PROGRAM SYSTEMS: THEORY AND APPLICATIONS ISSN 2079-3316
*Research Article* ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

UDC 004.93'11:616-073.75
10.25209/2079-3316-2026-17-2-147-190

# Integrating Multi-Scale Features and Attention Mechanisms for Colorectal Tumor Segmentation in CT Images

Yuqian **Wang**[1]✉, Sergey Vladimirovich **Aksenov**[2]

[1, 2] School of Information Technologies and Robotics, Tomsk Polytechnic University, Tomsk, Russia
[2] Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia
[1]✉*wangyuqian3333@gmail.com*

Abstract. In recent years, deep learning technologies have been widely applied in medical image analysis, demonstrating outstanding performance particularly in image segmentation tasks.

To address the issue of semantic information loss during the feature extraction stage in U-shaped networks, which limits the segmentation accuracy of colorectal tumors, this paper proposes a novel segmentation model based on the U-Net architecture, named MGA-UNet (Multi-scale Ghost Attention U-Net). The model integrates multi-scale feature extraction, dual channel and spatial attention mechanisms, and attention gating in skip connections. The specific improvements are as follows:

First, an enhanced Ghost module (combined with RFB) is adopted in the encoding stage to achieve extraction and fusion of multi-scale feature information.

Second, the CBAM (Convolutional Block Attention Module) is introduced into the encoding path to enhance the network's feature response to small-scale targets.

Third, attention gate units are embedded in the skip connections to suppress irrelevant background regions and highlight tumor features.

Experimental results on a colorectal tumor CT dataset demonstrate the high effectiveness of the proposed model. Compared with the classic U-Net, GhostNet, and the recent Mamba-UNet and U-SAM, the proposed model can delineate colorectal tumor regions more accurately and achieves superior segmentation performance. Furthermore, ablation studies and hyperparameter sensitivity analysis verify the effectiveness and stability of each proposed module. (*Linked article texts in English and in Russian*).

Key words and phrases: U-Net, attention mechanism, skip connection, image segmentation, MGA-UNet

2020 *Mathematics Subject Classification*: 68U10; 92C50

Acknowledgments: This work was supported by the China Scholarship Council (CSC) under grant No. 202008410491.

For citation: Yuqian Wang, Sergey V. Aksenov. *Integrating Multi-Scale Features and Attention Mechanisms for Colorectal Tumor Segmentation in CT Images*. Program Systems: Theory and Applications, 2026, **17**:2(71), pp. 147–190. (*In English, in Russian*).
https://psta.psiras.ru/read/psta2026_2_147-190.pdf

## Introduction

According to the 2024 global cancer statistics, colorectal cancer ranks third in incidence and second in mortality worldwide [1]. Clinical characteristics of the disease are closely related to the type, location, size, and number of colorectal polyps. Early detection and removal of polyps play a key role in reducing the incidence of colorectal cancer and significantly improve patient survival.

Currently, colonoscopy remains the primary method for detecting colorectal polyps. However, numerous systematic reviews and meta-analyses indicate that colonoscopy still has a substantial risk of missing polyps and adenomas [2]. Therefore, developing automated segmentation technologies for colorectal tumors not only significantly improves diagnostic accuracy but also reduces the burden on medical staff, facilitating the promotion of early screening and intervention programs.

In recent years, the rapid development of deep learning has provided powerful technical support for automatic medical image segmentation.

Shelhamer et al. proposed the Fully Convolutional Network (FCN), where fully connected layers of traditional CNNs are replaced by convolutional layers, and deconvolution operations are used to restore image resolution, while skip connections are introduced to merge semantic information from shallow and deep layers, yielding more detailed segmentation results [3].

Ben-Cohen et al. first applied FCN to liver segmentation and lesion detection, achieving an average Dice coefficient of 0.89 without complex preprocessing steps, outperforming traditional CNNs [4].

Isensee et al. developed nnU-Net, a self-configuring method that adapts to different medical image segmentation tasks [5].

Chen et al. presented EfficientNet-Lite UNet for biomedical image segmentation, providing high segmentation quality while saving computational resources [6].

Schenk et al. improved the FCN architecture by introducing long and short skip connections to transfer feature maps from the compression path to the expansion path, recovering details lost during downsampling and accelerating convergence [7].

Iqbal and Sharif proposed a semi-automatic breast tumor segmentation method using a U-shaped pyramid-dilated network, improving efficiency by leveraging both labeled and unlabeled data [8].

Ronneberger et al. introduced the classic U-Net model with a symmetric encoder-decoder structure, achieving excellent segmentation results through skip connections [9].

Zhou et al. developed U-Net++ based on U-Net, which extracts richer hierarchical information by reorganizing skip connections, minimizing the semantic gap between up- and down-sampling features [10].

Seo et al. proposed mu-U-Net, adding extra deconvolution layers and activation functions in skip connections to simultaneously extract high-level global features of small objects and high-resolution boundary information of large objects [11].

Wang et al. developed Retina UNet for head and neck tumor localization on PET/CT images, which not only improves segmentation accuracy but also predicts patient survival time [12].

In addition, lightweight convolutional networks such as GhostNet [13] and ShuffleNetV2 [14] can significantly reduce computational costs while maintaining high accuracy, offering new approaches for clinical deployment of medical image segmentation.

In colorectal tumor segmentation tasks, fluctuations in CT image quality hinder accurate identification of small tumor boundaries. To address this problem, we propose a segmentation method combining multi-scale convolutions and channel-spatial attention mechanisms.

(*1*) traditional convolutional operators are replaced by a multi-scale convolution module (Ghost+RFB) for multi-level feature extraction.

(*2*) a spatial-channel attention mechanism (CBAM) is embedded between the down-sampling and up-sampling paths to enhance the network's perception of small objects.

(*3*) attention gates are introduced into skip connections to filter and fuse key feature information.

Experimental data confirm that the proposed method effectively improves segmentation accuracy of small tumors.

Parallel to this, transformer-based models such as ViT [15], TransUNet [16], UNETR [17], and Medical Transformer [18] have also achieved success in medical image segmentation by using self-attention to capture global context. Our proposed MGA-UNet complements transformer-based methods by focusing on a combination of lightweight multi-scale convolutions and attention gates, making it more suitable for resource-constrained environments.

Recent studies, including MSGU-Net [19], GA-UNet [20], and MGTUNet [21], have also attempted to incorporate multi-scale structures or attention mechanisms into U-shaped architectures. However, these works either lack simultaneous channel and spatial attention or do not include learnable gating mechanisms in skip connections, and the models are mainly applied to skin lesion or liver tumor segmentation.

Our MGA-UNet is the first to seamlessly integrate the multi-scale Ghost+RFB module, dual CBAM attention, and attention gates. This effectively addresses issues such as small object size and low boundary contrast in colorectal CT images, achieving significantly improved segmentation accuracy with low computational cost.

## 1. Methods

### 1.1. Improved convolutional module

Implementing segmentation tasks with convolutional neural networks involves significant computational complexity due to the generation of a large number of redundant feature maps. This redundancy manifests in two aspects:

(*1*) some generated maps contain minimal relevant information, leading to inefficient resource usage

(*2*) high similarity among feature maps (duplication) substantially increases processing time.

To optimize the generation of high-quality features while minimizing hardware costs, Han et al. developed the lightweight GhostNet architecture [13]. The key component is the Ghost module, which first generates a base set of feature maps using standard convolution and then applies

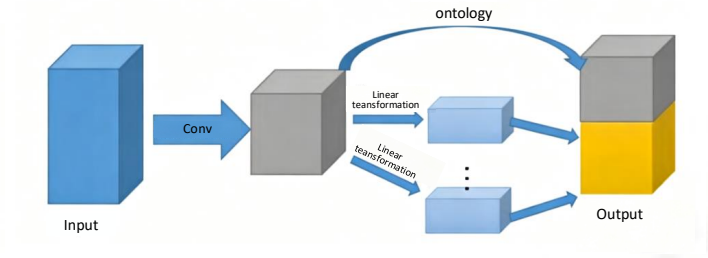



Figure 1. Structure of the Ghost convolution

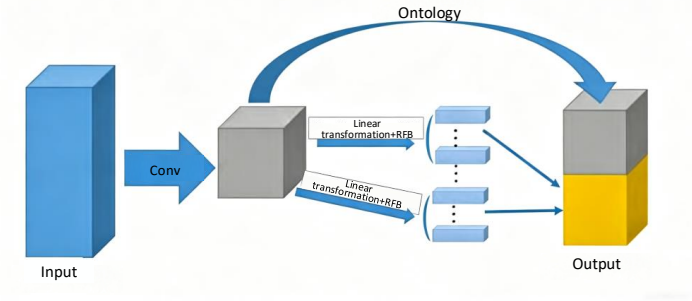



Figure 2. Improved Ghost+RFB convolution structure

cheap linear transformations to produce additional maps. This approach drastically reduces computational costs. The structure of the Ghost convolution is shown in Figure 1.

Compared to traditional convolution, the Ghost module effectively minimizes model complexity and reduces computational costs, accelerating training without significant loss of segmentation quality. However, the second stage of the Ghost module relies solely on linear transformations, limiting feature representativeness and diversity, potentially reducing final segmentation accuracy.

