfitting. In addition, 1000 images were used for the final performance evaluation and testing of the model. Examples from the panoramic dental radiography dataset are shown in Figure 1. In the figure, the images in the upper row show the panoramic dental radiography images of the cases, and the images in the lower row show the ground truths of the radiography images.

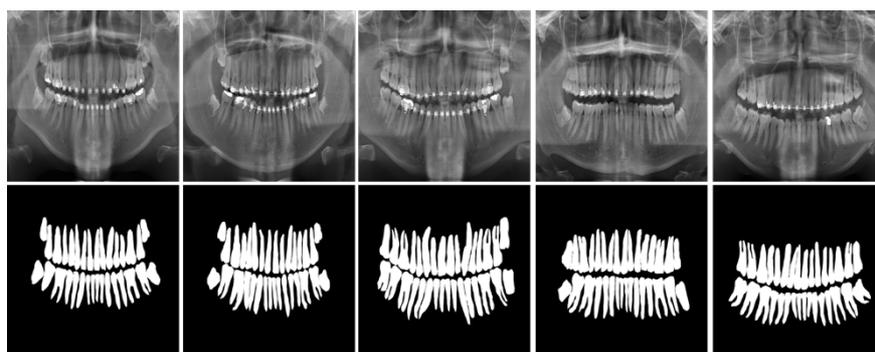


Figure 1. Examples from the dental panoramic radiography dataset(The top row is images, the bottom row is their ground truths.).

Before training and testing the model, all panoramic radiography images will be resized to 224 x 224 pixels, the input size required by the models in the dataset. In addition, the pixel values of the images will be normalized to the range [0, 1], facilitating faster and more effective learning. In addition, histogram equalization will be applied to increase the contrast of the teeth and simplify the segmentation process. Data augmentation techniques will improve the model's generalization ability. The methods to be used include random rotation (0° - 360°), horizontal and vertical flip, zooming, brightness adjustment, and noise addition. These augmentations will help the model become more robust to changing imaging conditions.

2.2. Proposed discriminator of the GAN

In this study, the segmentation models used as a Generator in the proposed generative adversarial network model were selected from the widely used models in the literature. The discriminator was designed most optimally by manually fine-tuning the hyperparameters. In the proposed study, five different segmentation models were used as Generators. The segmentation models are used as Generators in the GAN model. The models were designed as four layers from 32 to 512 depth, except for TransU-Net and SwinU-Net. TransU-Net and SwinUnet architectures

were reduced to four layers (32 to 256 depth) due to overfitting. The architectural structure of the GAN models, which consists of the generator and discriminator, is shown in Figure 2. 3 x 3 convolutional filters are used in all segmentation models. The filters of the model, the number of layers, the optimization algorithms to be used, and the activation function were determined as a result of 10-fold cross-validations to be performed on the dataset, and the random search and grid search algorithms ran on the designed model.

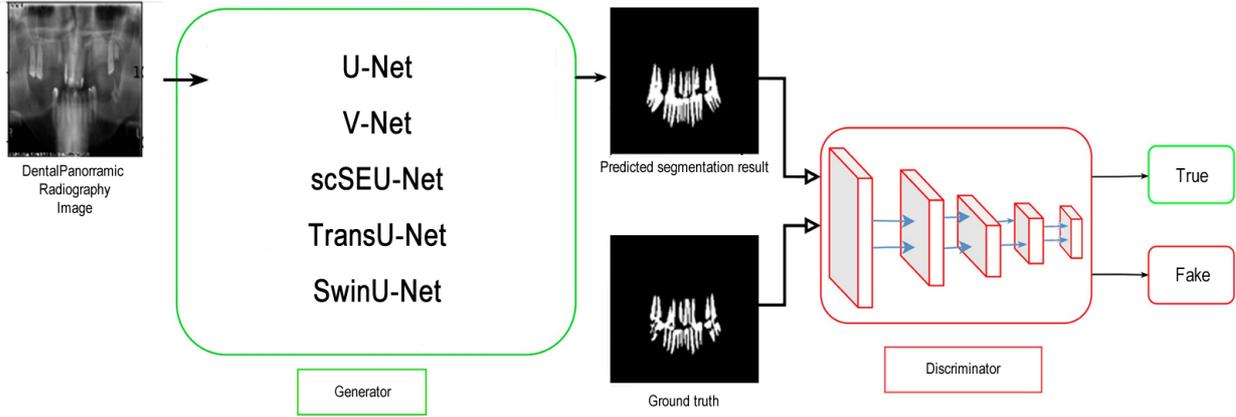


Figure 2. The GAN architecture used in the comparative analysis.

