НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.8

ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS FOR TOOTH SEGMENTATION ON PANORAMIC X-RAY IMAGES

Bohush V. I. – Master of the Department of Software Engineering and Information Technologies, Oles Honchar Dnipro National University, Dnipro, Ukraine. ROR: https://ror.org/00qk1f078. ORCID: https://orcid.org/0009-0007-0986-3653.

Ivanchenko M. G. – PhD, Associate Professor, Associate Professor of the Department of Software Engineering and Information Technologies, Oles Honchar Dnipro National University, Dnipro, Ukraine. ROR: https://ror.org/00qk1f078. ORCID: https://orcid.org/0000-0001-7795-0459.

ABSTRACT

Context. Semantic segmentation of teeth on panoramic X-ray images is an important task in dental diagnostics, as it allows for the automation of the diagnosis of dental diseases. However, panoramic X-rays have a complex structure, which complicates the task of segmentation. The use of convolutional neural networks shows high potential in solving this problem. In this context, it is relevant to study the best models and combine them into an ensemble in order to improve the quality of segmentation.

Objective. The aim of this work is to study the effectiveness of various convolutional neural network architectures in the task of semantic tooth segmentation on panoramic X-ray images and to develop an ensemble approach to improve the quality of the results.

Method. Various architectures of convolutional neural networks are used: U-Net, Attention U-Net, Residual U-Net, Residual Attention U-Net, R2 U-Net, U-Net++, U-Net 3+, USE-Net, Dense U-Net, and DeepLabV3+ with a pre-trained ResNet-101 backbone on the ImageNet dataset. An ensemble approach based on the best models is proposed, where the final segmentation mask is determined by majority voting. The models were trained on a preprocessed dataset of panoramic X-ray images with the application of augmentation techniques. The performance of the models was evaluated using the IoU, Dice, and Accuracy metrics.

Results. Various neural network models were investigated, and the best ones were combined into an ensemble. The conducted experiments confirmed that the ensemble approach improves segmentation accuracy compared to individual models. The best result was achieved by the ensemble combining the Dense U-Net, Attention U-Net, and U-Net 3+ architectures.

Conclusions. The proposed ensemble approach demonstrated high efficiency in the task of semantic tooth segmentation on panoramic X-ray images, outperforming the results of individual models. The scientific novelty of the study lies in the application of an ensemble approach that combines various architectures of convolutional neural networks for semantic tooth segmentation on X-ray images. The practical significance of the work is in the potential use of the developed approach for building automated diagnostic systems in dentistry. The obtained results can be applied to further automate the analysis of X-ray images and contribute to the development of intelligent medical systems.

KEYWORDS: deep learning, convolutional neural networks, tooth segmentation, semantic segmentation, U-Net, ensemble.

ABBREVIATIONS

BCE – Binary Cross-Entropy;
CNN – Convolutional Neural Network;
CUDA – Compute Unified Device Architecture;
FN – False Negative;
FP – False Positive;
GB – Gigabyte;
GPU – Graphics Processing Unit;
IoU – Intersection over Union;
ReLU – Rectified Linear Unit;
TN – True Negative;
TP – True Positive.

NOMENCLATURE

Accuracy is a proportion of correctly classified pixels out of the total number of pixels in the image;

C is a number of channels of the input image;

D is a full dataset;

Dice is a measure of overlap between the predicted mask and the ground-truth mask;

D_{train} is a training dataset;

D_{test} is a test dataset;

E is a number of models in the ensemble;

F is a neural network model;

F_m is a m-th model among all examined models;

FN is a proportion of class 1 (tooth) pixels incorrectly classified as class 0 (background);

DOI 10.15588/1607-3274-2026-2-6

FP is a proportion of class 0 (background) pixels incorrectly classified as class 1 (tooth);

G_{ens} is an ensemble containing E models;

G_e is an e-th model of the ensemble;

H is a height of the input image;

IoU is a ratio of the intersection area to the union area of the predicted and ground-truth masks;

L_{BCE} is a binary cross-entropy loss function;

L_{Dice} is a Dice-based loss function;

L is a loss function defined as the sum of L_{BCE} and L_{DICE};

M is a number of models in the study;

Metric is a function that denotes a quality metric;

N is a total number of images in the dataset;

N_{test} is a number of images in the test dataset;

O is a predicted binary segmentation mask;

TN is a proportion of pixels correctly classified as class 0 (background);

TP is a proportion of pixels correctly classified as class 1 (tooth);

U is a threshold function that returns 1 if the value is greater than 0.5, otherwise 0;

W is a width of the input image;

X is an input image;

Y is a binary mask with ground-truth values;

$\hat{Y}$ is a predicted mask containing probabilities of belonging to the target class 1 (tooth);

$\hat{Y}^e$ is a predicted probabilistic mask for the e-th model in the ensemble;

$\hat{Y}_{i,j}$ is a predicted probability of the target class for the pixel at position (i, j);

y_i is a ground-truth value of the i-th pixel;

$\hat{y}_i$ is a predicted probability of the i-th pixel;

ε is a small number to avoid division by zero;

Θ is a set of parameters of all models in the ensemble;

θ is a set of parameters of the neural network model;

θ_m is a set of parameters of the m-th model.

INTRODUCTION

In dentistry, X-ray images are actively used for the diagnosis and treatment of patients. The analysis of X-ray images is a complex and labor-intensive task for a dentist. One of the key tasks is the accurate segmentation of teeth on panoramic X-ray images, which aids in the detection of dental diseases.

In recent years, many approaches using neural networks have been proposed for solving the segmentation task in the medical field [1]. In particular, convolutional neural networks have shown high efficiency in the field of dentistry [2]. However, despite a significant number of studies in this area, most of them are focused on using single neural network architectures, which do not always provide the best results. This creates a need for the development of approaches that would enhance the reliability and accuracy of segmentation by combining the advantages of different models.

The relevance of the study is driven by the need to increase the accuracy and speed of diagnosing dental diseases, as well as the development of methods for computer analysis of medical images.

The object of the study is the process of semantic segmentation of teeth on panoramic X-ray images.

The subject of the study is convolutional neural network architectures and an ensemble approach for the semantic segmentation of teeth on panoramic X-ray images.

The purpose of the work is to investigate the effectiveness of various convolutional neural network architectures and to develop an ensemble approach for the semantic segmentation of teeth on panoramic X-ray images. To achieve this aim, it is necessary to investigate the effectiveness of modern convolutional neural network architectures and develop an ensemble model that combines several effective architectures. Furthermore, it is required to evaluate the segmentation accuracy using appropriate metrics and perform a comparative analysis of the obtained results.

