UDC 004.8:004.932:616.711(045)
DOI:10.18372/1990-5548.82.19365

[1]V. M. Sineglazov,
[2]O. I. Chumachenko,
[3]O. A. Pokhylenko

BAFUNET: HYBRID U-NET FOR SEGMENTATION OF SPINE MR IMAGES

[1]Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
[2,3]Department of Artificial Intelligence, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
E-mails: [1]svm@nau.edu.ua ORCID 0000-0002-3297-9060,
[2]chumachenko@tk.kpi.ua ORCID 0000-0003-3006-7460,
[3]o.pokhylenko@kpi.ua ORCID 0000-0002-1562-2051

Abstract—*The paper presents the development of a hybrid neural network architecture, BAFUNet, designed for the segmentation of spine MR images in the context of medical diagnostics. The architecture builds upon the classical U-Net, integrating atrous spatial pyramid pooling module in the bottleneck and a two-round fusion module in the skip connections to address challenges such as various object scales and unclear boundaries in medical images. The work describes the design of the proposed BAFUNet architecture, its implementation, and the experimental results. A comparative analysis was performed against classical U-Net and ResUNet++, demonstrating the relationship between the proposed architectural enhancements and segmentation performance. The evaluation was carried out using Dice score and Jaccard score metrics on the SPIDER dataset, a publicly available lumbar spine magnetic resonance imaging dataset. The results indicate that the BAFUNet architecture achieves a slight but consistent improvement in segmentation performance, with an average Dice Score increase of 0.003–0.005 compared to baseline models, highlighting its potential applicability in automated medical diagnostics.*

Index Terms—Hybrid neural network architecture; convolutional neural network; U-Net; image segmentation; spine MRI.

I. INTRODUCTION

Medical image segmentation is a foundational task in modern medical diagnostics, playing a vital role in disease detection, treatment planning, and surgical guidance. Segmentation of spine MR images, in particular, is crucial for diagnosing spine-related pathologies such as intervertebral disc degeneration, spondylolisthesis, and vertebral fractures. Traditionally, manual segmentation methods have been used, but they are time-intensive, prone to human error, and require significant expertise. These limitations have driven the development of automated segmentation techniques, which leverage machine learning and deep neural networks to provide faster and more consistent results.

One of the most influential advancements in this field is the U-Net architecture, introduced as a convolutional neural network (CNN) specifically designed for biomedical image segmentation. Due to its modular design and ability to handle limited training data effectively, U-Net has become a cornerstone model for medical image segmentation tasks. As highlighted in the review [1], U-Net's modular design and encoder-decoder structure make it adaptable to various medical imaging challenges, contributing to its widespread adoption in both academic and industrial settings. Over the years, several variants of U-Net have emerged, each addressing specific needs and extending its capabilities to tackle complex segmentation tasks.

Given the continued success of U-Net and its variants, the ongoing development of new models is essential to address the diverse challenges posed by medical image segmentation. Inspired by these advances, we propose a novel architecture, BAFUNet (Bottleneck ASPP & Fusion U-Net), aimed at improving spine MR image segmentation. This hybrid architecture builds upon the well-established U-Net framework by integrating advanced modules, including an Atrous Spatial Pyramid Pooling (ASPP) block in the bottleneck and a two-round fusion module in the skip connections, to enhance the network's ability to capture multi-scale features and improve segmentation accuracy. By designing BAFUNet with these enhancements, we aim to provide a more robust and effective solution for the segmentation of spine MR images, with the potential to contribute to the broader field of medical image analysis.

II. PROBLEM STATEMENT

Segmentation of spine MR images is essential for accurate diagnosis and treatment of spinal disorders. The main task of semantic segmentation of MR images is to partition an MR image $I \in \mathbb{R}^{h \times w}$, where w and h are the width and height of the image respectively ($w = h = 256$ in our case), into distinct anatomical regions corresponding to different structures, including vertebral bodies (VB), intervertebral discs (IVD), and the spinal canal (SC). This task presents several challenges, such as the complexity of anatomical structures, overlapping boundaries and blurred features. In addition, spine MR images can vary greatly depending on the patient, the MRI machine used, and the imaging protocols (e.g., T1-weighted and T2-weighted imaging). These issues make it difficult to achieve reliable and consistent segmentation results.