Discriminator: The discriminator section is designed to consist of 3 x 3 filters consisting of 5 layers from 32 channels to 512 channels. LeakyRelu is used as the activation function in the discriminator layer. 2-dimensional Global average pooling is added to the output of the layer. Dropouts with a coefficient of 0.1 and Dense(1) for binary classification are added to the output of the Global Average Pooling layer, respectively. The proposed Discriminator model is shown in Figure 3.

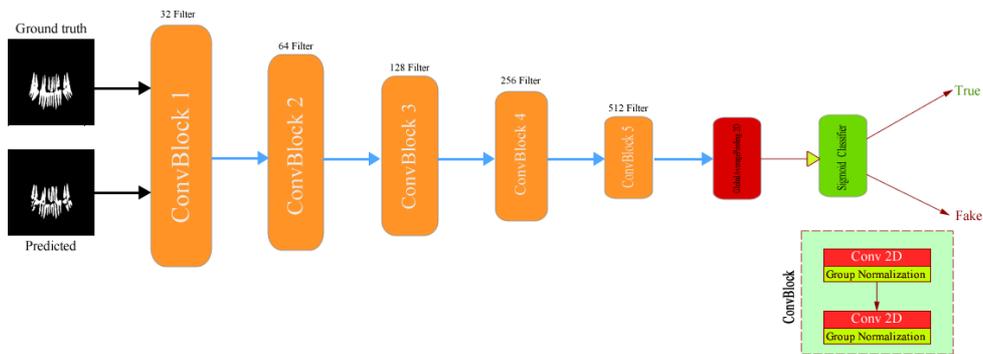


Figure 3. The architectural structure of the Discriminator model.

2.3 Generator models

U-Net: The U-Net architecture proposed by Ronneberger and his colleagues in the literature has been a groundbreaking study for later segmentation models with its success in medical image segmentation [18]. One of the most significant innovations in U-Net, which has a U-shaped architecture, is the skip-connections. The skip-connections ensure that the high-level features and spatial information lost in the pooling process in the encoder layer are fed to the decoder layer and preserved. While the encoder layer consists of convolutional blocks, the decoder layer consists of transposed convolutional blocks. 3x3 convolutional filters and Group normalization

method were used in the U-Net model used in comparative analysis. The architectural structure of the U-Net model is shown in Figure 4.

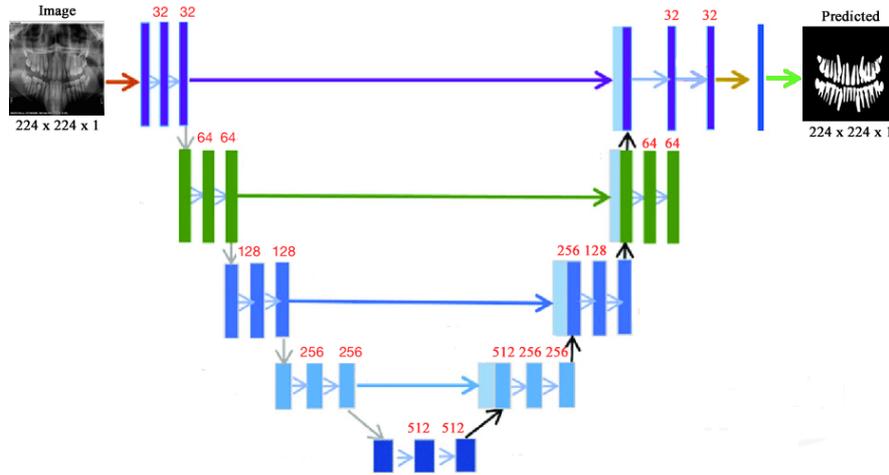


Figure 4. The architectural structure of the U-Net model.

V-Net: It is a deep learning model developed by Milletari et al., which uses convolutional layers instead of the pooling layer in the U-Net architecture and is proposed for robust volumetric segmentation of organs [28]. Fine-grained features are fed from the encoder layer to the decoder layer with the help of a skip-connection, thus preserving the edge information of the objects in the image, which is learned in the first layers and forgotten in the following layers. The V-Net segmentation model used as a generator in comparative analyses is shown in Figure 5.