To endow the Ghost module with multi-scale feature extraction capability, we integrate a Receptive Field Block (RFB) [22]. The modified Ghost+RFB structure is shown in Figure 2.

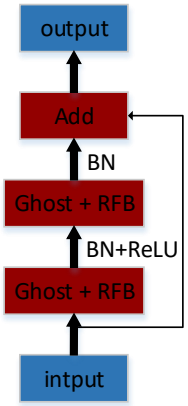



FIGURE 3. Improved convolutional block

The RFB block is implemented with four parallel branches: a $1 \times 1$ convolution and three types of atrous convolutions: $3 \times 3$ with dilation rate 1, $3 \times 3$ with dilation rate 3, and $5 \times 5$ with dilation rate 5. The outputs of all branches are concatenated along the channel dimension, then summed element-wise with the original output of the Ghost module. This design allows the model to simultaneously extract local details and contextual information over a broader receptive field.

Based on the improved Ghost+RFB module, we constructed an updated convolutional block designed to replace standard convolutional components in the classic U-Net architecture (see Figure 3).

Mathematically, for an input feature map $X$, traditional convolution uses $K$ filters to generate $M$ output maps, with computational complexity (FLOPs) $H \times W \times M \times k \times k \times C$, where $k$ is the kernel size. In contrast, the Ghost module adopts a two-stage strategy: first, it generates $M'$ intrinsic feature maps ($M' < M$); then it applies $s$ cheap linear transformations to each, synthesizing $M = M' \times s$ feature maps, reducing computational cost to approximately $1/s$ of standard convolution [**13**].

In our modification with the RFB block, the cheap transformations are replaced by atrous convolutions with varying dilation rates, substantially expanding the multi-scale receptive field [**23**].

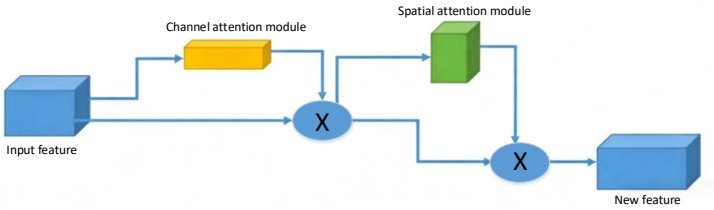



FIGURE 4. Structure of the CBAM module

As shown in Figure 3, the proposed convolutional block consists of two cascaded Ghost+RFB modules. A residual connection is integrated to mitigate degradation in deep models [24]. Batch normalization (BN) is applied after each Ghost+RFB module, while ReLU activation is used only after the first module. The output $Y$ is formalized as:

$$Y = \mathrm{ReLU}(G_2(G_1(X))) + X,$$

where $G_1, G_2$ denote the two cascaded Ghost+RFB modules.

### 1.2. CBAM attention mechanism

Attention mechanisms are effective tools for improving the performance of convolutional neural networks. By differential weighting, they suppress features from irrelevant regions and focus computational resources on the most informative data fragments. In this work, we integrate the Convolutional Block Attention Module (CBAM) [25], which sequentially extracts features along channel and spatial dimensions. By element-wise multiplication of the attention maps with the input features, adaptive feature refinement is achieved. The CBAM structure is shown in Figure 4.

For an intermediate feature map $F \in \mathbb{R}^{H\times W\times C}$, CBAM sequentially generates a 1D channel attention map $M_c \in \mathbb{R}^{1\times 1\times C}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{H\times W\times 1}$:

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F',$$

where $\otimes$ denotes the element-wise multiplication operation.

The channel attention mechanism aggregates spatial features using both average-pooling and max-pooling layers. The spatial attention block

generates adaptive weights based on global descriptors along the channel dimension. Integrating this dual-mode attention system allows the model to localize colorectal tumor regions with higher precision. Although the number of trainable parameters increases slightly, the gain in segmentation accuracy fully justifies the additional computational cost.

### 1.3. Modification of skip connections

To upgrade standard skip connections, we integrate the Attention Gate (AG) mechanism proposed by Oktay et al. [**26**]. The principle of AG is based on the joint analysis of the input feature vector and a gating signal, which undergo linear transformations for dimensionality alignment followed by element-wise summation. The result passes through a ReLU activation and a second linear transformation to reduce dimensionality. Then, using a sigmoid function, attention coefficients $\alpha$ are computed, which are used for element-wise weighting of the input feature vector $\hat{x}$.

The attention gate computation can be formalized as:

$$q = \Psi^T(\sigma_1(W_x^T x + W_g^T g + b_g)) + b_\psi,$$

where

$x$ are the encoder features passed through the skip connection,
$g$ is the gating signal from the decoder,
$\sigma_1$ is ReLU,
$\sigma_2$ is sigmoid,
$W_x, W_g, \Psi$ are learnable matrices,
$b_g, b_\psi$ are biases.

The attention gate can automatically adapt to different shapes and sizes of colorectal tumors, determining whether a pixel belongs to the region of interest via the attention coefficients, thereby suppressing irrelevant background areas and enhancing task-relevant features.

### 1.4. Network architecture

The architecture of the proposed model, shown in Figure 5, is based on the U-Net structure. The main modifications include replacing standard

convolutional blocks with improved Ghost+RFB modules, integrating attention gates into skip connections, and incorporating CBAM blocks in both the encoder and decoder paths.

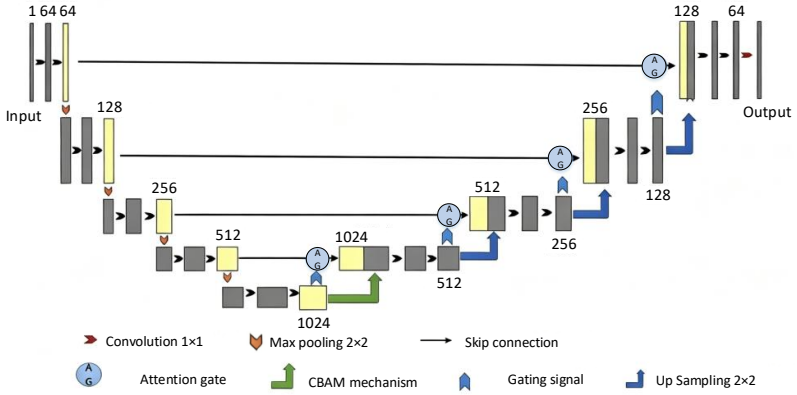



FIGURE 5. Architecture of the MGA-UNet network

In the classic U-Net, the compression path includes four down-sampling stages. The proposed model also implements four encoding levels, each consisting of two sequential Ghost+RFB modules. Within each module, input data first undergo a primary convolution, followed by cheap operations to generate redundant feature maps. In parallel, the RFB block uses kernels of various sizes to produce multi-scale features.

Batch normalization (BN) is applied for training stability, and ReLU activation is used only after the first Ghost+RFB module. The outputs of the two modules are combined through a residual connection. Each encoding stage ends with a max-pooling operation with stride 2.

The decoder contains four up-sampling operations. Each level includes a transposed convolution with stride 2 and two sequential Ghost+RFB modules, with normalization and ReLU applied between them. A residual block then merges shallow and deep features, preventing network degradation. Attention gates are added to the skip connections, assigning

different weights to features generated by the encoder and decoder, thereby highlighting information relevant to the colorectal tumor region.

The overall loss function is a combination of binary cross-entropy (BCE) and Dice loss:

$$L = \beta \cdot L_{\text{BCE}} + (1 - \beta) \cdot L_{\text{Dice}}, \quad \beta = 0.5,$$

where

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log p_i + (1 - y_i) \log(1 - p_i)],$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum y_i p_i + \epsilon}{\sum y_i + \sum p_i + \epsilon},$$

with

| | |
|---|---|
| $y_i$ | the true pixel label, |
| $p_i$ | the predicted probability, |
| $N$ | the total number of pixels, |
| $\beta$ | the balancing weight ($\beta = 0.5$ in this work), |
| $\epsilon$ | a small constant to avoid division by zero. |

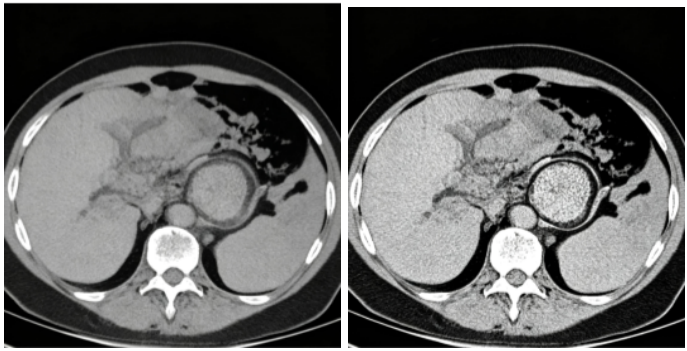
The combined loss function helps mitigate class imbalance and improves boundary delineation accuracy.

## 2. Experimental results and analysis

### 2.1. Dataset and preprocessing

The experimental studies were conducted on a CT image dataset of rectal cancer collected at the First Affiliated Hospital of Henan University of Science and Technology (Luoyang, China). The original sample includes 2D CT scans of 108 patients, of which 1693 images contain expert tumor annotations (masks). Data augmentation expanded the training set to 3057 images. Preprocessing steps included:

- Cropping: images were cropped to $512 \times 512$ centered on the tumor region.
- Intensity correction: Hounsfield unit range was limited to $[-200, 200]$ to improve soft tissue visualization [27].
- Contrast enhancement: histogram equalization was used to increase tumor tissue differentiation.