In this paper the task of segmentation of spine MR images is formalized as the classification of each pixel $p(x, y)$ of the MR image I into one of the following classes:

$$\hat{C}(x, y) = \{VB, IVD, SC\},$$

where $\hat{C}(x, y)$ is the predicted label for the pixel at position (x, y) in the image I.

In such cases, a multi-label segmentation approach can be used, where the model must predict a set of probabilities for each pixel, indicating the likelihood of each class at that pixel:

$$\hat{P}(x, y) = \left[\hat{P}_{VB}(x, y), \hat{P}_{IVD}(x, y), \hat{P}_{SC}(x, y)\right],$$

where each $\hat{P}_C$ represents the predicted probability of the pixel $p(x, y)$ belonging to class c.

The predicted probabilities are computed independently for each class, so the output of the model for each pixel is a vector of probabilities for the three classes. This formulation allows to independently learn the likelihood of each class for each pixel and to make multi-class predictions, which can be useful in regions with overlapping or blurred boundaries between anatomical structures.

III. RELATED WORK

A. U-Net-Based Systems for Segmentation of Spine MR Images

Over the past years many researchers proposing different solutions to the segmentation of spine MRI images using deep learning, particularly the U-Net architecture. Several studies have explored different U-Net-based models for segmentation of spine MR images, each addressing unique challenges and offering specific solutions.

In article [2], the authors used U-Net++ (a modified U-Net architecture) and Yolov5x architectures, leveraging transfer learning. For U-Net++, the model was initialized with weights pre-trained on the ImageNet dataset, and Yolov5x was pre-trained on the COCO 2017 dataset. While data augmentation was applied to reduce overfitting due to limited training data, the study observed the negative impact of noise on input images, which reduced segmentation accuracy. Despite this, U-Net++ achieved promising results on the validation set, with a Dice score of 0.93 for vertebrae segmentation and 0.96 for intervertebral discs.

In study [3], the authors proposed using Sequential Conditional Reinforcement Learning in conjunction with anatomical modeling, ResNet for bounding box detection, and Y-Net (an extended U-Net architecture) for segmentation. The proposed model achieved good results, with Jaccard and Dice scores of 0.923 and 0.926, respectively. However, the performance was slightly lower than the results in [2], likely due to a smaller and more diverse dataset, as MR images from different machines were used.

The study [4] introduced Spine Explorer, a software that uses U-Net to segment VB, IVD, and the SC. Despite training the U-Net model with only 50 MRI images for both training and testing, the study achieved notable results, with a Jaccard index of 0.947 for vertebrae and 0.926 for discs.

Although the results from these studies are promising, the differences in dataset size and quality make direct comparisons challenging. These models show the variety of approaches and highlight the challenges faced when segmenting spine MRIs, such as limited datasets, noise, and the need for robust methods that generalize well.

B. Modifications to U-Net for Improved Medical Image Segmentation

U-Net [5] is one of the most popular architectures for medical image segmentation, owing to its fully convolutional nature, which makes it flexible and computationally efficient. The architecture consists of an encoder that extracts features through convolutional layers and Max Pooling, which reduces resolution, while the decoder upsamples these features to reconstruct the output. U-Net is known for its use of skip-connections, which help preserve spatial information during segmentation. The activation function typically used is ReLU, and the use of convolutional layers instead of fully connected layers significantly reduces the number of parameters in the model.

U-Net's flexibility and modular design have led to numerous hybrid solutions and architectural modifications to address specific challenges in biomedical segmentation tasks. These modifications aim to enhance the performance of U-Net by incorporating new blocks or adjusting the architecture in response to the needs of the task at hand. One of the simplest modifications involves adding batch normalization layers after each convolutional layer, which was demonstrated in [6] to improve training on small-sized MRI datasets.