1 PROBLEM STATEMENT

Let $D = \{(X_i, Y_i) \mid i = 1, \ldots, N\}$ be a set of panoramic X-ray images and their corresponding segmentation masks. The task is to find a function (model) $F: X \to \hat{Y}$ that maps an input image X to a predicted segmentation mask $\hat{Y} \in [0,1]^{H \times W \times C}$, which contains the probabilities of each pixel belonging to the target class 1 (the tooth class). To obtain a binary mask $O \in \{0, 1\}^{H \times W \times C}$, where 0 denotes the background and 1 denotes a tooth, a threshold function must be applied to each value $\hat{Y}_{i,j}$ according to the formula (1):

$$U(Y_{i,j}) = \begin{cases} 1, & \text{if} \quad Y_{i,j} > 0.5; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

It is necessary to train M models from the set $\{F_m \mid m = 1, \ldots, M\}$, which have different architectures. For each model F_m, it is required to find the optimal set of parameters θ_m that minimizes the loss function L on the training set D_{train} (2):

$$\theta_m = \arg\min_{\theta_m} \sum_{(X_i, Y_i) \in D_{train}} L(F_m(X_i; \theta_m), Y_i). \tag{2}$$

From the set of models $\{F_m \mid m = 1, \ldots, M\}$, it is necessary to select the E best models for the ensemble $\{G_e \mid e = 1, \ldots, E\}$ based on their quality metric values on the test set D_{test}. Dice and IoU can be used as quality metrics.

As a result of applying the models from the ensemble G to an input image X, a set of predicted masks $\{\hat{Y}^e \mid e = 1, \ldots, E\}$ will be obtained. The final binary mask for the ensemble of models is determined by the majority voting method according to the formula (3):

$$O_{i,j} = \begin{cases} 1, & \text{if} \quad \sum_{e=1}^{E} U(\hat{Y}_{i,j}^e) > \frac{E}{2}; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

DOI 10.15588/1607-3274-2026-2-6

The main task is to find the optimal parameters of the ensemble $\Theta = \{\theta_1, \ldots, \theta_E\}$ that maximize the chosen quality metric on the test dataset (4):

$$\Theta = \arg\max_{\Theta} \frac{1}{N_{test}} \sum_{(X_i, Y_i) \in D_{test}} Metric(G_{ens}(X_i; \Theta), Y_i). \quad (4)$$

2 REVIEW OF THE LITERATURE

In recent years, many studies have been conducted applying various approaches to solve the task of segmentation in dentistry using neural networks. In the task of semantic segmentation, the U-Net architecture [3] has proven to be particularly effective. Since the creation of U-Net, many modifications of this architecture have been proposed, including Attention U-Net [4], Residual U-Net [5], Residual Attention U-Net [6], R2 U-Net [7], USE-Net [8], Dense U-Net [9], U-Net++ [10], and U-Net 3+ [11]. For the segmentation of objects at different scales, the DeepLabV3+ architecture [12] has shown high results.

In the paper [13], convolutional neural networks based on the U-Net architecture were investigated for solving the task of semantic segmentation of teeth on panoramic X-ray images. The dataset contained 1500 images. As a result, the Dice coefficient reached 0.934. The authors also used an ensemble of several models, which increased the Dice coefficient value to 0.936.

In the paper [14], a Multi-Scale Location Perception Network (MSLPNet) was used for tooth segmentation. This model focused on improving the segmentation quality of teeth with fuzzy root boundaries. The dataset contained 1500 panoramic X-ray images. As a result, the Dice coefficient reached 0.9301.

In the paper [15], a Two-Stage Attention Segmentation Network (TSASNet) on dental panoramic X-ray images was proposed to address the problems arising when performing the task of tooth boundary and tooth root segmentation. The dataset contained 1500 panoramic X-ray images. As a result, the Dice coefficient reached 0.9272.

The paper [16] is notable for describing the creation of a Knowledge Consistency Neural network (KCNet) using knowledge distillation, which allows for obtaining lightweight models for the purpose of deployment on edge devices. In this paper, they propose an attempt to obtain reliable data from a teacher network using a knowledge network. The dataset contained 1321 panoramic dental images. They chose U-Net and ESPNet-v2 as the student and teacher networks, respectively. As a result, the Dice coefficient was 0.89. The KCNet model was compared with others, where although it has slightly worse quality, it requires significantly fewer parameters.

In the paper [17], the SWin-U-Net model was used for tooth segmentation on panoramic X-ray images. SWin-U-Net is an encoder-decoder system that uses a transformer and has a U-shape with skip connections. In SWin-U-Net, a symmetric encoder-decoder structure is built using jump connections. It uses a local to a global strategy for self-attention. A patch-expanding layer is also created to increase sampling and feature dimension without using convolution or interpolation methods. The dataset contained 100 panoramic X-ray images. As a result, the values for the F1 (Dice), IoU, and Accuracy metrics were 0.6372, 0.4689, and 0.8852, respectively. This model was compared with U-Net, Link-Net, and FPN models, where SWin-U-Net outperforms all other models.

In the paper [18], the Teeth U-Net model is presented for tooth segmentation on panoramic X-ray images. This model uses a Squeeze-Excitation Module in the encoder, and a dense skip connection between the encoder and decoder is proposed to reduce the semantic gap. To solve problems related to irregular tooth shape and low image contrast, a Multi-scale Aggregation attention Block (MAB) is applied in the bottleneck layer, which can effectively extract tooth shape features and adaptively fuse multi-scale features. A Dilated Hybrid self-Attentive Block (DHAB) is also formulated to capture dental feature information across a broader perceptual field. The dataset included 1500 panoramic X-ray images. The proposed model demonstrates high results in three comparative experiments, showing high metrics for Accuracy, Precision, Recall, and Dice, which are 0.9853, 0.9562, 0.9451, and 0.9428, respectively.

The paper [19] compares varieties of U-Net models for tooth segmentation on panoramic X-ray images, namely the following models: U-Net, Attention U-Net, Dense U-Net, R2 U-Net, Residual U-Net, SE U-Net (USE-Net). In these variants, configurations with two and three convolutional layers are used in both the encoder and decoder blocks. The dataset contained 389 images. For models with two convolutional layers, U-Net demonstrated the highest results, achieving a Dice coefficient and an IoU score of 0.8911 and 0.88, respectively. For models with three convolutional layers, the Dense U-Net model proved to be the best, achieving a Dice coefficient and an IoU score of 0.9033 and 0.8907, respectively. It was found that for each of the models, the quality metrics were higher with three convolutional layers.

The authors of the paper [20] presented the CariSeg neural network ensemble, which contains four models. The first model in the CariSeg system, trained using the U-Net architecture, segments the region with teeth in order to subsequently crop the panoramic X-ray image. The next component segments carious lesions and is an ensemble containing three models with different architectures: U-Net, Feature Pyramid Network, and DeeplabV3. For tooth segmentation, the dataset included 1116 panoramic images, and for caries segmentation – 150 images, from which only 108 were selected. As a result, for tooth segmentation, the Dice coefficient and IoU score were 0.889 and 0.798, respectively, and for caries segmentation, the Dice coefficient and IoU score were 0.6820 and 0.5180, respectively.