(*a*) before (*b*) after

Figure 6. Preprocessing of colorectal tumor CT images

- Normalization and resizing: data normalization minimized inter-individual variability, after which all scans were resized to $256 \times 256$ pixels.

The dataset was split into training, validation, and test sets in an 8:1:1 ratio. A visual comparison of images before and after preprocessing is shown in Figure 6.

### 2.2. Metrics and experimental setup

All experiments were implemented in Python 3.9 using PyTorch 1.12 [28] in the PyCharm IDE. Computations were performed on Windows 10 with an AMD Ryzen 5900X CPU and an NVIDIA GeForce RTX 3080 GPU (10 GB memory). Model weights were initialized using Kaiming Normal without pretrained weights.

Optimization used the AdamW algorithm with an initial learning rate of $10^{-4}$ and weight decay $10^{-5}$, batch size 8. The loss function was a weighted combination of BCE and Dice loss with $\beta = 0.5$. The learning rate schedule followed cosine annealing ($T_{max} = 100$ epochs), dropout rate was 0.5. Early stopping was applied: training stopped if the Dice metric on the validation set did not improve for 15 consecutive epochs.

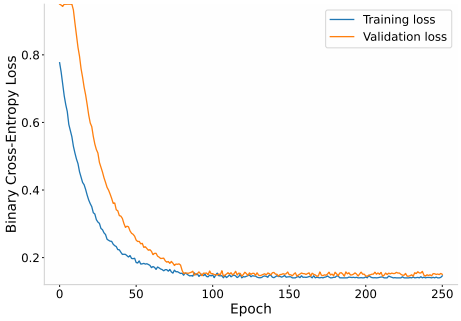



FIGURE 7. Training and validation loss curves

For objective verification of colorectal tumor segmentation effectiveness, we used the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). These metrics range in $[0, 1]$, with values near 1 indicating maximum agreement with ground truth. Here $A$ is the predicted mask, $B$ is the ground truth mask. IoU measures the relative overlap of two contours. The mathematical expressions are:

$$\text{Dice(A, B)} = \frac{2|A \cap B|}{|A| + |B|},$$

Here A is the predicted mask, B is the ground truth mask. IoU measures the relative overlap of two contours.

$$\text{IoU(A, B)} = \frac{|A \cap B|}{|A \cup B|}.$$

The dynamics of loss and Dice during training are shown in Figure 7 and Figure 8. For visualization of the global convergence trend, raw loss values were smoothed using a Savitzky–Golay filter (window=5, order=2). Analysis of the learning curves together with the statistics in Table 1 shows high convergence speed in the first 80 epochs.

The final loss on the training set stabilized at 0.14, and on the validation set around 0.15. The close proximity of the curves (gap less than 0.05) with no signs of oscillation or increase after epoch 80 confirms the high generalization ability of the model and absence of overfitting. The validation Dice reached a stable plateau around 96% by approximately epoch 80.

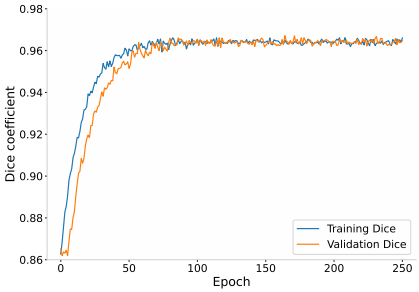



FIGURE 8. Change in Dice coefficient during training

TABLE 1. Raw loss statistics (unsmoothed)

| Epoch range | Train loss mean±std | Train max/min | Val loss mean±std | Val max/min |
|---|---|---|---|---|
| 1–20 | 0.47±0.12 | 0.85/0.31 | 0.52±0.10 | 0.98/0.38 |
| 21–40 | 0.28±0.06 | 0.37/0.19 | 0.31±0.05 | 0.39/0.24 |
| 41–60 | 0.19±0.03 | 0.24/0.14 | 0.22±0.04 | 0.28/0.16 |
| 61–80 | 0.16±0.02 | 0.18/0.11 | 0.17±0.03 | 0.22/0.13 |
| 81–100 | 0.14±0.01 | 0.15/0.10 | 0.15±0.02 | 0.18/0.12 |

The loss curves in Figure 7 are presented in smoothed form. To provide a complete description of the training process, Table 1 presents segmented statistics of raw (unsmoothed) loss values by epoch range. Analysis of the raw data also confirms stable convergence without anomalous jumps or sharp fluctuations.

The model code is available in an open repository: `https://github.com/Wangqian33/MGA-UNet`.

### 2.3. Comparative experiment results

To validate the effectiveness of the proposed algorithm, a comparative analysis was performed between MGA-UNet and baseline U-Net, as well as recent solutions: GhostNet, Attention U-Net, U-Net++, and the recently published Mamba-UNet and U-SAM. All compared models were trained from scratch without pretrained weights under identical hyperparameters.

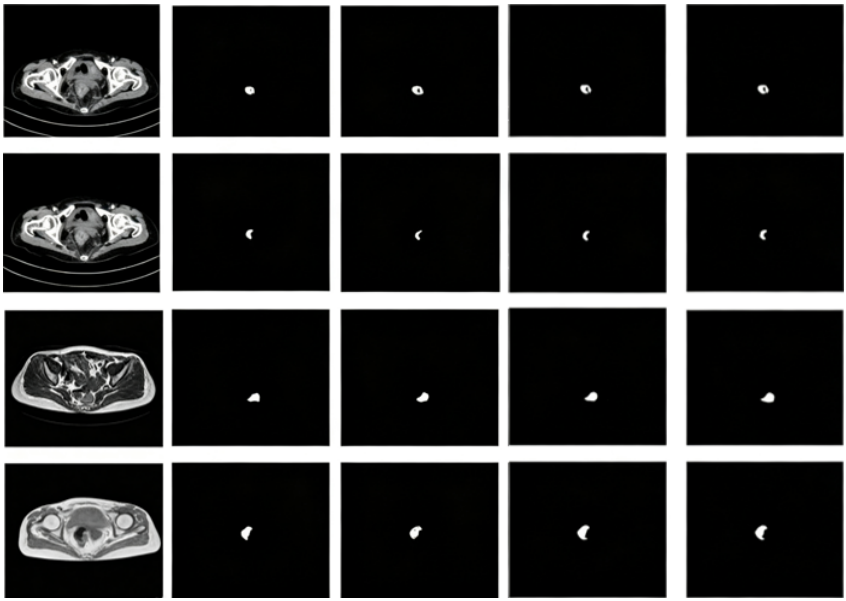


FIGURE 9. Visualization of segmentation results by different models: each row presents original image, U-Net, GhostNet, MGA-UNet, ground truth

Figure 9 demonstrates the colorectal tumor segmentation results. The first column contains original images, the second and third show results from U-Net and GhostNet respectively, the fourth shows results from our proposed model, and the fifth is the ground truth. Visual analysis shows that U-Net tends to over-segment and localizes boundaries less accurately. In contrast, MGA-UNet provides more precise tumor boundary segmentation, showing high agreement with the ground truth.

Quantitative results in Table 2 show that on our private dataset, the Dice metric of the proposed model exceeds that of U-Net, GhostNet, Attention U-Net, U-Net++, Mamba-UNet, and U-SAM by 4.45%, 1.52%, 0.87%, 0.64%, 1.32%, and 0.53%, respectively. For IoU, the improvements are 3.93%, 1.41%, 0.79%, 0.58%, 1.08%, and 0.47%, respectively. According to a paired t-test, the difference in Dice between MGA-UNet and baseline

Table 2. Performance comparison of different models (%)

| Model | Dice | IoU | Params (M) | FLOPs (G) | Time (ms) |
|---|---|---|---|---|---|
| U-Net [9] | 91.76 | 89.65 | 31.0 | 48.2 | 12.5 |
| GhostNet [13] | 94.69 | 92.17 | 8.5 | 12.3 | 5.8 |
| Attention U-Net [26] | 95.34 | 92.79 | 34.9 | 52.6 | 13.2 |
| U-Net++ [10] | 95.57 | 93.00 | 36.6 | 55.1 | 14.0 |
| Mamba-UNet [29] | 94.89 | 92.50 | 18.5 | 28.4 | 9.6 |
| U-SAM [30] | 95.68 | 93.11 | 12.4 | 17.8 | 8.4 |
| **MGA-UNet** | **96.21** | **93.58** | **10.2** | **15.6** | **7.2** |

Note: U-SAM (Zhang et al., 2025) is a SAM-based model; only adapter parameters are shown.

Table 3. Comparison with related architectures

| Model | Ghost | Multi-scale | CBAM | Attention gate | Params (M) | Application |
|---|---|---|---|---|---|---|
| MSGU-Net [19] | ✓ | ✓ (branches) | — | ✓ | ≈12 | Skin lesions |
| GA-UNet [20] | — | — | ✓ | ✓ | ≈15 | Liver tumors |
| MGTUNet [21] | ✓ | Transformer | — | ✓ | ≈25 | Multi-organ |
| **MGA-UNet** | ✓ | ✓ (RFB+atrous) | ✓ | ✓ | **10.2** | Colorectal cancer |

U-Net is statistically significant ($p < 0.01$). All evaluation metrics confirm the superiority of the proposed method over current alternatives.