Hybrid architectures combine several paradigms or structural elements (blocks) to address the limitations of individual models. These modifications are particularly useful in the case of U-Net, where enhancements are often made to specific components, such as the encoder, decoder, bottleneck, and/or skip-connections (Fig. 1). By integrating different blocks, these hybrid architectures aim to improve performance and efficiency. Below, we discuss several notable modifications to the U-Net architecture that have been proposed to address the challenges posed by medical image segmentation tasks.

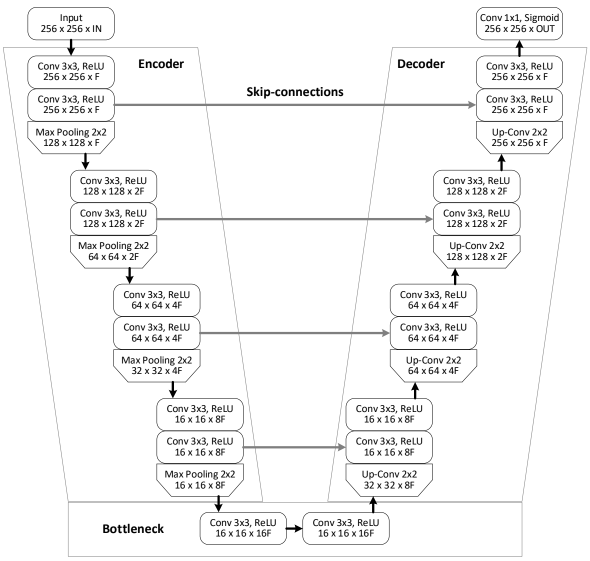

Fig. 1. Components of four-level U-Net architecture

DENSE-Inception U-Net: In article [7], the authors proposed using Inception-Res blocks in both the encoder and decoder, and Dense-Inception blocks in the bottleneck. The Dense-Inception block comprises several Inception-Res blocks, each containing skip-connections and batch normalization layers. While this approach led to improved feature extraction, the large number of Dense-Inception blocks increased the model's parameters, potentially slowing down training without guaranteeing a substantial performance improvement.

FusionU-Net: To address the semantic gap between the encoder and decoder, [8] introduced the FusionU-Net model, which incorporates a fusion module to bridge the gap before applying the skip-connections. This module, consisting of specialized blocks (Fig. 2) such as DownFuse and UpFuse, ensures bi-directional information exchange between adjacent encoder layers. The proposed two-round fusion process showed promising results in improving segmentation performance.

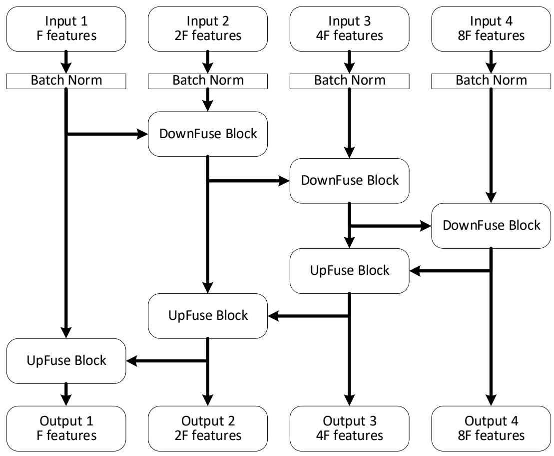

Fig. 2. Fuse block architecture

HDA-ResUNet: [9] proposed the HDA-ResUNet model, which integrates attention mechanisms into the skip-connections. By using channel attention blocks and a hybrid dilated attention convolutional layer in the bottleneck, the model enhances its focus on more relevant features while suppressing less informative ones.

ResUNet++: In article [10], the authors introduced a variant of U-Net that combines residual blocks (Res), Squeeze-and-Excitation (SE) blocks, Atrous Spatial Pyramid Pooling (Fig. 3), and attention blocks. The ResUNet++ architecture performs well on smaller datasets, as it effectively combines feature extraction and attention mechanisms to enhance the model's ability to capture relevant details.