The authors of the paper [21] proposed the S2AgScUNet model to solve the task of semantic segmentation of teeth, which uses a pre-trained SAM2 model as an encoder combined with ScConv modules and gated attention mechanisms. To simplify deployment on edge

DOI 10.15588/1607-3274-2026-2-6

devices, they created the LightUNet model using knowledge distillation to achieve quality comparable to the U-Net model. In the study, the proposed models were compared with U-Net, SegFormer, MaNet, and CeNet models. The dataset contained 1500 panoramic X-ray images, from which 425 images were selected for the study. As a result, the S2AgScUNet model showed the highest result with a Dice coefficient and an IoU score of 0.9254 and 0.8612, respectively. The LightUNet model achieved a Dice coefficient and an IoU score of 0.9156 and 0.8443, respectively, which does not differ much from the results of the U-Net model.

In the paper [22], the authors used U-Net, U-Net++, U-Net with MobileNetV3 as an encoder, and DeepLabV3 architectures for the semantic segmentation of teeth on panoramic X-ray images. The U-Net and U-Net++ models contained ResNet 50 as the base model. The study was conducted on two datasets, where the first contained 1000 images, and the second – 2500 images. On the first dataset, the Dice coefficient for U-Net, U-Net++, U-Net-MobileNetV2, and DeepLabV3 models was 0.9121, 0.9014, 0.8761, and 0.6412, respectively. On the second dataset, the Dice coefficient for U-Net and U-Net++ models was 0.9253 and 0.9258, respectively.

The authors of the paper [23] proposed a dual-stream dental segmentation model based on Transformer heterogeneous feature complementarity for tooth segmentation on panoramic X-ray images. This parallel architecture includes a Transformer semantic parsing branch and a convolutional neural network detail capturing pathway, achieving collaborative optimization of global context modeling and local feature extraction. The dataset contained 1500 images. Experimental results showed that the proposed method achieved a Dice coefficient and an IoU score of 0.9454 and 0.9149, respectively. The results were compared with U-Net, U-Net++, Attention U-Net, CE-Net, and GT U-Net models, where the proposed method showed the highest result.

Based on the literature review conducted, it can be concluded that the application of neural networks for tooth segmentation on X-ray images is a promising direction in modern dentistry. Despite significant achievements, most works in this field focus on using single architectures, whereas the results demonstrate considerable variability depending on the model structure, sample size, and image characteristics. In this regard, there is a growing interest in ensemble methods, which combine the results of several models to obtain a generalized prediction.

Thus, the analysis indicates that, despite significant advancements in the field of medical image segmentation, there remains a need to investigate the effectiveness of ensemble approaches specifically for tooth segmentation on panoramic X-ray images. This opens the prospect of creating more reliable and accurate systems for supporting dental diagnostics.

3 MATERIALS AND METHODS

In this study, ten convolutional neural network architectures are considered: U-Net, Attention U-Net, Residual U-Net, Residual Attention U-Net, R2 U-Net, USE-Net, Dense U-Net, U-Net++, U-Net 3+, and DeepLabV3+ (with the ResNet-101 model, pre-trained on the ImageNet dataset, as the backbone network). The choice of these models is determined by their proven effectiveness in the task of medical semantic segmentation.

U-Net has a symmetric structure consisting of an encoder and a decoder. The encoder gradually reduces the spatial dimensions of the image, extracting high-level features, while the decoder restores the spatial dimension of the image using transposed convolution operations. A characteristic feature of U-Net is the presence of skip connections between the corresponding levels of the encoder and decoder, which allows combining contextual and spatial information for more accurate segmentation. Fig. 1 illustrates the U-Net architecture.

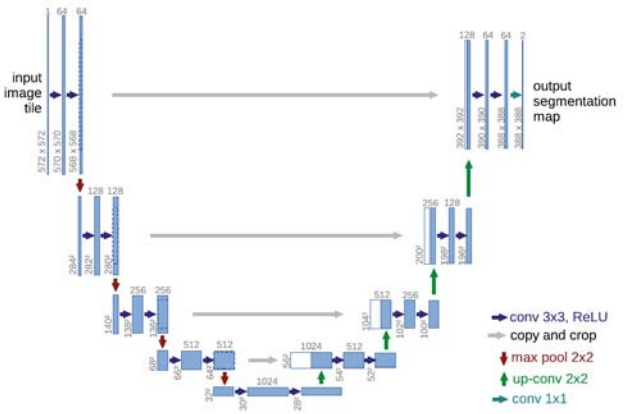

Figure 1 – U-Net architecture [3]

Other architectures modify U-Net through various mechanisms:

1. Attention U-Net uses attention gates, which automatically identify the most informative regions of the image while suppressing irrelevant signals.
2. Residual U-Net extends the base architecture with residual connections, which facilitate the propagation of gradients during training. This contributes to more stable training of very deep models.
3. Residual Attention U-Net combines the advantages of residual blocks and attention mechanisms, allowing the model to simultaneously maintain depth and focus on the most important image features.
4. R2 U-Net integrates recurrent convolutional blocks and residual connections, which allows for the consideration of spatial-contextual information within the image.
5. USE-Net (U-Net with Squeeze-and-Excitation blocks) performs adaptive weighting of feature maps, enabling the model to amplify significant features and suppress less relevant ones.
6. Dense U-Net uses dense blocks. In these blocks, each layer receives feature maps from all previous layers within the block as input. This encourages feature reuse, improves gradient flow, and allows for high performance with fewer parameters.

DOI 10.15588/1607-3274-2026-2-6

7. U-Net++ adds intermediate nodes between the encoder and decoder, forming a cascaded system of skip connections. This ensures deeper integration of features from different levels and contributes to more accurate object contour reconstruction. U-Net 3+ uses dense multi-level connections between the encoder and decoder, ensuring the aggregation of information from different scales.

8. U-Net 3+ offers full-scale skip connections. Each decoder node receives input not only from the corresponding encoder level, but also from all encoder levels (smaller size) and all decoder levels (larger size). This ensures feature aggregation at absolutely all scales for generating an accurate segmentation mask.

DeepLabV3+ with the ResNet-101 backbone network applies dilated convolutions and an ASPP (Atrous Spatial Pyramid Pooling) module to extract features at various spatial scales. This approach allows the model to consider both local and global contextual dependencies during tooth segmentation.

In all architectures, ReLU is used as the activation function, which ensures fast and stable training. Additionally, for better training, batch normalization is used after the convolutional layers.

At the output of each model, the SoftMax function is used, which allows interpreting the values of the output mask as the probabilities of each pixel belonging to the target class.