For a more detailed justification of the architectural innovations, Table 3 provides a comparative analysis of the key modules of MGA-UNet against MSGU-Net, GA-UNet, and MGTUNet. As can be seen, MGA-UNet is the only architecture that simultaneously integrates multi-scale Ghost+RFB convolutions, the bimodal CBAM attention mechanism (channel and spatial), and attention gates. Moreover, the proposed model has the smallest number of trainable parameters, confirming its specialization and efficiency for colorectal CT image processing.

### 2.4. Ablation experiments

To evaluate the individual contribution of each proposed component to overall system performance, a series of ablation experiments was conducted. The baseline was the standard U-Net architecture, to which the improved

TABLE 4. Ablation experiment results (%)

| Configuration | Dice | IoU |
|---|---|---|
| Baseline U-Net | 91.76 | 89.65 |
| + improved convolutional module (Ghost+RFB) | 93.88 | 91.02 |
| + improved module + CBAM | 95.14 | 92.46 |
| + improved module + CBAM + attention gates (MGA-UNet) | **96.21** | **93.58** |

convolutional block (Ghost+RFB), CBAM attention, and attention gates in skip connections were added sequentially. Quantitative results are presented in Table 4.

According to the presented data, each integrated module made a positive contribution to segmentation accuracy. Introducing the multi-scale Ghost+RFB block increased Dice by 2.12% and IoU by 1.37%. Adding CBAM further increased Dice by 1.26% and IoU by 1.44%. Including attention gates contributed an additional 1.07% and 1.12%, respectively. The results indicate a synergistic effect among the components, with the attention gate mechanism having the most significant impact on tumor boundary detail [**26**]. A paired t-test confirmed that the improvements in Dice and IoU provided by each module are statistically significant ($p < 0.05$).

### 2.5. Hyperparameter sensitivity analysis

To determine the optimal configuration of the proposed model, a sensitivity analysis was performed on the weight coefficient $\beta$ in the combined loss function. As shown in Figure 10, experiments were conducted over $\beta \in [0, 1]$. The results demonstrate that the highest performance (Dice $\approx 96.2\%$) is achieved at $\beta = 0.5$. For small values ($\beta < 0.2$), the insufficient contribution of Dice loss hinders precise boundary segmentation, leading to lower metrics. Conversely, at excessively high values ($\beta > 0.8$), the reduced weight of binary cross-entropy also causes a moderate decline in accuracy.

Additionally, the influence of pooling methods and initial learning rate on model effectiveness was investigated. Replacing max-pooling with

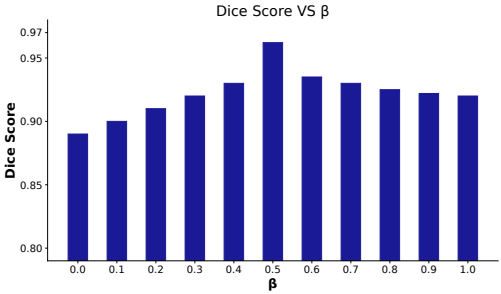



FIGURE 10. Dependence of Dice on the coefficient $\beta$

average-pooling during down-sampling led to a decrease in Dice of about 0.8%. Increasing the initial learning rate to $10^{-3}$ caused destabilization (oscillations) of the training process, reducing Dice by 1.2%. Therefore, this work justifies the use of max-pooling with stride 2 and a learning rate of $10^{-4}$.

Table 5 presents the results of an extended hyperparameter analysis, including weight decay, dropout probability, and learning rate scheduling strategies. Experiments confirm that deviations from the chosen configuration (lr=1e-4, $s = 4$, WD=1e-5, dropout=0.5, cosine annealing) reduce segmentation accuracy within 0.5%–1.5%. This indicates that the model has acceptable sensitivity to hyperparameter settings, maintaining consistently high performance near the optimum.

## 3. Conclusion

In this work, to address the insufficient accuracy of existing models for colorectal tumor segmentation, we proposed an improved MGA-UNet architecture based on U-Net. In the developed model, standard convolutional layers are replaced by a combination of Ghost and RFB modules, allowing the network to extract more diversified features during down-sampling. Integrating the CBAM attention mechanism between encoder and decoder focuses the model on key features, while embedding attention gates in skip connections enhances relevant local characteristics.

Experimental results on a private colorectal cancer CT dataset show that the proposed method outperforms existing analogues, achieving a Dice

TABLE 5. Extended hyperparameter sensitivity analysis (Dice %)

| Hyperparameter | Value | Dice (%) | FLOPs (G) | Time (ms) |
|---|---|---|---|---|
| Learning rate | $10^{-5}$ | 94.82 | 15.6 | 7.2 |
| | $5 \cdot 10^{-5}$ | 95.62 | 15.6 | 7.2 |
| | **$10^{-4}$** | **96.21** | 15.6 | 7.2 |
| | $5 \cdot 10^{-4}$ | 95.35 | 15.6 | 7.2 |
| Compression factor $s$ | 2 | 95.67 | 19.2 | 8.5 |
| | 3 | 96.01 | 17.1 | 7.8 |
| | **4** | **96.21** | 15.6 | 7.2 |
| | 5 | 95.98 | 14.2 | 6.7 |
| Weight decay | 0 | 95.83 | 15.6 | 7.2 |
| | **$10^{-5}$** | **96.21** | 15.6 | 7.2 |
| | $10^{-4}$ | 95.92 | 15.6 | 7.2 |
| Dropout | 0.0 | 96.24 | 15.6 | 7.2 |
| | 0.3 | 96.18 | 15.6 | 7.2 |
| | **0.5** | **96.21** | 15.6 | 7.2 |
| | 0.7 | 95.86 | 15.6 | 7.2 |
| Scheduler | Fixed | 95.44 | 15.6 | 7.2 |
| | Step | 95.91 | 15.6 | 7.2 |
| | **Cosine annealing** | **96.21** | 15.6 | 7.2 |

coefficient of 96.21%. It should be noted that this high Dice score was obtained on a limited sample (108 patients) after strict preprocessing (tumor centering, intensity windowing), so these results may require additional verification when extrapolated to more variable clinical data. With only 10.2M parameters and an inference time of 7.2 ms per image, the model is promising for use in real-time clinical diagnostic scenarios.

Directions for further research include:

- extending the proposed model to 3D CT volumes [31] with development of 3D versions of the Ghost and CBAM modules;
- implementing semi-supervised learning strategies [32] to leverage unlabeled data and improve network generalization;
- testing the model on public datasets to confirm its robustness.

## Acknowledgments

The authors thank the staff of the First Affiliated Hospital of Henan University of Science and Technology (Luoyang, China) for providing the data.

## References


[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, A. Jemal. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA: A Cancer Journal for Clinicians*, **74**:3 (2024), pp. 229–263. ↑148

[2] S. Zhao, S. Wang, P. Pan, T. Xia, X. Chang, X. Yang, L. Guo, Q. Meng, F. Yang, W. Qian, Z. Xu, Y. Wang, Z. Wang, L. Gu, R. Wang, F. Jia, J. Yao, Z. Li, Y. Bai. "Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis", *Gastroenterology*, **156**:6 (2019), pp. 1661–1674.e11. ↑148

[3] J. Long, E. Shelhamer, T. Darrell. "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA, USA, June 07–12, 2015), IEEE, 2015, ISBN 978-1-4673-6964-0, pp. 3431–3440. arXiv 1411.4038 ↑148

[4] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan. "Fully convolutional network for liver segmentation and lesions detection", *International Workshop on Deep Learning in Medical Image Analysis*, DLMIA 2016, LABELS 2016 (Athens, Greece, October 21, 2016), Lecture Notes in Computer Science (LNIP), vol. **10008**, Springer, Cham, 2016, ISBN 978-3-319-46975-1, pp. 77–85. arXiv 1606.06650 ↑148

[5] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, K. H. Maier-Hein. *nnU-Net: Self-adapting framework for U-Net-based medical image segmentation*, 2018, 11 pp. arXiv 1809.10486 ↑148

[6] T. Chen, Y. Son, A. Park, S.-J. Baek. "Baseline correction using a deep-learning model combining ResNet and UNet", *Analyst*, **147**:19 (2022), pp. 4285–4292. ↑148

[7] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, H. Meine. "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing", *Scientific Reports*, **8**:1 (2018), id. 15497, 7 pp. ↑148

[8] A. Iqbal, M. Sharif. "UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network", *Expert Systems with Applications*, **221**:1 (2023), id. 119718. URL ↑149

[9] O. Ronneberger, P. Fischer, T. Brox. "U-Net: Convolutional networks for biomedical image segmentation", *International Conference on Medical Image Computing and Computer-Assisted Intervention*. V. III, MICCAI (Munich, Germany, October 5–9, 2015), Lecture Notes in Computer Science (LNIP), vol. **9351**, Springer, Cham, 2015, ISBN 978-3-319-24573-7, pp. 234–241. arXiv 1505.04597 ↑149, 161

[10] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang. "UNet++: A nested U-Net architecture for medical image segmentation", *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, DLMIA 2018, ML-CDS 2018 (Granada, Spain, September 20, 2018), Lecture

Notes in Computer Science (LNIP), vol. **11045**, Springer, Cham, 2018, ISBN 978-3-030-00888-8, pp. 3–11. arXiv 1807.10165 ↑149, 161