MS-TransUNet++: Lastly, the MS-TransUNet++ model [11] incorporates transformer-based blocks in the bottleneck, demonstrating the growing trend of using transformers to improve segmentation tasks. It also proposes flexible fusion schemes for skip-connections, allowing information to be shared between the encoder and decoder.

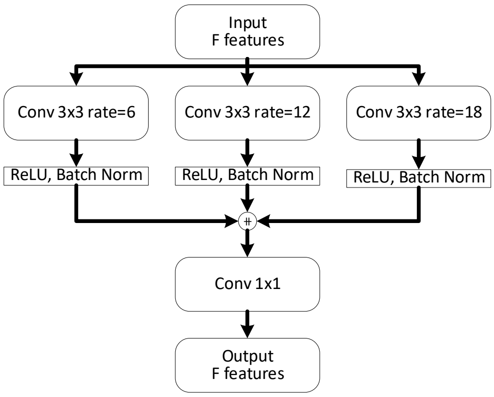

Fig. 3. ASPP block architecture

IV. METHODOLOGY

Based on the review of hybrid solutions for medical image segmentation, we propose a novel architecture called BAFUNet (Bottleneck ASPP & Fusion U-Net). It is a custom hybrid architecture for segmentation of spine MR images designed to address challenges such as multi-scale object segmentation and accuracy on small datasets.

BAFUNet is based on a four-level U-Net architecture (see Fig. 1) with batch normalization layers following each convolutional layer [6]. Its key architectural modifications include bottleneck replacement and enhanced skip-connections.

The bottleneck of the base U-Net is replaced with ASPP block (see Fig. 3), inspired by ResUNet++ [10]. This modification improves multi-scale feature extraction, enhancing the model's ability to handle structural defects of varying sizes.

To reduce semantic gaps, skip-connections are augmented using a two-round fusion module inspired by FusionU-Net [8]. This module employs two fuse blocks (see Fig. 2) for better integration of spatial and semantic features.

The design of the architecture, including the number of layers, block configurations, and specific parameters, was determined empirically. Various configurations, such as the number of convolutional layers, and the incorporation of ASPP and fusion modules, were tested. These configurations showed promising results, suggesting that the proposed modifications improve segmentation performance of spine MR images. The results of these experiments are detailed in the following section.

The resulting hybrid architecture is depicted in Fig 4. It is designed to efficiently extract objects with complex or unclear boundaries and operate effectively on small datasets. This makes it suitable for segmenting VB, IVD, and the SC in MR images, where object boundaries may be ambiguous or challenging to define.

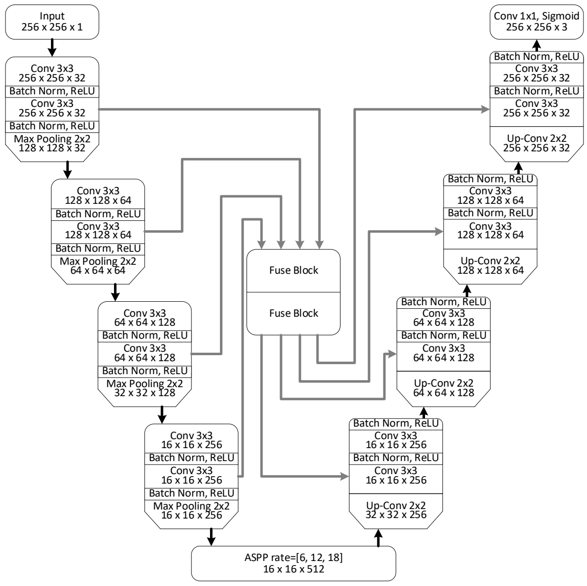

Fig. 4. BAFUNet architecture

V. EXPERIMENTS AND RESULTS

The experiments were conducted using a publicly available dataset of 218 patients' MRI images of the spine (both T1-weighted and T2-weighted), detailed in [12]. The dataset contains 447 3D MRI images, in the MHA format, collected from four different hospitals. Each image includes segmentations of the VB, IVD, and the SC, generated through an iterative semi-automated method. In this process, an initial segmentation was performed automatically, followed by manual verification and adjustment.