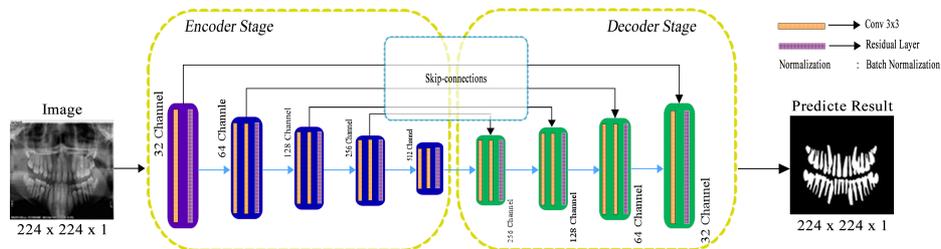


Figure 5. The architectural structure of the V-Net model.

scSEU-Net: It is a deep learning model obtained using the spatial and channel Squeeze and Excitation module proposed by Roy et al. in the decoder layer of the U-Net model [29]. Thanks to the SE blocks, higher segmentation success is achieved by focusing on the important regions in the image, just like in the Bottleneck attention module (BAM). Thanks to the scSE module, feature space and channel are calculated simultaneously. This way, it aims to emphasize high-level features, including dental regions, and suppress low-level features by recalibrating panoramic radiography images. The scSEU-Net model is shown in Figure 6.

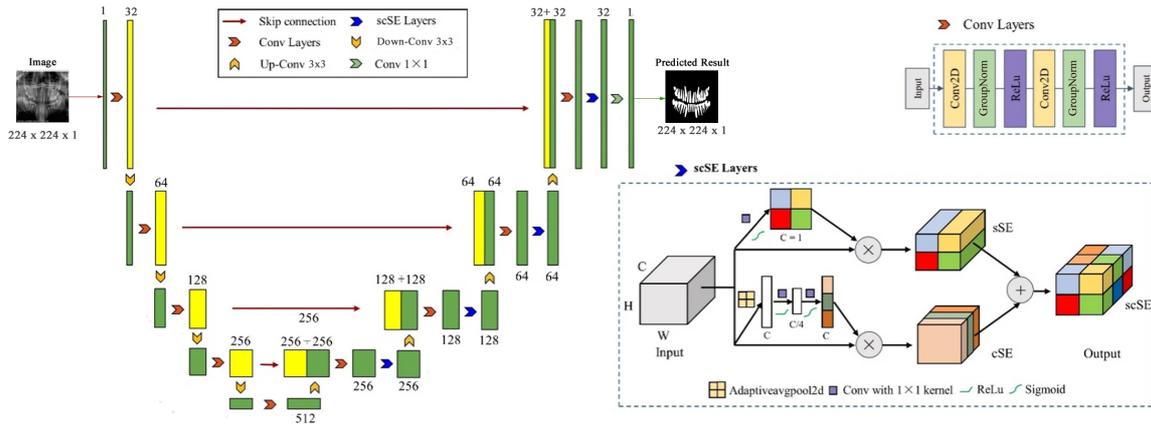


Figure 6. The architectural structure of the scSEU-Net model.

TransU-Net: Chen et al. proposed the TransU-Net model with self-attention modules to predict the contexts between pixels in an image [30]. To prevent the loss of feature resolution in high-resolution images due to the narrow receptive field network of transformers, both the success of CNN architectures in learning spatial information in high-resolution images and the high performance of Transformers in global context extraction are combined. Transformers in the TransU-Net architecture extract the contexts between features in an image with the help of self-attention modules and feed them to CNN layers to obtain high-level features in the image. This way, edge information, low-level features extracted in the first layers and forgotten in the following layers, is preserved, and a better segmentation result is obtained. In TransU-Net architecture, $embed_dim=512$, $num_mlp=256$, $num_heads=6$, and $num_transformer=6$ were determined as a result of the experimental studies. The scSEU-Net model is shown in Figure 7.