[11] H. Seo, C. Huang, M. Bassenne, R. Xiao, L. Xing. "Modified U-Net (mU-Net) with incorporation of object-dependent high-level features for improved liver and liver-tumor segmentation in CT images", *IEEE Transactions on Medical Imaging*, **39**:5 (2020), pp. 1316–1325. ↑149

[12] Y. Wang, E. Lombardo, L. Huang, C. Belka, M. Riboldi, Ch. Kurz, G. Landry. "Head and neck cancer localization with Retina UNet for automated segmentation and time-to-event prognosis from PET/CT images", *Head and Neck Tumor Segmentation and Outcome Prediction*, HECKTOR 2022 (Singapore, September 22, 2022), Lecture Notes in Computer Science, vol. **13626**, Springer, Cham, 2023, ISBN 978-3-031-27419-0, pp. 202–211. ↑149

[13] K. Han, Y. Wang, Q. Tian, J. Guo, Chu. Xu, Cha. Xu. "GhostNet: More features from cheap operations", *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR (Seattle, WA, USA, June 13–19, 2020), IEEE, 2020, ISBN 978-1-7281-9360-1, pp. 1580–1589. arXiv 1911.11907 ↑149, 150, 152, 161

[14] N. Ma, X. Zhang, H.-T. Zheng, J. Sun. "ShuffleNet V2: Practical guidelines for efficient CNN architecture design", *Computer Vision — ECCV 2018*, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP), vol. **11218**, Springer, Cham, 2018, ISBN 978-3-030-01263-2, pp. 122–138. % arXiv 1807.11164 ↑149

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, Th. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby. "An image is worth 16 × 16 words: Transformers for image recognition at scale", *9th International Conference on Learning Representations*, ICLR 2021 (Virtual Event, Austria, May 3–7, 2021), 2021, ISBN 979-8-3313-0008-1, pp. 611, 21 pp. arXiv 2010.11929 URL ↑150

[16] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou. *TransUNet: Transformers make strong encoders for medical image segmentation*, 2021, 13 pp. arXiv 2102.04306 ↑150

[17] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, D. Xu. "UNETR: Transformers for 3D medical image segmentation", *2022 IEEE/CVF Winter Conference on Applications of Computer Vision*, WACV (Waikoloa, HI, USA, January 03–08, 2022), IEEE, 2022, ISBN 978-1-6654-0916-2, pp. 1748–1758. arXiv 2103.10504 ↑150

[18] P. A. Suxov, S. S. Danilyuk. "The use of transformers for segmentation of medical images", *Vestnik nauki*, **1**:6 (75) (2024), pp. 1539–1546, id. 230 (in Russian). URL ↑150

[19] H. Cheng, Y. Zhang, H. Xu, D. Li, Z. Zhong, Y. Zhao, Zh. Yan. "MSGU-Net: A lightweight multi-scale ghost U-Net for image segmentation", *Frontiers in Neurorobotics*, **18** (2025), id. 1480055. ↑150, 161

[20] B. Pang, L. Chen, Q. Tao, E. Wang, Y. Yu. "GA-UNet: A lightweight ghost and attention U-Net for medical image segmentation", *Journal of Imaging Informatics in Medicine*, **37**:4 (2024), pp. 1874–1888. ↑150, 161

[21] L. Pan, L. Wang, Zh. Feng, Zh. Xu, L. Xu, Sh. Peng. *MGTUNet: A new UNet for colon nuclei instance segmentation and quantification*, 2022, 5 pp. doi arXiv 2210.10981 ↑150, 161

[22] S. Liu, D. Huang, Y. Wang. "Receptive field block net for accurate and fast object detection", *Computer Vision — ECCV 2018*. V. XI, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP), vol. **11215**, 2018, ISBN 978-3-030-01251-9, pp. 404–419. doi arXiv 1711.07767 ↑151

[23] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille. "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**:4 (2018), pp. 834–848. doi arXiv 1606.00915 ↑152

[24] K. He, X. Zhang, S. Ren, J. Sun. "Deep residual learning for image recognition", *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR (Las Vegas, NV, USA, June 27–30, 2016), IEEE, 2016, ISBN 978-1-4673-8851-1, pp. 770–778. doi% doi arXiv 1512.03385 ↑153

[25] S. Woo, J. Park, J.-Y. Lee, I.-S. Kweon. "CBAM: Convolutional block attention module", *Computer Vision — ECCV 2018*, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP), vol. **11211**, Springer, Cham, 2018, ISBN 978-3-030-01233-5, pp. 3–19. doi% doi arXiv 1807.06521 ↑153

[26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert. *Attention U-Net: Learning where to look for the pancreas*, 2018, 10 pp. doi arXiv 1804.03999 ↑154, 161, 162

[27] G. Fei, B. Yan, J. Chen, K. Qiao, P. Ning, D. Shi. "Liver tumor segmentation based on dilated convolution of stacked tree aggregation structure", *Acta Optica Sinica*, **41**:18 (2021), id. 1810002. doi URL ↑156

[28] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala. "PyTorch: An imperative style, high-performance deep learning library", *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*, NeurIPS 2019 (Vancouver, BC, Canada, December 8–14, 2019), 2019, ISBN 9781713807933, pp. 8024–8035. arXiv 1912.01703 doi URL ↑157

[29] Z. Wang, J. Q. Zheng, Y. Zhang, G. Cui, L. Li. *Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation*, 2024, 12 pp. doi arXiv 2402.05079 ↑161

[30] H. Zhang, W. Guo, Sh. Wan, B. Zou, W. Wang, C. Qiu, K. Liu, P. Jin, J. Yang. "Tuning vision foundation models for rectal cancer segmentation from CT scans", *Communications Medicine*, **5**:1 (2025), id. 256, 11 pp. doi ↑161

[31] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger. "3D U-Net: Learning dense volumetric segmentation from sparse annotation", *International Conference on Medical Image Computing and Computer-Assisted*

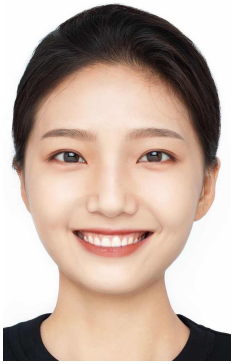
*Intervention*. V. II, MICCAI 2016 (Athens, Greece, October 17–21, 2016), Lecture Notes in Computer Science (LNIP), vol. **9901**, Springer, Cham, 2016, ISBN 978-3-319-46722-1, pp. 424–432. doi% doi arXiv 1606.06650 ↑164

[32] Y. Zhou, Y. Wang, P. Tang, S. Bai, W. Shen, E. K. Fishman, A. L. Yuille. "Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training", *2019 IEEE Winter Conference on Applications of Computer Vision*, WACV (Waikoloa, HI, USA, January 07–11, 2019), IEEE, 2019, ISBN 978-1-7281-1975-5, pp. 121–140. doi arXiv 1804.02586 ↑164

| | |
|---|---|
| *Received* | 19.04.2026; |
| *approved after reviewing* | 28.04.2026; |
| *accepted for publication* | 29.05.2026; |
| *published online* | 20.06.2026. |

Recommended by *PhD. A. N. Vinogradov*

## Information about the authors:





Yuqian Wang

PhD student at Tomsk Polytechnic University. Research interests: system analysis, information processing, medical informatics, deep learning, image segmentation.

0000-0003-3755-0586

*e-mail:* *wangyuqian3333@gmail.com*

Sergey Vladimirovich Aksenov

Associate Professor at Tomsk Polytechnic University, PhD in technical sciences. Research interests: system analysis, software systems, medical informatics.

0000-0002-1251-7133

*e-mail:* *axyonov@tpu.ru*

*The authors contributed equally to this article.*
*The authors declare no conflicts of interests.*

RU EN ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ ISSN 2079-3316
*научная статья* ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

УДК 004.93'11:616-073.75
10.25209/2079-3316-2026-17-2-147-190

# Интеграция многомасштабных признаков и механизмов внимания для сегментации колоректальных опухолей на компьютерных томограммах

Юйцянь **Ван**[1]✉, Сергей Владимирович **Аксёнов**[2]

[1,2] Школа информационных технологий и робототехники, Томский политехнический университет, Томск, Россия
[2] Томский государственный университет систем управления и радиоэлектроники, Томск, Россия

[1]✉*wangyuqian3333@gmail.com*

Аннотация. В последние годы технологии глубокого обучения получили широкое применение в области анализа медицинских изображений, демонстрируя выдающиеся результаты, особенно в задачах сегментации.

В связи с проблемой потери семантической информации на этапе извлечения признаков в U-образных сетях, что ограничивает точность сегментации опухолей прямой кишки, в данной работе на основе архитектуры U-Net предложена новая модель сегментации, получившая название MGA-UNet (Multi-scale Ghost Attention U-Net). Модель объединяет многомасштабное извлечение признаков, механизмы двойного (канального и пространственного) внимания и блоки управления вниманием (Attention Gate) в пропускных соединениях. Основные улучшения заключаются в следующем:

Во-первых, на этапе кодирования используется усовершенствованный модуль Ghost (интегрированный с RFB) для извлечения и слияния признаков на различных масштабах.