The models were trained using a combined loss function of Binary Cross-Entropy (BCE) and Dice loss.

BCE evaluates the pixel-wise difference between the predicted mask and the true labels, ensuring stable weight updates in the early stages of training, but it can be ineffective in cases of class imbalance. BCE is defined by the formula (5):

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]. \quad (5)$$

The Dice loss function evaluates the similarity between the predicted and true masks, which is particularly important in cases of class imbalance, where the number of pixels belonging to the target class is much smaller than the number of background pixels. This function is defined by the formula (6):

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \varepsilon}, \quad (6)$$

where ε is a small number to avoid division by zero.

The total loss function is defined as a linear combination of (5) and (6) according to the following formula (7):

$$L = L_{BCE} + L_{Dice}. \quad (7)$$

The combination of these loss functions allows for the simultaneous consideration of both pixel-wise error and mask similarity, which ensures more effective training.

After training and evaluating all models, the three best models are selected based on the IoU and Dice metrics. An ensemble is constructed based on these models, which forms the final segmentation mask using the majority voting method. For each pixel, the number of models that classified it as class 1 (tooth) is calculated. A pixel is labeled as 1 (tooth) if two or more models from the ensemble predicted this class; otherwise, it is labeled as 0 (background). This approach allows for reducing the influence of errors from individual models and increasing the stability of the segmentation results.

4 EXPERIMENTS

The study was conducted on a public dataset [24], containing 1500 panoramic X-ray images with annotated tooth masks. As many images were poorly annotated, 892 images were selected. The dataset was randomly divided into three subsets, where 70% (624 images) was the training set, 15% (133 images) was the test set, and 15% (135 images) was the validation set.

All images were resized to a uniform size of 256x256 using bilinear interpolation (for images) and nearest-neighbor interpolation (for masks). The pixel intensity values were normalized to a normal distribution with a mean of 0.5 and a standard deviation of 0.5. All images were in grayscale format, i.e., with a single channel.

To enhance the generalization ability of the models and prevent overfitting, the following augmentation techniques were applied to the training set during training (all with a probability of 0.5):

– horizontal flips;
– rotations by 10°;
– scaling by 10%;
– shifting along the x and y axes by 5%;
– brightness and contrast adjustments by 20%.

The validation and test sets were used without augmentations; only resizing and normalization were applied.

For the data samples, a batch size of 16 was chosen for training all models, except for the Dense U-Net model, for which the batch size was 8. This was due to insufficient video memory on the graphics card used, but this is not critical, and during training, this model can still effectively reach a local minimum of the loss function.

All models were implemented, trained, and tested using the PyTorch framework in the Python programming language.

The experiments were conducted on a computer equipped with an NVIDIA GeForce 5090 graphics card with 32 GB of video memory, an AMD EPYC 7713 processor with 64 cores, and 32 GB of RAM. The use of this graphics card allowed for the utilization of CUDA technology, which significantly accelerated the training process.

To ensure a fair comparison, all models were trained with almost identical parameters. The loss function used was according to the formula (7) for all models. The

DOI 10.15588/1607-3274-2026-2-6

Adam optimizer was chosen with a learning rate of 0.001 and L2 regularization with a coefficient of 0.001. For all models, 100 epochs were set for training, and early stopping was applied if the loss function value on the validation set did not decrease for 15 epochs. The best weights were saved based on the lowest validation loss; therefore, even in the presence of overfitting, the weights from before overfitting were selected.

For the R2 U-Net model, which tends to overfit, dropout was applied to the feature maps with probabilities of 0.1 in the second encoder block, 0.2 in the third and fourth blocks, and 0.3 in the bottleneck. The DeepLabV3+ model uses dropout in the decoder in the convolutional layers when refining features with a probability of 0.1 after the first convolutional layer and a probability of 0.5 after the second convolutional layer. Besides, at the end of the ASPP module of this model, a dropout with a probability of 0.1 is used.

All U-Net-based models used two convolutional layers in a single block, except for Dense U-Net, which used four layers in the Dense Block, and R2 U-Net, which used a recurrent loop twice.

Three main metrics were used to evaluate the segmentation quality: Accuracy, IoU, Dice. The Accuracy metric does not take into account class imbalance, so the IoU and Dice metrics are taken into account more during the study.

The Accuracy metric is defined by the formula (8):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (8)$$

The IoU metric is defined by the formula (9):

$$IoU = \frac{TP}{TP + FP + FN}. \quad (9)$$

The Dice metric is defined by the formula (10):

$$Dice = \frac{2TP}{2TP + FP + FN}. \quad (10)$$

5 RESULTS

Fig. 2 shows the training curves for all models, displaying the loss function values for the training and validation sets. The number of epochs may vary for different models, as early stopping was applied.

As can be seen from the graph, there is no overfitting for all models, except for the R2 U-Net model, which exhibits slight overfitting, even though corresponding overfitting prevention techniques were applied to it. As a result, the models were saved over many epochs, and the best ones were selected.

Table 1 presents the quality metrics for all models on the training and test sets.

On both the training and test sets, the best quality metric values were shown by the Dense U-Net, U-Net 3+, and Attention U-Net models, which were combined into an ensemble. The best single model achieved a Dice coefficient of 0.9371 on the training set and 0.9341 on the test set. The ensemble exceeds the performance of the individual models, achieving a Dice coefficient of 0.9377 on the training set and 0.9351 on the test set, which confirms the appropriateness of using such an approach.

Fig. 3 shows an example of tooth segmentation for each model and the ensemble in comparison with the ground truth mask.

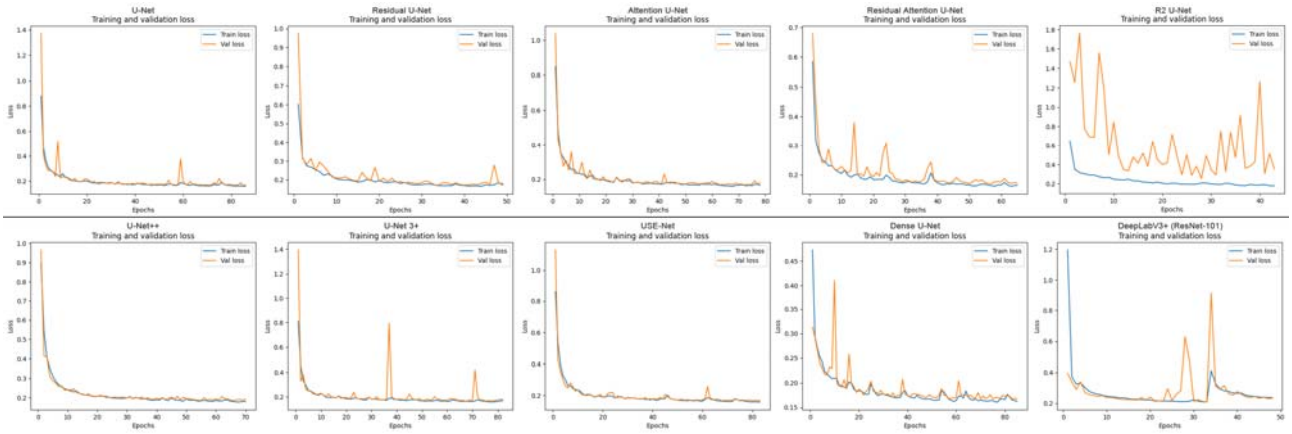

Figure 2 – Loss curves for model training

DOI 10.15588/1607-3274-2026-2-6

Table 1 – Performance metrics of all the models on the training and test datasets