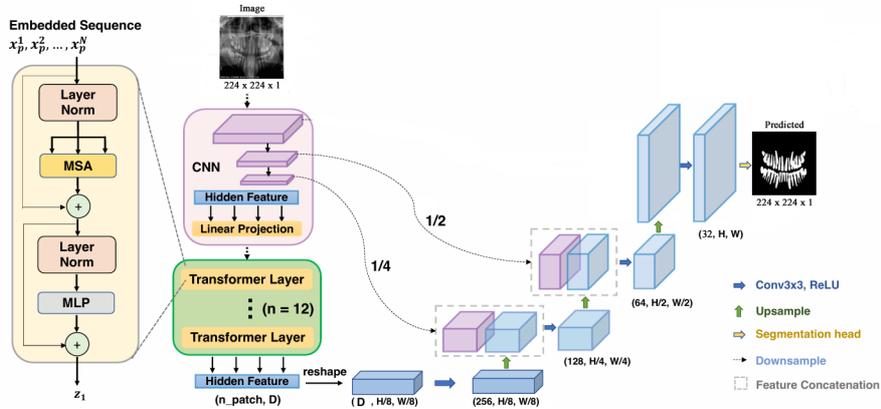


Figure 7. The architectural structure of the TransU-Net model.

SwinU-Net: The architecture proposed by Cao et al. uses a Transformer-based U-shaped model [31]. The proposed methodology has tokenized image patches and skip-connections where local global features from the encoder layer are fed to the decoder layer. Specifically, Swin Transformers with shifted windows are used in the encoder layer to provide context-related feature extraction. Additionally, researchers have reported that the model with four encoder and decoder layers each has higher performance in organ segmentation than pure transformer-based architecture and traditional CNN architectures. In the SwinU-Net architecture, as a result of the experimental studies, $patch_size=(2, 2)$, $num_heads=[4, 8, 8, 8]$, $window_size=[4, 2, 2, 2]$, $num_mlp=128$ were determined. The architectural structure of the proposed model is shown in Figure 8.

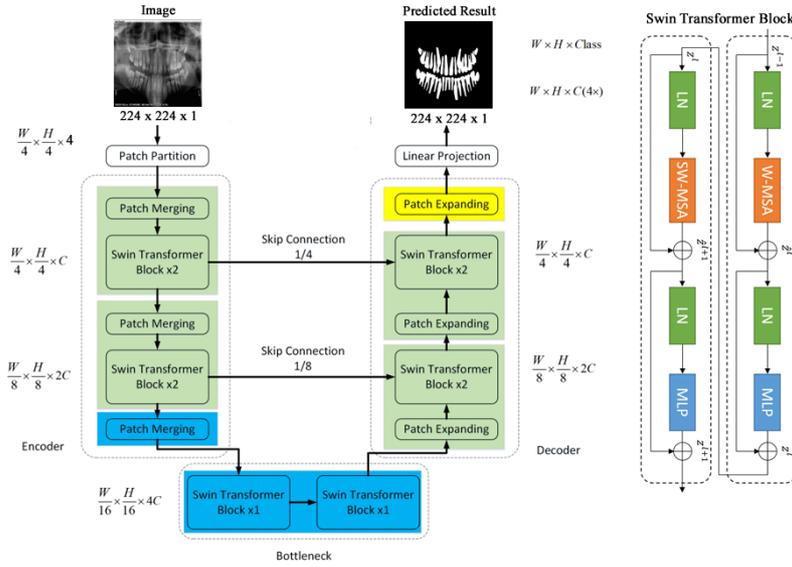


Figure 8. The architectural structure of the SwinU-Net model.