Во-вторых, в путь кодирования внедрён модуль внимания CBAM для усиления отклика сети на малоразмерные целевые объекты.

В-третьих, в пропускные соединения встроены блоки управления вниманием для подавления нерелевантных фоновых областей и выделения характеристик опухоли.

Результаты тестирования на наборе KT-данных колоректального рака показали высокую эффективность предложенной модели. По сравнению с классическими моделями U-Net, GhostNet, а также современными архитектурами Mamba-UNet и U-SAM, предложенная модель обеспечивает более точную локализацию границ опухоли прямой кишки и демонстрирует превосходное качество сегментации. Эффективность и стабильность каждого модуля подтверждены в ходе абляционных исследований и анализа чувствительности гиперпараметров. (*Связанные тексты статьи на английском и на русском языках*)

Ключевые слова и фразы: U-Net, механизм внимания, пропускные соединения, сегментация изображений, MGA-UNet

Благодарности: Работа выполнена при поддержке Китайского стипендиального совета (CSC) (грант № 202008410491).

Для цитирования: Ван Ю., Аксёнов С. В. *Интеграция многомасштабных признаков и механизмов внимания для сегментации колоректальных опухолей на компьютерных томограммах* // Программные системы: теория и приложения. 2026. **Т. 17**. № 2(71). С. 147–190. (*Англ.+русс.*) https://psta.psiras.ru/read/psta2026_2_147-190.pdf

## Введение

Согласно отчёту о мировой статистике рака за 2024 год, колоректальный рак занимает третье место в мире по показателям заболеваемости и второе место среди причин смертности от онкологических заболеваний [1]. Клинические характеристики заболевания тесно связаны с типом, локализацией, размером и количеством колоректальных полипов. Раннее выявление и удаление полипов играет ключевую роль в снижении заболеваемости колоректальным раком и значительно улучшает прогноз выживаемости пациентов.

В настоящее время колоноскопия является основным методом обнаружения колоректальных полипов. Однако результаты многочисленных систематических обзоров и мета-анализов показывают, что при колоноскопии сохраняется существенный риск пропуска полипов и аденом [2]. Таким образом, разработка технологий автоматизированной сегментации колоректальных опухолей не только способствует значительному повышению точности диагностики, но и снижает нагрузку на медицинский персонал, содействуя продвижению программ раннего скрининга и вмешательства.

В последние годы стремительное развитие глубокого обучения обеспечило мощную технологическую поддержку для автоматической сегментации медицинских изображений.

Shelhamer и др. предложили полносвёрточную сеть (FCN), в которой полносвязные слои традиционных CNN заменены свёрточными слоями, а для восстановления разрешения изображений используются операции деконволюции (обратной свёртки), при этом вводятся пропускные соединения (skip connections) для слияния семантической информации мелких и глубоких слоев, что позволяет получать более детализированные результаты сегментации [3].

Ben-Cohen и др. впервые применили FCN для сегментации печени и обнаружения ее повреждений, достигнув среднего коэффициента Dice 0,89 без использования сложных этапов предобработки, что превзошло результаты традиционных CNN [4].

Isensee и др. разработали метод nnU-Net с автоматической конфигурацией, способный адаптивно настраиваться под различные задачи сегментации медицинских изображений [5].

Chen и др. представили сеть EfficientNet-Lite UNet для сегментации биомедицинских изображений, которая обеспечивает высокое качество сегментации при экономии вычислительных ресурсов [6].

Schenk и др. усовершенствовали архитектуру FCN, введя длинные и короткие пропускные соединения для передачи карт признаков из пути сжатия в путь расширения, что позволило восстановить детали, утерянные в процессе субдискретизации, и ускорить сходимость обучения [7].

Iqbal и Sharif предложили метод полуавтоматической сегментации опухолей молочной железы на основе U-образной пирамидальной сети с расширенными свёртками (dilation networks), повысив эффективность сегментации за счет совместного использования размеченных и неразмеченных данных [8].

Ronneberger и др. представили классическую модель U-Net с симметричной структурой кодировщика-декодировщика, которая достигла отличных результатов сегментации благодаря интеграции карт признаков через пропускные соединения [9].

Zhou и др. на основе U-Net разработали U-Net++, в которой за счёт реорганизации пропускных соединений извлекается более богатая иерархическая информация, что минимизирует семантический разрыв между признаками на этапах повышения и понижения дискретизации [10].

Seo и др. предложили сеть mu-U-Net, добавив дополнительные слои деконволюции и функции активации в пропускные соединения для одновременного извлечения глобальных признаков высокого уровня малых объектов и информации о границах высокого разрешения крупных объектов [11].

Wang и др. разработали модель Retina UNet для локализации опухолей головы и шеи на изображениях ПЭТ/КТ, которая не только повышает точность сегментации, но и прогнозирует время выживания пациентов [12].

Кроме того, лёгкие свёрточные сети, такие как GhostNet [13] и ShuffleNetV2 [14], способны значительно снизить вычислительные затраты при сохранении высокой точности, предлагая новые подходы для клинического внедрения сегментации медицинских изображений.

В задачах сегментации колоректальных опухолей колебания качества КТ-изображений затрудняют точную идентификацию границ небольших новообразований. Для решения этой проблемы в данной работе предложен метод сегментации, сочетающий многомасштабные свёртки и механизмы канально-пространственного внимания:

(*1*) для извлечения многоуровневых признаков традиционные свёрточные операторы заменены модулем многомасштабной свёртки (Ghost+RFB).

(*2*) между путями понижения и повышения дискретизации встроен механизм пространственно-канального внимания (CBAM) для усиления способности сети воспринимать малые объекты.

(*3*) в пропускные соединения внедрены блоки управления вниманием (Attention Gate) для фильтрации и слияния ключевой признаковой информации.

Экспериментальные данные подтверждают, что предложенный метод эффективно повышает точность сегментации небольших опухолей.

Параллельно с этим модели на основе трансформеров, такие как ViT [15], TransUNet [16], UNETR [17] и Medical Transformer [18], также достигли успеха в сегментации медицинских изображений, используя механизмы самовнимания для захвата глобального контекста.

Предложенная в данной работе модель MGA-UNet дополняет методы на основе трансформеров, фокусируясь на сочетании легковесных многомасштабных свёрток и блоков управления вниманием, что делает ее более подходящей для сред с ограниченными вычислительными ресурсами.

Недавние исследования, такие как MSGU-Net [19], GA-UNe [20] и MGTUNet [21], также предпринимали попытки внедрения многомасштабных структур или механизмов внимания в U-образную архитектуру. Однако в этих работах либо отсутствует одновременное сочетание канального и пространственного внимания, либо не предусмотрены обучаемые механизмы управления в пропускных соединениях, а сами модели в основном применяются для сегментации поражений кожи или опухолей печени.

Предложенная модель MGA-UNet впервые обеспечивает бесшовную интеграцию многомасштабного модуля Ghost+RFB, двойного внимания CBAM и блоков Attention Gate. Это позволяет эффективно решать такие проблемы КТ-изображений колоректальных опухолей, как малый размер объектов и низкая контрастность границ, обеспечивая значительное повышение точности сегментации при низких вычислительных затратах.

## 1. Методы

### 1.1. Улучшенный свёрточный модуль

Реализация задач сегментации с использованием свёрточных нейронных сетей сопряжена со значительной вычислительной трудоёмкостью, что обусловлено генерацией большого объема избыточных карт признаков. Избыточность проявляется в двух аспектах:

(*1*) часть сформированных карт содержит минимальный объем релевантной информации, что ведёт к нерациональному расходу ресурсов.

(*2*) высокая степень сходства между отдельными картами признаков (дублирование) существенно увеличивает время обработки данных.

С целью оптимизации процесса генерации высококачественных признаков при минимизации аппаратных затрат Han и др. разработали облегчённую архитектуру GhostNet [13]. Ключевым элементом данной сети является модуль Ghost, принцип работы которого основан на первоначальном формировании базового набора карт признаков посредством стандартной свёртки с последующим применением экономичных линейных преобразований для получения их аналогов. Такой подход позволяет

радикально сократить вычислительные издержки (структурная схема представлена на рисунке 1).

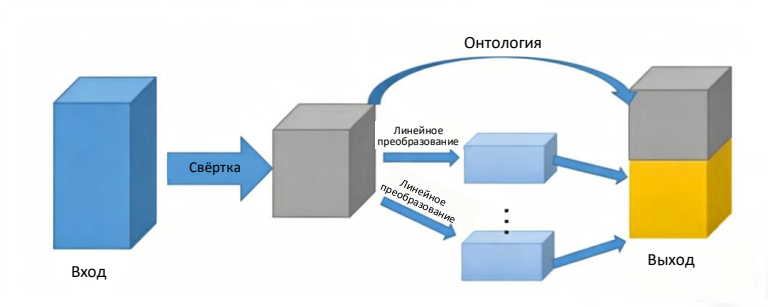



Рисунок 1. Структура свёртки Ghost

В сравнении с традиционными свёрточными операциями, модуль Ghost позволяет эффективно минимизировать структурную сложность модели и снизить вычислительные затраты, ускоряя процесс обучения без существенной потери качества сегментации. Тем не менее, на втором этапе работы модуля Ghost генерация карт признаков детерминирована исключительно линейными преобразованиями, что ограничивает репрезентативность и разнообразие признаков, потенциально снижая итоговую точность сегментации.