Dataset	Model	Accuracy	IoU	Dice
Training dataset	U-Net	0.9712	0.8654	0.9276
	Attention U-Net	0.9736	0.8778	0.9347
	Residual U-Net	0.9729	0.8751	0.9332
	Residual Attention U-Net	0.9698	0.8616	0.9254
	R2 U-Net	0.9598	0.8184	0.8993
	U-Net++	0.9693	0.8580	0.9234
	U-Net 3+	0.9738	0.8793	0.9356
	USE-Net	0.9720	0.8714	0.9311
	Dense U-Net	0.9745	0.8820	0.9371
	DeepLabV3+ (ResNet-101 backbone)	0.9617	0.8312	0.9076
	Ensemble (Dense U-Net, U-Net3+, Attention U-Net)	0.9747	0.8830	0.9377
Test dataset	U-Net	0.9682	0.8608	0.9251
	Attention U-Net	0.9708	0.8725	0.9318
	Residual U-Net	0.9708	0.8732	0.9321
	Residual Attention U-Net	0.9694	0.8678	0.9291
	R2 U-Net	0.9585	0.8248	0.9036
	U-Net++	0.9659	0.8519	0.9199
	U-Net 3+	0.9715	0.8756	0.9335
	USE-Net	0.9691	0.8662	0.9282
	Dense U-Net	0.9717	0.8766	0.9341
	DeepLabV3+ (ResNet-101 backbone)	0.9590	0.8299	0.9069
	Ensemble (Dense U-Net, U-Net3+, Attention U-Net)	0.9721	0.8783	0.9351

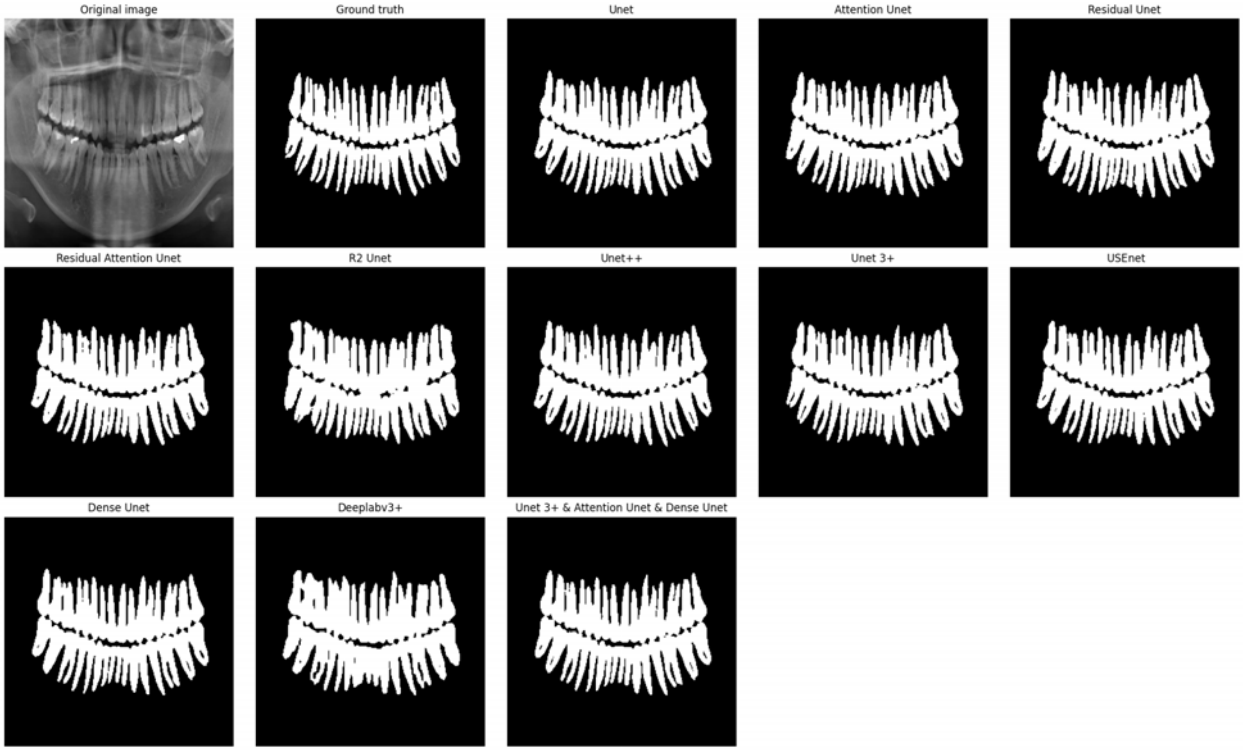

Figure 3 – Results of tooth segmentation for all the models

6 DISCUSSION

The obtained results confirm the effectiveness of using CNNs in the task of tooth segmentation on panoramic X-ray images. Among the investigated architectures, the Dense U-Net, U-Net++, and Attention U-Net models achieved the highest quality metric scores. Their advantage is explained by advanced feature propagation mechanisms. Specifically, Dense U-Net benefits from dense connections between layers, which promote effective feature reuse. Attention U-Net is effective due to the use of attention mechanisms, allowing the model to focus on relevant areas of the image. U-Net 3+ provides deep multi-level context aggregation, which improves the restoration of spatial details.

The key result of the work is the confirmation of the effectiveness of the proposed ensemble approach. Theoretically, this is explained by the fact that the selected three best models have different architectures and may have learned different features and, as a consequence, make different types of errors. The majority voting method allows for reducing the number of errors from individual models. Thus, the ensemble demonstrates better generalization ability.

DOI 10.15588/1607-3274-2026-2-6

The obtained results are consistent with the conclusions of other researchers, who also note the high efficiency of these U-Net architecture modifications.

Table 2 presents the results of our approach in comparison with the methods of other authors using the Dice metric. As can be seen from the results, our proposed ensemble outperforms most of the models from other authors.