Training details of the models: In this study, comparative analyses of U-Net-based GAN, V-Net-based GAN, scSEU-Net-based GAN, TransU-Net-based GAN, and SwinU-Net based GAN models were performed on Adult Dental Panoramic Radiography dataset. Images were fed to the models in $224 \times 224 \times 1$ dimensions. Models were trained for 7500 epochs. By using the early stopping parameter in the tensorflow-keras 2.12 library based on Python 3.8, which was created in the anaconda ecosystem used for training the models, training was stopped if the validation loss of the models did not progress more than five epochs. The training of SwinU-Net and TransU-Net generator models stopped because there was no progress in validation loss after 3500 epochs. As a result of the experimental studies, it was seen that 3550 and 7500 epochs were sufficient for separate training of all generator and discriminator models. Learning_rate=0.0001, weight_decay=6e-8, activation_functon=ReLU, and batch_size=16 were selected for the generator part of the models. For the discriminator, learning_rate=0.0001 and batch_size=16 were used. The training of the GAN models was determined as 160 epochs. However, the training of scSEU-Net-based GAN 50 epochs, V-Net-based GAN 56 epochs, and TransU-Net-based GAN models stopped after 75 epochs because there was no progress in the validation loss. Adam optimizer was used in the proposed models [32]. A hardware structure consisting of a Tesla P100 graphics card, 64 GB RAM, and an Intel i5-8500H processor running on the Windows operating system was used for comparative analysis. **Loss Functions:** While the BinaryFocalCrossentropy loss function is used in the Generator modules of Gan architectures, the Mean Square Error (MSE) loss function is used in the Discriminator section. The mathematical equations of the loss functions are seen in Equation 1 and Equation 2.

$$\frac{-1}{N} \sum_{i=1}^N t_i \cdot \log(p(t_i)) + (1 - t_i) \cdot \log(1 - p(t_i)) \quad (1)$$

$$\text{MSE} = \sum (y_i - \pi_i)^2 n \quad (2)$$

In Equation 1, $p(t_i)$ is the probability of one, $1-p(t_i)$ is the probability of zero. In Equation 2, y_i is the i th observed value, π_i is the corresponding predicted value for y_i , and n is the number of observations. Σ indicates that a summation is performed over all values of i .

Performance metrics: Various metric-based algorithms will be used to determine the segmentation performance of the model. The performance metrics in Equations 3, 4, 5, 6, and 7 were used for training and testing the model on publicly available datasets. The most commonly used metrics in the literature for testing two-dimensional images are the Dice(DSC) and Intersection over Union (IoU-Jaccard) metrics. In addition, a threshold of 0.3 was added to these two metrics. Threshold aims to get a more robust model that undercuts a prediction score

of 0.3 to 0 in metrics. Thresholded Dice and Thresholded IoU metrics are shown in Equation 3 and Equation 4, respectively. These metrics were imported from a Python library called MedPy. In Equations 3 and 4, TP stands for True Positive, FP stands for False Positive, and FN stands for False Negative. The mathematical equation of the Accuracy (ACC) metric is seen in Equation 5. The Recall (REC) metric in Equation 6 measures our model's correct identification of True Positives. The performance metric Precision (PREC) given in Equation 7 is the ratio between TP and all positives.

$$\text{Thresh-DSC} = \frac{2*TP}{2*TP+FP+FN} > 0.3 \quad (3)$$

$$\text{Thresh-IoU} = \frac{TP}{TP+FN+FP} > 0.3 \quad (4)$$

$$\text{ACC} = \frac{TN+TP}{TN+TP+FN+FP} \quad (5)$$

$$\text{REC} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{PREC} = \frac{TP}{TP+FP} \quad (7)$$

3. Experimental Results and Discussion

This study performs GAN-based segmentation of dental images in the panoramic radiography dataset. Various data augmentation techniques are applied to the dataset, and 7,000 images and 28 thousand augmented data are obtained. In the generator part of the GAN models, the models' segmentation performances obtained using high-impact architectures in the literature, such as U-Net, V-Net, scSEU-Net, SwinU-Net, and TransU-Net, are evaluated. The obtained prediction results are analyzed comparatively with performance metrics widely used in the literature, such as Thresh-DSC, Thresh-IoU, Acc, Prec, and Recall. The quantitative analysis results are shown in Table 1. The quantitative analyses show that scSEU-Net-based GAN architecture obtained the highest prediction results.

Table 1. Quantitative performance analysis results of models.

Dataset	Model[Method]	Prm(M)	Thresh-DSC	Thresh-IoU	ACC	PREC	REC
Dental Panoramic Radiography Dataset	U-Net-based GAN ^[18]	5.4	0.8512	0.7413	0.9723	0.8904	0.8630
	V-Net-based GAN ^[28]	58.6	0.8434	0.7160	0.9740	0.8885	0.8835
	TransU-Net-based GAN ^[30]	42.3	0.8656	0.7633	0.9761	0.9268	0.8848
	SwinU-Net-based GAN ^[31]	2.4	0.7031	0.5183	0.9437	0.8098	0.6622
	scSEU-Net-based GAN ^[29]	22.3	0.8826	0.7901	0.9805	0.9268	0.9001

As can be seen from Table 1, the best result was obtained by the scSEU-Net-based GAN architecture. The second-best result was obtained by the TransU-Net architecture. The worst segmentation scores were obtained by the SwinU-Net architecture. This situation is thought to be due to the difficulty in learning local and global features in the image due to the narrowness of the receptive field network in the transformers. TransU-Net and SwinU-Net architectures were designed as four layers with 32 to 256 filters to cope with the overfitting. In addition, the multi-layer perceptrons in the self-attention modules were set to 256.