For the purpose of 2D segmentation, 2D slices were extracted from the 3D MRI volumes and used for training and validation of the models.

The model was implemented using the PyTorch machine learning library, along with the torchvision package for computer vision tasks. Data preprocessing was performed with popular scientific computing libraries such as NumPy and scikit-learn. Matplotlib was used for visualizing the results, and OpenCV was employed for processing the 2D images.

Training was conducted over 40 epochs with the Adam optimizer and an initial learning rate of 0.001, which was reduced by a factor of 10 every 20 epochs. A batch size of 8 was used for the training process. The Jaccard Loss function (Fig. 5), based on the Intersection over Union metric, was chosen as the loss function, while the model with the highest Dice score (F_1 score) on the validation set was saved after each epoch.

The models compared in this study include U-Net, ResUNet++, and the proposed BAFUNet architecture. The goal of the experiments was to evaluate the performance of each architecture in terms of segmentation accuracy and robustness, particularly in extracting structures such as VB, IVD, and the SC.

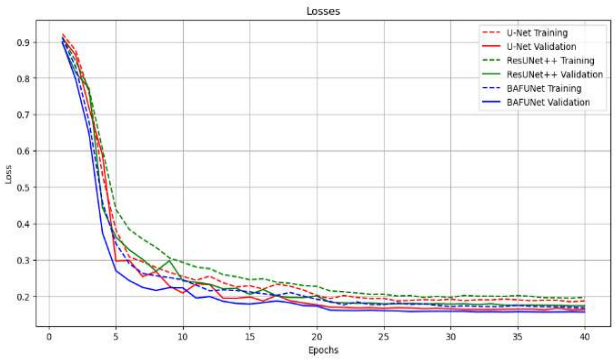

Fig. 5. Losses for compared models

The training results demonstrated that all models reached a plateau towards the end of the training process. However, BAFUNet outperformed the other architectures, with the highest Dice score observed during the final epochs (Fig. 6). Specifically, the BAFUNet model achieved a Dice score of 0.916 that was 0.003-0.005 higher than that of the second-best model, indicating better overall accuracy in segmentation.

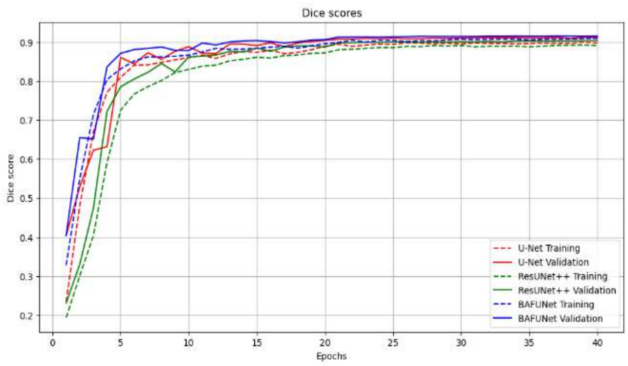

Fig. 6. Dice scores for compared models

The segmentation result obtained using the trained BAFUNet model, shown in Fig. 7, is generally quite accurate. However, there are some noticeable segmentation errors, such as the absence of small segments of the vertebral bodies in the 2D slices and some challenges in defining the boundaries of highly deformed intervertebral discs.