Despite the positive results obtained, certain limitations of this work must be considered. First, the study was conducted on a limited public dataset. In this field, there are quite few public datasets with high-quality annotated images; many authors use private datasets during their research. A large and high-quality annotated dataset could increase the quality of the models in this study. Furthermore, due to the need for simultaneous training and storage of gradients for deep neural network architectures, the input images were downsampled to a size of 256x256 pixels. Such a reduction in dimensionality inevitably leads to the loss of high-frequency spatial information. Many authors used larger image sizes, which can improve segmentation quality, and the quality metrics may correspondingly be better. For example, the authors of paper [23] used images of size 512x512, which may yield higher metric values compared to our approach, where we used images of size 256x256.

Table 2 – Comparison of our results with the results of other authors

Author	Number of images	Dice
Koch et al. [13]	1500	0.936
Chen et al. [14]	1500	0.9301
Zhao et al. [15]	1500	0.9272
Lin at al. [16]	1321	0.89
Sheng et al. [17]	100	0.6372
Hou et al. [18]	1500	0.9428
Zannah et al. [19]	389	0.9033
Mărginean et al. [20]	1116	0.889
Li et al. [21]	425	0.9254
Bhat et al. [22]	2500	0.9258
Ma et al. [23]	1500	0.9454
Our results	892	0.9351

The obtained results have significant practical importance for dental diagnostics. The high accuracy of tooth segmentation achieved by the Dense U-Net, Attention U-Net, and U-Net 3+ models allows for the automation of the analysis stage of panoramic X-ray images. This can substantially reduce the workload on dentists, shorten image analysis time, and lower the probability of missing pathological changes. Furthermore, the segmentation results can be used as input data for systems of automatic caries detection or other dental lesions, which creates prerequisites for comprehensive computer-aided support in dental diagnostics.

Further research could be directed towards investigating other ensemble methods, particularly probability map averaging and weighted averaging. Another promising direction involves the evaluation of more advanced neural network architectures, particularly transformer-based models. Transformers are capable of capturing long-range spatial dependencies more effectively than classical convolutional architectures, which may lead to improved recognition of complex dental structures on panoramic X-rays. Studying transformer models both individually and as components of an ensemble could help identify even more optimal architectural combinations for dental segmentation tasks. Additionally, based on the conducted research, prerequisites are being created for the development of lightweight model variants, as the ensemble may be too large for deployment on edge devices.

CONCLUSIONS

In this work, the scientific and practical problem of increasing the accuracy and stability of semantic segmentation of teeth on panoramic X-ray images was solved by analyzing modern CNN architectures and applying an ensemble approach. An experimental study of the U-Net, Attention U-Net, Residual U-Net, Residual Attention U-Net, R2 U-Net, Dense U-Net, USE-Net, U-Net++, U-Net 3+, and DeepLabV3+ (ResNet-101 backbone) models was conducted using the Accuracy, IoU, and Dice metrics. The highest quality indicators were demonstrated by the Dense U-Net, Attention U-Net, and U-Net 3+ architectures, which indicates the effectiveness of their feature propagation mechanisms. An ensemble based on these three best models, utilizing majority voting, was proposed. Using this approach, a Dice coefficient and an IoU score of 0.9351 and 0.8783, respectively, were achieved on the test dataset, surpassing the results of all individual models.

The scientific novelty of the work lies in the comprehensive comparative analysis of modern CNN architectures for the task of tooth segmentation on panoramic X-ray images and the identification of the most effective architectures. The effectiveness of an ensemble consisting of Dense U-Net, Attention U-Net, and U-Net 3+ models was substantiated for the first time. This approach allows for reducing the number of errors and improving segmentation quality.

The practical significance of the obtained results lies in the possibility of using the proposed models in computer-aided diagnostic systems, which allow for automating the analysis of panoramic X-ray images, reducing the influence of the human factor, and accelerating the diagnostic process.

Prospects for further research include the investigation of other neural network architectures and their combination into an ensemble, the use of other ensemble methods, and the study of methods for creating a more lightweight model.

ACKNOWLEDGEMENTS

This work was conducted as an independent research project by the authors. The study was performed without financial support.

DECLARATIONS

Conflict of interest: The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that

DOI 10.15588/1607-3274-2026-2-6

could affect the research and its results presented in this paper.

Authors' contributions: Vasyl Bohush: building models, experimental model research; Maryna Ivanchenko: model selection, results analysis.

Data availability: The dataset used in this study is openly available in the Kaggle repository at https://www.kaggle.com/datasets/truthisneverlinear/childrens-dental-panoramic-radiographs-dataset.

Software availability: The manuscript has no associated software.

Use of artificial intelligence tools: The authors confirm that they did not use artificial intelligence technologies in creating the submitted work.