Для наделения модуля Ghost способностью к извлечению многомасштабных признаков в настоящей работе интегрирован блок рецептивного поля (RFB) [22]. Модифицированная структура свёртки Ghost, предложенная авторами, представлена на рисунке 2.

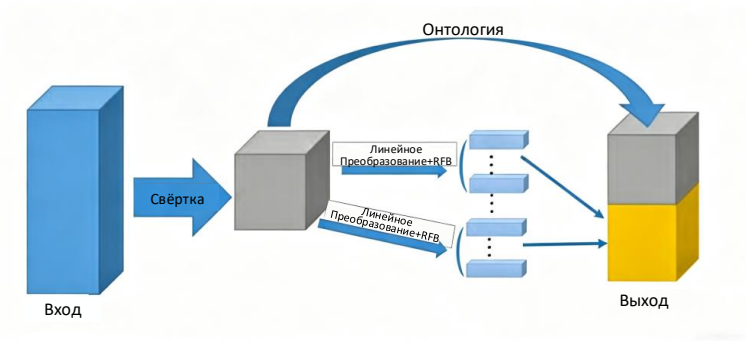



Рисунок 2. Улучшенная свёрточная структура Ghost+RFB

Архитектура блока RFB реализована в виде четырёх параллельных ветвей, включающих свёртку1 × 1, а также три типа расширенных (atrous) свёрток: 3 × 3 с коэффициентом расширения (dilation rate) 1, 3 × 3 с коэффициентом 3 и 5 × 5 с коэффициентом 5. Выходные данные всех ветвей объединяются по канальной размерности (конкатенация), после чего выполняется их поэлементное суммирование с исходным выходом модуля Ghost. Подобная проектная стратегия позволяет модели одновременно экстрагировать локальные детали и учитывать контекстуальную информацию в более широком диапазоне рецептивного поля.

На основе модернизированного модуля Ghost+RFB в данной работе сконструирован обновлённый свёрточный блок, предназначенный для замены стандартных свёрточных компонентов в классической архитектуре U-Net (подробная схема представлена на рисунке 3).

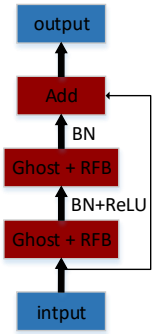



Рисунок 3. Улучшенная свёрточная структура

В математическом представлении, при входной карте признаков $X$, традиционная свёртка задействует $K$ фильтров для генерации $M$ выходных карт. Вычислительная сложность данного процесса (FLOPs) выражается формулой $H \times W \times M \times k \times k \times C$, где $k$ обозначает пространственный размер ядра свёртки. В отличие от этого, модуль Ghost реализует двухэтапную стратегию: на первом этапе формируется $M'$ базовых карт признаков ($M' < M$), а на втором – к каждой из них применяется $s$ экономичных линейных преобразований. В результате синтезируется $M = M' \times s$ карт признаков, при этом вычислительные затраты сокращаются примерно до $1/s$ от объема стандартной свёртки [**13**].

В предложенной модификации с блоком RFB этап экономичных преобразований заменён расширенными свёртками с варьируемыми коэффициентами, что существенно расширяет многомасштабное рецептивное поле [**23**].

Как показано на рисунке 3, предложенный свёрточный блок состоит из двух последовательно соединённых модулей Ghost+RFB. Для нивелирования проблемы деградации глубоких моделей в структуру интегрирована остаточная связь [24]. В процессе реализации после каждого модуля Ghost+RFB применяется пакетная нормализация (BN), при этом функция активации ReLU задействована исключительно после первого модуля. Выход данной структуры $Y$ формализуется следующим образом:

$$Y = \mathrm{ReLU}(G_2(G_1(X))) + X$$

где $G_1, G_2$ обозначают каскадную операцию двух модулей Ghost+RFB.

### 1.2. Механизм внимания CBAM

Механизмы внимания являются эффективным инструментом повышения производительности свёрточных нейронных сетей. За счет дифференцированного взвешивания они позволяют подавлять признаки нерелевантных областей, фокусируя вычислительные ресурсы на наиболее информативных фрагментах данных. В данной работе интегрирован модуль внимания CBAM (Convolutional Block Attention Module) [25], который последовательно экстрагирует признаки в канальном и пространственном измерениях. Путем поэлементного умножения полученных карт внимания с входными признаками достигается адаптивное уточнение (refinement) признаковых представлений. Структурная схема модуля CBAM представлена на рисунке 4.

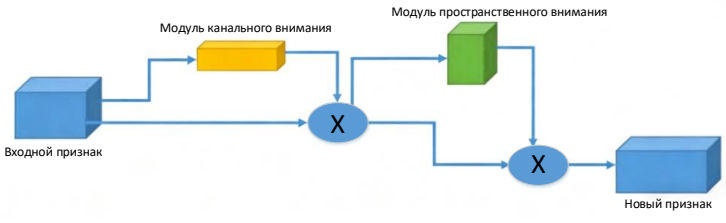



Рисунок 4. Структура модуля CBAM

Для промежуточной карты признаков $F \in \mathbb{R}^{H\times W\times C}$ модуль CBAM последовательно генерирует одномерную карту канального внимания $M_c \in \mathbb{R}^{1\times 1\times C}$ и двумерную карту пространственного внимания $M_s \in \mathbb{R}^{H\times W\times 1}$. Процесс обработки описывается следующими выражениями:

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F'$$

где $\otimes$ обозначает операцию поэлементного умножения.

Механизм канального внимания осуществляет агрегацию пространственных признаков посредством совместного использования слоев

усредняющего (Average Pooling) и максимального (Max Pooling) пула. В свою очередь, блок пространственного внимания генерирует адаптивные весовые коэффициенты, опираясь на глобальные дескрипторы вдоль канальной размерности. Интеграция подобной бимодальной системы внимания позволяет модели с более высокой точностью локализовать области колоректальных новообразований. Несмотря на незначительное увеличение количества обучаемых параметров, прирост точности сегментации, обеспечиваемый модулем CBAM, полностью оправдывает дополнительные вычислительные затраты.

### 1.3. Модификация пропускных соединений

Для модернизации стандартных пропускных соединений в данной работе интегрирован механизм управления вниманием (Attention Gate, AG), предложенный Oktay и др. [26]. Принцип работы AG основан на совместном анализе вектора входных признаков и стробирующего сигнала (gating signal), которые подвергаются линейному преобразованию для выравнивания размерности с последующим поэлементным суммированием. Полученный результат проходит через функцию активации ReLU и вторичное линейное преобразование для снижения размерности. Далее, с помощью сигмоидной функции рассчитываются коэффициенты внимания $\alpha$, которые используются для поэлементного взвешивания входного вектора признаков$\hat{x}$.

Вычисление вентиля внимания можно формализовать следующим образом:

$$q = \Psi^T(\sigma_1(W_x^T x + W_g^T g + b_g)) + b_\psi,$$

где

$x$ – признаки кодировщика, передаваемые по пропускному соединению,
$g$ – управляющий сигнал из декодировщика,
$\sigma_1$ – ReLU,
$\sigma_2$ – сигмоид,
$W_x, W_g, \Psi$ – обучаемые матрицы,
$b_g, b_\Psi$ – базовые смещения.

Вентиль внимания способен автоматически подстраиваться под различные формы и размеры колоректальных опухолей, определяя по коэффициентам внимания, принадлежит ли пиксель области интереса, тем самым подавляя нерелевантные фоновые области и усиливая полезные для конкретной задачи признаки.

### 1.4. Архитектура сети

Архитектура разработанной модели, представленная на рисунке 5. базируется на структуре U-Net. Основные модификации включают

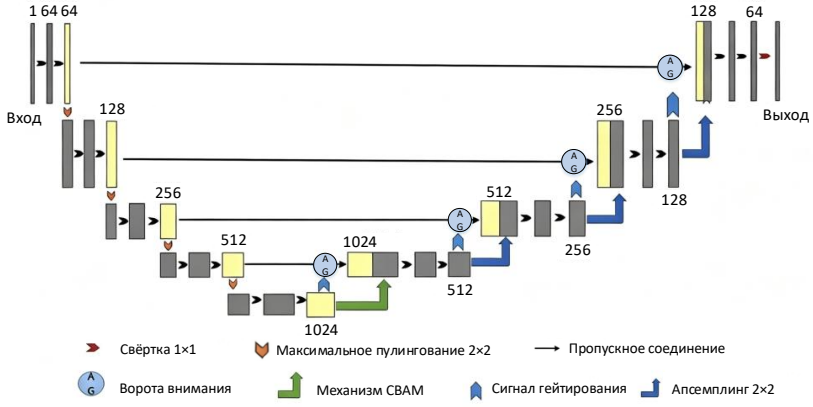



Рисунок 5. Архитектура сети MGA-UNet

замену стандартных свёрточных блоков на усовершенствованные модули Ghost+RFB, интеграцию механизмов управления вниманием (Attention Gate) в пропускные соединения, а также внедрение блоков CBAM в пути кодирования и декодирования.

В классической сети U-Net процесс сжатия включает четыре этапа субдискретизации. В предложенной модели также реализованы четыре уровня кодирования, однако каждый из них состоит из двух последовательных модулей Ghost+RFB. В рамках каждого модуля входные данные сначала подвергаются первичной свёртке, после чего применяются «экономичные» операции для генерации избыточных карт признаков. Параллельно с этим блок RFB, использующий ядра различных размеров, формирует многомасштабные признаки.