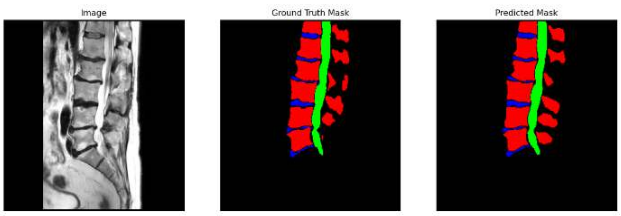

Fig. 7. Segmentation result

VI. Conclusions

We introduced BAFUNet, a novel hybrid architecture designed to enhance the segmentation of spine MR images. By integrating the ASPP block in the bottleneck and a two-round fusion module in the skip-connections, BAFUNet effectively addresses the challenges of multi-scale feature extraction and boundary clarity, which are particularly critical in medical image segmentation tasks.

The experimental results demonstrate that BAFUNet outperforms classical models such as U-Net and ResUNet++ in terms of segmentation accuracy, as measured by Dice score and Jaccard index. Specifically, BAFUNet achieved an average Dice score of 0.916, which was consistently higher (by 0.003–0.005) than the other models evaluated. These improvements suggest that BAFUNet's hybrid design leads to better handling of complex spinal structures in MR images, particularly in cases of unclear boundaries and varying object scales.

Despite its improvements, BAFUNet still faces challenges in segmenting some structures, which highlights the need for further refinements. However, the results suggest that BAFUNet has potential for application in automated medical diagnostics, such as diagnosing intervertebral disc degeneration and other spinal pathologies.

Further work can be focused on optimizing the architecture for even better generalization across diverse datasets, exploring additional enhancements to improve segmentation performance on smaller or highly deformed structures, and evaluating BAFUNet's applicability to other medical image segmentation tasks.

In summary, BAFUNet demonstrates promising results for automated segmentation of spine MR images and may offer a useful tool for clinical applications, though further refinements are needed.

References

[1] R. Azad et al., “Medical Image Segmentation Review: The Success of U-Net,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. https://doi.org/10.1109/TPAMI.2024.3435571

[2] S. Guinebert et al., “Automatic semantic segmentation and detection of vertebras and intervertebral discs by neural networks,” *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100055, 2022. https://doi.org/10.1016/j.cmpbup.2022.100055

[3] D. Zhang, B. Chen, and S. Li, “Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy,” *Medical Image Analysis*, vol. 67, p. 101861, 2021. https://doi.org/10.1016/j.media.2020.101861

[4] J. Huang et al., "Spine Explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images," *The Spine Journal*, vol. 20, no. 4, pp. 590–599, 2020. https://doi.org/10.1016/j.spinee.2019.11.010

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention--MICCAI 2015: 18th international conference*, Munich, Germany, October 5-9, 2015, proceedings, part III 18, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[6] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in biology and medicine*, vol. 109, pp. 218–225, 2019. https://doi.org/10.1016/j.compbiomed.2019.05.002

[7] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Computer methods and programs in biomedicine*, vol. 192, p. 105395, 2020. https://doi.org/10.1016/j.cmpb.2020.105395

[8] Z. Li, H. Lyu, and J. Wang, "FusionU-Net: U-Net with Enhanced Skip Connection for Pathology Image Segmentation," in *Asian Conference on Machine Learning*, 2024, pp. 694–706. https://doi.org/10.48550/arXiv.2310.10951

[9] Z. Wang, Y. Zou, and P. X. Liu, "Hybrid dilation and attention residual U-Net for medical image segmentation," *Computers in biology and medicine*, vol. 134, p. 104449, 2021. https://doi.org/10.1016/j.compbiomed.2021.104449

[10] D. Jha et al., "ResUNet++: An Advanced Architecture for Medical Image Segmentation," 2019, *IEEE International Symposium on Multimedia (ISM)*, San Diego, CA, USA, 2019, pp. 225–2255. https://doi.org/10.1109/ISM46123.2019.00049

[11] B. Wang, F. Wang, P. Dong, and C. Li, "Multiscale TransUNet++: dense hybrid u-net with transformer for medical image segmentation," *Signal, Image and Video Processing*, vol. 16, no. 6, pp. 1607–1614, 2022. https://doi.org/10.1007/s11760-021-02115-w

[12] J. W. van der Graaf et al., "Lumbar spine segmentation in MR images: a dataset and a public benchmark," *Scientific Data*, vol. 11, no. 1, p. 264, 2024. https://doi.org/10.1038/s41597-024-03090-w