Figure 9 shows the qualitative analyses of the models whose comparative analyses were performed in the study. As can be understood from the images, scSEU-Net-based GAN compressed the features both channel and spatially. It brought the essential features to the forefront, achieving higher performance values compared to other models. Then, the best result was obtained by TransU-Net-based GAN architecture. U-Net-based GAN and V-Net-based GAN architectures, which achieved high performance in volumetric image segmentation, achieved similar results. SwinU-Net-based GAN, on the other hand, showed poor performance.

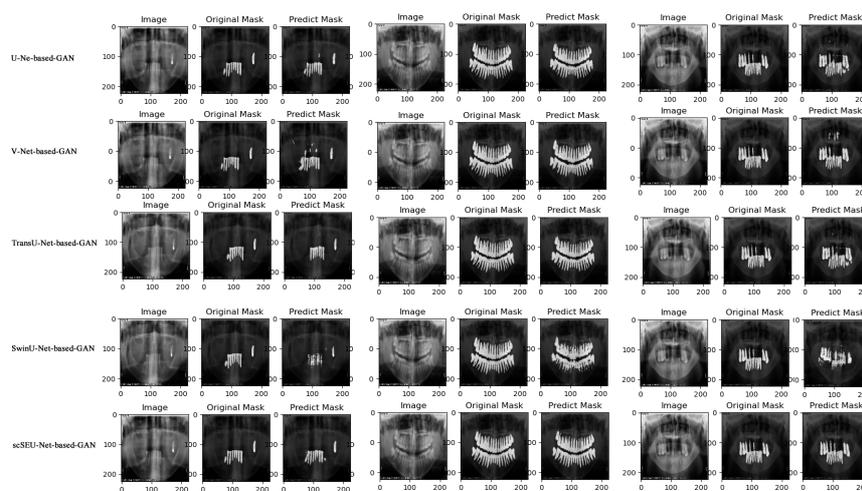


Figure 9. Examples of Qualitative analysis results of the GAN models used in my study on the Dental Panoramic Radiography dataset.

4. Conclusions

In this study, U-Net, V-Net, scSEU-Net, TransU-Net, and SwinU-Net architectures, which have achieved high performance in the segmentation of medical images in the literature, were used as generators in GAN models, and the segmentation results were analyzed comparatively. Comparative analyses of GAN models were performed using Thresh-DSC, Thresh-IoU, ACC, PREC, and REC performance metrics. For the Generator models, the BinaryFocalCrossentropy loss function was used, while the Discriminator model used the MSE loss function. In the comparative analysis, the highest performance result was obtained by the scSEU-Net-based GAN architecture. The TransU-Net-based GAN architecture obtained the second-highest performance result. However, while the TransU-Net-based GAN architecture achieved these performance results, it used almost twice as many parameters as the scSEU-Net-based GAN architecture. The high resource usage of transformer-based architectures is still a significant challenge. Future studies will investigate ways to achieve high performance with low parameters in Transformer-based GAN architectures.

References

- [1] Arsiwala-Scheppach LT, Chaurasia A, Mueller A, Krois J, Schwendicke F. Machine learning in dentistry: a scoping review. *J Clin Med* 2023; 12(3), 937.
- [2] Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, Schwendicke F. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep* 2019; 9(1), 8495.
- [3] Ekert T, Krois J, Meinhold L, Elhennawy K, Emar R, Golla T, Schwendicke F. Deep learning for the radiographic detection of apical lesions. *J Endod* 2019; 45(7), 917-922.
- [4] Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018; 77, 106-111.