REFERENCES

1. Rayed M. E., Islam S. M. S., Niha S. I. et al. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked,* 2024, Vol. 47, P. 101504. DOI: 10.1016/j.imu.2024.101504
2. Huang C., Wang J., Wang S. et al. A review of deep learning in dentistry. *Neurocomputing*, 2023, Vol. 554, P. 126629. DOI: 10.1016/j.neucom.2023.126629
3. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*, 2015, Vol. 9351, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28
4. Oktay O., Schlemper J., Folgoc L. L. et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999,* 2018. DOI: 10.48550/arXiv.1804.03999
5. Zhang Z., Liu Q., Wang Y. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters,* 2018, Vol. 15, № 5, pp. 749–753. DOI: 10.1109/LGRS.2018.2802944
6. Ni Z.-L., Bian G.-B., Zhou X.-H. et al. RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments. *Lecture Notes in Computer Science,* 2019, Vol. 11954, pp. 139–149. DOI: 10.1007/978-3-030-36711-4_13
7. Alom M. Z., Yakopcic C., Hasan M. et al. Recurrent residual U-Net for medical image segmentation. *Journal of medical imaging,* 2019, Vol. 6, № 1, pp. 014006–014006. DOI: 10.1117/1.JMI.6.1.014006
8. Rundo L., Han C., Nagano Y. et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing,* 2019, Vol. 365, pp. 31–43. DOI: 10.1016/j.neucom.2019.07.006
9. Cai S., Tian Y., Lui H. et al. Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative imaging in medicine and surgery,* 2020, Vol. 10, № 6, P. 1275. DOI: 10.21037/qims-19-1090
10. Zhou Z., Rahman Siddiquee M. M., Tajbakhsh N. et al.Unet++: A nested u-net architecture for medical image segmentation. *Lecture Notes in Computer Science,* 2018, Vol. 11045, pp. 3–11. DOI: 10.1007/978-3-030-00889-5_1
11. Huang H., Lin L., Tong R. et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *Acoustics, Speech and Signal Processing (ICASSP) : IEEE International Conference, 4–8 May 2020 : proceedings*. Barcelona: IEEE, 2020, pp. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405
12. Chen L.-C., Zhu Y., Papandreou G. et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Computer Vision – ECCV 2018 : 15th European Conference, 8–14 September 2018 : proceedings.* Munich, Springer-Verlag, 2018, pp. 833–851. DOI: 10.1007/978-3-030-01234-2_49
13. Koch T. L., Perslev M., Igel C. et al. Accurate Segmentation of Dental Panoramic Radiographs with U-NETS. *Biomedical Imaging (ISBI 2019) : IEEE 16th International Symposium, 8–11 April 2019 : proceedings.* Venice, IEEE, 2019, pp. 15–19. DOI: 10.1109/ISBI.2019.8759563
14. Chen Q., Zhao Y., Liu Y. et al. MSLPNet: multi-scale location perception network for dental panoramic X-ray image segmentation. *Neural Computing and Applications,* 2021, Vol. 33, № 16, pp. 10277–10291. DOI: 10.1007/s00521-021-05790-5
15. Zhao Y., Li P., Gao C. et al. TSASNet: Tooth segmentation on dental panoramic X-ray images by Two-Stage Attention Segmentation Network. *Knowledge-Based Systems,* 2020, Vol. 206, P. 106338. DOI: 10.1016/j.knosys.2020.106338
16. Lin S., Hao X., Liu Y. et al. Lightweight deep learning methods for panoramic dental X-ray image segmentation. *Neural Computing and Applications,* 2023, Vol. 35, № 11, pp. 8295–8306. DOI: 10.1007/s00521-022-08102-7
17. Sheng C., Wang L., Huang Z. et al. Transformer-Based Deep Learning Network for Tooth Segmentation on Panoramic Radiographs. *Journal of systems science and complexity,* 2022, Vol. 36, № 1, pp. 257–272. DOI: 10.1007/s11424-022-2057-9
18. Hou S., Zhou T., Liu Y. et al. Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancemen. *Computers in Biology and Medicine,* 2023, Vol. 152, P. 106296. DOI: 10.1016/j.compbiomed.2022.106296
19. Zannah R., Bashar M., Mushfiq R. B. et al. Semantic Segmentation on Panoramic Dental X-Ray Images Using U-Net Architectures. *IEEE Access,* 2024, Vol. 12, pp. 44598–44612. DOI: 10.1109/ACCESS.2024.3380027
20. Mărginean A. C., Mureşanu S., Hedeşiu M. et al. Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks' ensemble. *Heliyon,* 2024, Vol. 10, № 10, P. e30836. DOI: 10.1016/j.heliyon.2024.e30836
21. Li Z., Tang W., Gao S. et al. Adapting SAM2 Model from Natural Images for Tooth Segmentation in Dental Panoramic X-Ray Images. *Entropy*, 2024, Vol. 26, № 12, P. 1059. DOI: 10.3390/e26121059
22. Bhat S., Birajdar G. K., Patil M. D.Tooth segmentation in panoramic dental radiographs using deep convolution neural network -Insights from subjective analysis. *Discover Applied Sciences,* 2025, Vol. 7, № 4, P. 279. DOI: 10.1007/s42452-025-06606-0
23. Ma T., Li J., Dang Z. et al. A Dual-Stream Dental Panoramic X-Ray Image Segmentation Method Based on Transformer Heterogeneous Feature Complementation. *Technologies,* 2025, Vol. 13, № 7, P. 293. DOI: 10.3390/technologies13070293
24. Children's Dental Panoramic Radiographs Dataset [Electronic resource]. Access mode: https://www.kaggle.com/datasets/truthisneverlinear/childrens-dental-panoramic-radiographs-dataset

Received 12.12.2025.
Accepted 08.04.2026.
Published 26.06.2026.

DOI 10.15588/1607-3274-2026-2-6

УДК 004.8

АНСАМБЛЬ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ СЕГМЕНТАЦІЇ ЗУБІВ НА ПАНОРАМНИХ РЕНТГЕНІВСЬКИХ ЗНІМКАХ

Богуш В. І. – магістр кафедри інженерії програмного забезпечення та інформаційних технологій, Дніпровський національний університет імені Олеся Гончара, Дніпро, Україна. ROR: https://ror.org/00qk1f078. ORCID: https://orcid.org/0009-0007-0986-3653.

Іванченко М. Г. – канд. техн. наук, доцент, доцент кафедри інженерії програмного забезпечення та інформаційних технологій, Дніпро, Україна. ROR: https://ror.org/00qk1f078. ORCID: https://orcid.org/0000-0001-7795-0459.

АНОТАЦІЯ

Актуальність. Семантична сегментація зубів на панорамних рентгенівських знімках є важливим завданням у галузі стоматологічної діагностики, оскільки дозволяє автоматизувати процес діагностики стоматологічних захворювань. Проте панорамні рентгенівські знімки мають складну структуру, що ускладнює задачу сегментації. Застосування згорткових нейронних мереж показує високий потенціал у розв'язанні цієї задачі. У цьому контексті актуальним є дослідження кращих моделей та поєднання їх у ансамбль з метою підвищення якості сегментації.

Мета роботи – дослідження ефективності різних архітектур згорткових нейронних мереж у задачі семантичної сегментації зубів на панорамних рентгенівських знімках та розроблення ансамблевого підходу для покращення якості результатів.

Метод. Використовуються різні архітектури згорткових нейронних мереж: U-Net, Attention U-Net, Residual U-Net, Residual Attention U-Net, R2 U-Net, U-Net++, U-Net 3+, USE-Net, Dense U-Net, DeepLabV3+ з попередньо навченою моделлю ResNet-101 на датасеті ImageNet. Запропоновано ансамблевий підхід з кращих моделей, де фінальна маска сегментації визначається за методом мажоритарного голосування. Навчання моделей виконувалося на попередньо підготовленому наборі панорамних рентгенівських знімків із застосуванням методів аугментації. Оцінювання якості моделей проводилося за метриками IoU, Dice та Accuracy.

Результати. Досліджено різні моделі нейромереж, найкращі з яких були об'єднані в ансамбль. Проведені експерименти підтвердили, що ансамблевий підхід забезпечує покращення точності сегментації порівняно з окремими моделями. Найкращий результат показав ансамбль, який поєднує архітектури Dense U-Net, Attention U-Net та U-Net 3+.

Висновки. Запропонований ансамблевий підхід продемонстрував високу ефективність у задачі семантичної сегментації зубів на панорамних рентгенівських знімках, перевершуючи результати окремих моделей. Наукова новизна роботи полягає в застосуванні ансамблевого підходу з використанням різних архітектур згорткових нейронних мереж для семантичної сегментації зубів на рентгенівських знімках. Практична значущість полягає у можливості використання розробленого підходу для створення систем автоматизованої діагностики у стоматології. Отримані результати можуть бути застосовані для подальшої автоматизації процесів аналізу рентгенівських знімків і розвитку інтелектуальних медичних систем.