Received October 21, 2024

Sineglazov Victor. ORCID 0000-0002-3297-9060. Doctor of Engineering Science. Professor. Head of the Department of Aviation Computer-Integrated Complexes.
Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant, artificial intelligence.
Publications: more than 700 papers.
E-mail: svm@nau.edu.ua

Chumachenko Olena. ORCID 0000-0003-3006-7460. Doctor of Engineering Science. Professor.
Department of Artificial Intelligence, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Georgian Polytechnic Institute, Tbilisi, Georgia, (1980).
Research area: system analysis, artificial neural networks.
Publications: more than 80 papers.
E-mail: chumachenko@tk.kpi.ua

Pokhylenko Oleksandr. ORCID 0000-0002-1562-2051. Master of Computer Science.
Department of Artificial Intelligence, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, (2024).
Research Interests: intelligent systems, artificial intelligence, artificial neural networks.
Publications: more than 10 papers.
E-mail: o.pokhylenko@kpi.ua

В. М. Синєглазов, О. І. Чумаченко, О. А. Похиленко. BAFUNet: гібридна U-Net для сегментації МРТ-зображень хребта

В роботі представлено розробку гібридної архітектури нейронної мережі BAFUNet, призначеної для сегментації МРТ-зображень хребта в контексті медичної діагностики. Архітектура заснована на класичній мережі U-Net, і включає модуль розширеного просторового пірамідального об’єднання у вузькому місці та двораундовий модуль злиття у пропускних з’єднаннях для вирішення таких проблем, як різні масштаби об’єктів і нечіткі межі на медичних зображеннях. У роботі описано дизайн запропонованої архітектури BAFUNet, її реалізацію та експериментальні результати. Було проведено порівняльний аналіз із класичною U-Net і ResUNet++, що продемонструвало зв'язок між запропонованими архітектурними вдосконаленнями та ефективністю сегментації. Оцінку було проведено за допомогою коефіцієнту подібності Дайса та індексу Жаккара на наборі даних SPIDER – загальнодоступному наборі даних магнітно-резонансної томографії поперекового відділу хребта. Результати показують, що архітектура BAFUNet досягає незначного, але постійного покращення продуктивності сегментації, із збільшенням середнього коефіцієнту Дайса на 0,003–0,005 порівняно з базовими моделями, що підкреслює потенційну можливість її застосування в автоматизованій медичній діагностиці.

Ключові слова: гібридна архітектура нейронної мережі; згорткова нейронна мережа; U-Net; сегментація зображень; МРТ хребта.

Синєглазов Віктор Михайлович. ORCID 0000-0002-3297-9060. Доктор технічних наук. Професор. Завідувач кафедри авіаційних комп'ютерно-інтегрованих комплексів.

Факультет аеронавігації, електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.

Освіта: Київський політехнічний інститут, Київ, Україна, (1973).

Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки, штучний інтелект.

Кількість публікацій: більше 700 наукових робіт.

E-mail: svm@nau.edu.ua

Чумаченко Олена Іллівна. ORCID 0000-0003-3006-7460. Доктор технічних наук. Професор.

Кафедра штучного інтелекту, Навчально-науковий інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.

Освіта: Грузинський політехнічний інститут, Тбілісі, Грузія, (1980).

Напрям наукової діяльності: системний аналіз, штучні нейронні мережі.

Кількість публікацій: більше 80 наукових робіт.

E-mail: chumachenko@tk.kpi.ua

Похиленко Олександр Андрійович. ORCID 0000-0002-1562-2051. Магістр комп'ютерних наук.

Кафедра штучного інтелекту, Навчально-науковий інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Освіта: Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, (2024).

Напрям наукової діяльності: інтелектуальні системи, штучний інтелект, штучні нейронні мережі.

Кількість публікацій: більше 10 наукових робіт.

E-mail: o.pokhylenko@kpi.ua