- [5] Schwendicke FA, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 2020; 99(7), 769-774.
- [6] Mateusz B, Atsuto M, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural nets* 2018; 106, 249-259.
- [7] Bhat S, Birajdar GK, Patil MD. A comprehensive survey of deep learning algorithms and applications in dental radiograph analysis. *Healthc Anal* 2023; 100282.
- [8] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2014; Ohio, USA, 580–587.
- [9] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE PAMI* 2015; 37, 1904–1916.
- [10] Girshick R. F R-CNN. *ICCV* 2015; Santiago, Chile, 1440–1448.
- [11] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *NIPS* 2015; Montreal, Canada, 91–99.
- [12] Szegedy C, Iofe S, Vanhoucke V. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *AAAI* 12 2017; San Fransisco, USA, 1602-1625.
- [13] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *CVPR IEEE* 2016; Las Vegas, USA, 770–778.
- [14] Xu X, Liu C, Zheng Y. 3D tooth segmentation and labeling using deep convolutional neural networks. *IEEE Trans Vis Comput Graph* 2018; 25(7):2336-2348.
- [15] Tian S, Dai N, Zhang B, Yuan F, Yu Q, Cheng X. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. *IEEE Access* 2019; 7, 84817-84828.
- [16] Sampath V, Maurtua I, Aguilar Martin JJ, Gutierrez A. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data* 2021; 8, 1-59.
- [17] Fatima A, Shafi I, Afzal H, Mahmood K, Diez IDLT, Lipari V, Ashraf I. Deep learning-based multiclass instance segmentation for dental lesion detection. *Healthcare MDPI* 2023, Basel, Balgium, Vol. 11, No. 3, p. 347.
- [18] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *MICCAI* 2015; Munich, Germany 234-241.
- [19] Tyas MJ, Anusavice KJ, Frencken JE, Mount GJ. Minimal intervention dentistry—a review. *Int Dent J* 2000; 50(1), 1-12.
- [20] Featherstone JD. The caries balance: the basis for caries management by risk assessment. *Oral Health Prev Dent* 2004; 2, 259-264.
- [21] Li Z, Tang W, Gao S, Wang Y, Wang S. Adapting SAM2 Model from Natural Images for Tooth Segmentation in Dental Panoramic X-Ray Images. *Entropy* 2024; 26(12), 1059.
- [22] Şahin ME, Ulutaş H, Süzgen EE. Automated Segmentation of Dental Structures in Panoramic Radiographs Using U-Net 3+. *IDAP IEEE* 2024; Malatya, Türkiye, 1-6.
- [23] Xing Y, Liao P, Alasleh RA, Khampatee V, Alizadeh-Shabdiz F. Dental X-ray Segmentation and Auto Implant Design Based on Convolutional Neural Network. *MIPR IEEE* 2024; San Jose, CA, USA, 243-246.
- [24] Özçelik STA, Üzen H, Şengür A, Fırat H, Türkoğlu M, Çelebi A, Sobahi NM. Enhanced Panoramic Radiograph-Based Tooth Segmentation and Identification Using an Attention Gate-Based Encoder–Decoder Network. *Diagnostics* 2024; 14(23), 2719.
- [25] Zhou W, Lu X, Zhao D, Jiang M, Fan L, Zhang W, Liu X. A dual-labeled dataset and fusion model for automatic teeth segmentation, numbering, and state assessment on panoramic radiographs. *BMC Oral Health* 2024; 24(1), 1201.
- [26] Kong V, Lee EY, Kim KA, Shon HS. Integrating Super-Resolution with Deep Learning for Enhanced Periodontal Bone Loss Segmentation in Panoramic Radiographs. *Bioeng.* 2024; 11(11), 1130, (2024).
- [27] Altundağ G, Öcal H. A Comparison of shcU-Net Based GAN and U-net Based GAN in Adult Dental Segmentation. *UBMK IEEE* 2024; Antalya, Türkiye, 1040-1045.
- [28] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *3DV IEEE* 2016; Stanford University, CA, USA; 565-571).
- [29] Roy AG, Navab N, Wachinger C. Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks. *IEEE Trans Med Imaging* 2018; 38(2), 540-549.
- [30] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* 2021.
- [31] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. Swin-unet: Unet-like pure transformer for medical image segmentation. *ECCV* 2022; Switzerland, 205-218.
- [32] Ba J, Kingma P. Adam: a method for stochastic optimization. *ICLR* 2015; San Diego, USA, 1–11.