КЛЮЧОВІ СЛОВА: глибоке навчання, згорткові нейронні мережі, сегментація зубів, семантична сегментація, U-Net, ансамбль.

ЛІТЕРАТУРА

1. Deep learning for medical image segmentation: State-of-the-art advancements and challenges / [M. E. Rayed, S. M. S. Islam, S. I. Niha et al.] // Informatics in Medicine Unlocked. – 2024. – Vol. 47. – P. 101504. DOI: 10.1016/j.imu.2024.101504
2. A review of deep learning in dentistry / [C. Huang, J. Wang, S. Wang et al.] // Neurocomputing. – 2023. – Vol. 554. – P. 126629. DOI: 10.1016/j.neucom.2023.126629
3. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // Lecture Notes in Computer Science. – 2015. – Vol. 9351. – P. 234–241. DOI: 10.1007/978-3-319-24574-4_28
4. Attention U-Net: Learning Where to Look for the Pancreas / [O. Oktay, J. Schlemper, L. L. Folgoc et al.] // arXiv preprint arXiv:1804.03999. – 2018. DOI: 10.48550/arXiv.1804.03999
5. Zhang Z. Road Extraction by Deep Residual U-Net / Z. Zhang, Q. Liu, Y. Wang // IEEE Geoscience and Remote Sensing Letters. – 2018. – Vol. 15, № 5. – P. 749–753. DOI: 10.1109/LGRS.2018.2802944
6. RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments / [Z.-L. Ni, G.-B. Bian, X.-H. Zhou et al.] // Lecture Notes in Computer Science. – 2019. – Vol. 11954. – P. 139–149. DOI: 10.1007/978-3-030-36711-4_13
7. Recurrent residual U-Net for medical image segmentation / [M. Z. Alom, C. Yakopcic, M. Hasan et al.] // Journal of medical imaging. – 2019. – Vol. 6, № 1. – P. 014006–014006. DOI: 10.1117/1.JMI.6.1.014006
8. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets / [L. Rundo, C. Han, Y. Nagano et al.] // Neurocomputing. – 2019. – Vol. 365. – P. 31–43. DOI: 10.1016/j.neucom.2019.07.006
9. Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network / [S. Cai, Y. Tian, H. Lui et al.] // Quantitative imaging in medicine and surgery. – 2020. – Vol. 10, № 6. – P. 1275. DOI: 10.21037/qims-19-1090
10. Unet++: A nested u-net architecture for medical image segmentation / [Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh et al.] // Lecture Notes in Computer Sci-

DOI 10.15588/1607-3274-2026-2-6

ence. – 2018. – Vol. 11045. – P. 3–11. DOI: 10.1007/978-3-030-00889-5_1
11. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation / [H. Huang, L. Lin, R. Tong et al.] // Acoustics, Speech and Signal Processing (ICASSP) : IEEE International Conference, 4–8 May 2020 : proceedings. – Barcelona : IEEE, 2020. – P. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405
12. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation / [L.-C. Chen, Y. Zhu, G. Papandreou et al.] // Computer Vision – ECCV 2018 : 15th European Conference, 8–14 September 2018 : proceedings. – Munich : Springer-Verlag, 2018. – P. 833–851. DOI: 10.1007/978-3-030-01234-2_49
13. Accurate Segmentation of Dental Panoramic Radiographs with U-NETS / [T. L. Koch, M. Perslev, C. Igel et al.] // Biomedical Imaging (ISBI 2019) : IEEE 16th International Symposium, 8–11 April 2019 : proceedings. – Venice : IEEE, 2019. – P. 15–19. DOI: 10.1109/ISBI.2019.8759563
14. MSLPNet: multi-scale location perception network for dental panoramic X-ray image segmentation / [Q. Chen, Y. Zhao, Y. Liu et al.] // Neural Computing and Applications. – 2021. – Vol. 33, № 16. – P. 10277–10291. DOI: 10.1007/s00521-021-05790-5
15. TSASNet: Tooth segmentation on dental panoramic X-ray images by Two-Stage Attention Segmentation Network / [Y. Zhao, P. Li, C. Gao et al.] // Knowledge-Based Systems. – 2020. – Vol. 206. – P. 106338. DOI: 10.1016/j.knosys.2020.106338
16. Lightweight deep learning methods for panoramic dental X-ray image segmentation / [S. Lin, X. Hao, Y. Liu et al.] // Neural Computing and Applications. – 2023. – Vol. 35, № 11. – P. 8295–8306. DOI: 10.1007/s00521-022-08102-7
17. Transformer-Based Deep Learning Network for Tooth Segmentation on Panoramic Radiographs / [C. Sheng, L. Wang, Z. Huang et al.] // Journal of systems science and complexity. – 2022. – Vol. 36, № 1. – P. 257–272. DOI: 10.1007/s11424-022-2057-9
18. Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement / [S. Hou, T. Zhou, Y. Liu et al.] // Computers in Biology and Medicine. – 2023. – Vol. 152. – P. 106296. DOI: 10.1016/j.compbiomed.2022.106296
19. Semantic Segmentation on Panoramic Dental X-Ray Images Using U-Net Architectures / [R. Zannah, M. Bashar, R. B. Mushfiq et al.] // IEEE Access. – 2024. – Vol. 12. – P. 44598–44612. DOI: 10.1109/ACCESS.2024.3380027
20. Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks' ensemble / [A. C. Mărginean, S. Mureşanu, M. Hedeşiu et al.] // Heliyon. – 2024. – Vol. 10, № 10. – P. e30836. DOI: 10.1016/j.heliyon.2024.e30836
21. Adapting SAM2 Model from Natural Images for Tooth Segmentation in Dental Panoramic X-Ray Images / [Z. Li, W. Tang, S. Gao et al.] // Entropy. – 2024. – Vol. 26, № 12. – P. 1059. DOI: 10.3390/e26121059
22. Bhat S. Tooth segmentation in panoramic dental radiographs using deep convolution neural network -Insights from subjective analysis / S. Bhat, G. K. Birajdar, M. D. Patil // Discover Applied Sciences. – 2025. – Vol. 7, № 4. – P. 279. DOI: 10.1007/s42452-025-06606-0
23. A Dual-Stream Dental Panoramic X-Ray Image Segmentation Method Based on Transformer Heterogeneous Feature Complementation / [T. Ma, J. Li, Z. Dang et al.] // Technologies. – 2025. – Vol. 13, № 7. – P. 293. DOI: 10.3390/technologies13070293
24. Children's Dental Panoramic Radiographs Dataset [Electronic resource]. – Access mode: https://www.kaggle.com/datasets/truthisneverlinear/childrens-dental-panoramic-radiographs-dataset

DOI 10.15588/1607-3274-2026-2-6