Для стабилизации обучения применяется пакетная нормализация (BN), а функция активации ReLU задействована только после первого модуля Ghost+RFB. Результаты двух модулей объединяются через остаточную связь (residual connection). Завершается каждый этап кодирования операцией максимального пула (Max Pooling) с шагом 2.

В декодировщике содержатся четыре операции повышающей дискретизации. Каждый уровень включает транспонированную свёртку с шагом 2 и два последовательных модуля Ghost+RFB, между которыми применяются нормализация и ReLU. Затем остаточный блок объединяет мелкие и глубокие признаки, предотвращая деградацию сети. В пропускные соединения добавлен механизм внимания, который присваивает различные веса признакам, генерируемым кодировщиком и декодировщиком, тем

самым выделяя информацию, относящуюся к области колоректальной опухоли.

Функция потерь всей сети представляет собой комбинацию бинарной кросс-энтропии (BCE) и Dice-потери:

$$L = \beta \cdot L_{\text{BCE}} + (1 - \beta) \cdot L_{\text{Dice}}, \quad \beta = 0.5,$$

где

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log p_i + (1 - y_i) \log(1 - p_i)],$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum y_i p_i + \epsilon}{\sum y_i + \sum p_i + \epsilon}$$

где

$y_i$ – истинная метка пикселя,
$p_i$ – предсказанная вероятность,
$N$ – общее число пикселей,
$\beta$ – балансировочный вес (в данной работе $\beta$=0.5),
$\epsilon$ – малая константа для предотвращения деления на ноль.

Комбинированная функция потерь способствует смягчению проблемы дисбаланса классов и повышению точности выделения границ.

## 2. Экспериментальные результаты и анализ

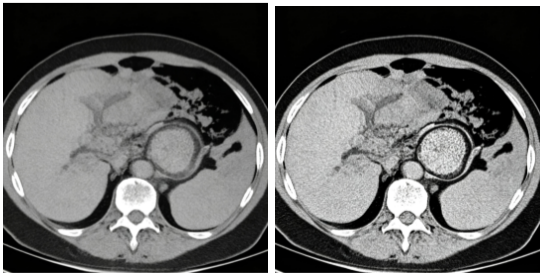
### 2.1. Набор данных и предварительная обработка

Экспериментальные исследования в данной работе проводились на базе набора данных КТ-изображений рака прямой кишки, собранного в Первой аффилированной больнице Университета науки и технологий Хэнань (Лоян, Китай). Исходная выборка включает 2D КТ-сканы 108 пациентов, из которых 1693 изображения содержат экспертную разметку (маски) опухолей. В результате применения методов аугментации данных объем обучающей выборки был расширен до 3057 изображений.Процедура предварительной обработки включала следующие этапы:

- Кадрирование: изображения были обрезаны до размера $512 \times 512$ с центрированием по области новообразования.
- Коррекция интенсивности: диапазон значений Хаунсфилда (HU) был ограничен пределами $[-200, 200]$ для улучшения визуализации мягких тканей [27].
- Улучшение контрастности: для повышения дифференциации опухолевой ткани использовался метод выравнивания гистограммы (histogram equalization).

- Нормировка: для минимизации межиндивидуальной вариативности применена нормализация данных, после чего разрешение всех сканов было приведено к стандарту $256 \times 256$ пикселей.

Набор данных был разделен на обучающую, валидационную и тестовую выборки в соотношении $8 : 1 : 1$ соответственно. Визуальное сравнение изображений до и после этапа предобработки представлено на рисунке 6.



(*а*) до (*б*) после

Рисунок 6. Предобработка КТ-изображений колоректальной опухоли

## 2.2. Метрики и настройки эксперимента

Все экспериментальные исследования реализованы на языке Python 3.9 с использованием фреймворка PyTorch 1.12 [28] в среде разработки PyCharm. Вычисления проводились под управлением ОС Windows 10 на аппаратной платформе, включающей процессор AMD Ryzen 5900X и графический ускоритель NVIDIA GeForce RTX 3080 (10 ГБ видеопамяти). Инициализация весов модели осуществлялась по методу Кайминга (Kaiming Normal) без использования предобученных весов.

Для оптимизации применялся алгоритм AdamW с начальной скоростью обучения $10^{-4}$ и коэффициентом затухания весов $10^{-5}$ при размере пакета (batch size) 8. В качестве функционала потерь использована взвешенная комбинация бинарной кросс-энтропии (BCE) и потерь Dice с коэффициентом $\beta = 0.5$. Изменение скорости обучения регулировалось по стратегии косинусного отжига ($T_max = 100$ эпох), коэффициент Dropout составил 0.5. Для предотвращения переобучения применена стратегия ранней остановки (Early Stopping): обучение прекращалось, если метрика Dice на валидационном наборе не демонстрировала роста в течение 15 последовательных эпох.

Для объективной верификации эффективности сегментации колоректальных опухолей использованы коэффициент сходства Дайса (Dice Similarity Coefficient, DSC) и индекс пересечения над объединением (Intersection over Union, IoU). Данные метрики варьируются в диапазоне $[0, 1]$, где значения, близкие к 1, свидетельствуют о максимальном совпадении результатов сегментации с экспертной разметкой (Ground Truth). Математические выражения для расчета имеют следующий вид:

$$\mathrm{Dice(A,\,B)} = \frac{2|A \cap B|}{|A| + |B|}$$

где A – предсказанная маска, B – эталонная маска. IoU измеряет относительную величину перекрытия двух контуров:

$$\mathrm{IoU(A,\,B)} = \frac{|A \cap B|}{|A \cup B|}$$

Динамика функции потерь и метрики Dice в процессе обучения представлена на рисунках 7 и 8. Для визуализации глобального тренда

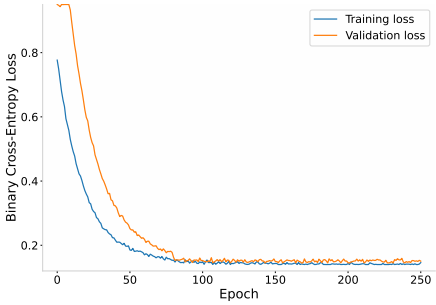



Рисунок 7. Кривые потерь на обучении и валидации

сходимости исходные значения потерь были сглажены с помощью фильтра Савицкого-Голея (окно=5, порядок=2). Анализ кривых обучения в совокупности со статистическими данными таблица 1 показывает высокую скорость сходимости модели на первых 80 эпохах.

Итоговое значение функции потерь на обучающей выборке стабилизировалось на уровне 0.14, на валидационной – около 0.15. Плотное прилегание кривых (разрыв менее 0.05) без признаков осцилляции или роста после 80-й эпохи подтверждает высокую обобщающую способность модели и отсутствие эффекта переобучения. Значение Dice на валидационном наборе достигло стабильного плато на уровне 96% примерно к 80-й эпохе.

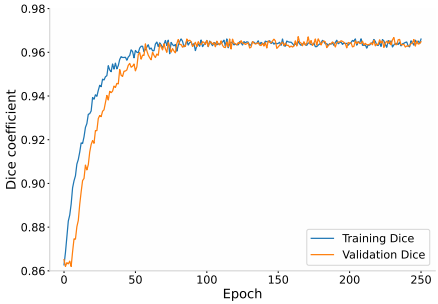



Рисунок 8. Изменение коэффициента Dice в процессе обучения

Таблица 1. Исходная статистика потерь (несглаженная)

| Диапазон эпох | Среднее ± std обуч. | Макс/мин обуч. | Среднее ± std вал. | Макс/мин вал. |
|---|---|---|---|---|
| 1–20 | 0.47±0.12 | 0.85/0.31 | 0.52±0.10 | 0.98/0.38 |
| 21–40 | 0.28±0.06 | 0.37/0.19 | 0.31±0.05 | 0.39/0.24 |
| 41–60 | 0.19±0.03 | 0.24/0.14 | 0.22±0.04 | 0.28/0.16 |
| 61–80 | 0.16±0.02 | 0.18/0.11 | 0.17±0.03 | 0.22/0.13 |
| 81–100 | 0.14±0.01 | 0.15/0.10 | 0.15±0.02 | 0.18/0.12 |

Кривые потерь на рисунке 7 представлены в сглаженном виде. Для обеспечения полноты описания процесса обучения в таблице 1 приведена сегментированная статистика исходных (необработанных) значений функции потерь по эпохам. Анализ первичных данных также подтверждает стабильную сходимость модели без аномальных скачков или резких флуктуаций.

Код модели доступен в открытом репозитории: https://github.com/Wangqian33/MGA-UNet.

### 2.3. Результаты сравнительных экспериментов

Для валидации эффективности предложенного алгоритма был проведён сравнительный анализ модели MGA-UNet с базовой архитектурой U-Net, а также современными решениями: GhostNet, Attention U-Net, U-Net++ и недавно опубликованными моделями Mamba-UNet и U-SAM. Все сопоставляемые модели обучались «с нуля» без использования предобученных весов при идентичных гиперпараметрах.

На рисунке 9 продемонстрированы результаты сегментации колоректальных опухолей. Первый столбец содержит исходные изображения,

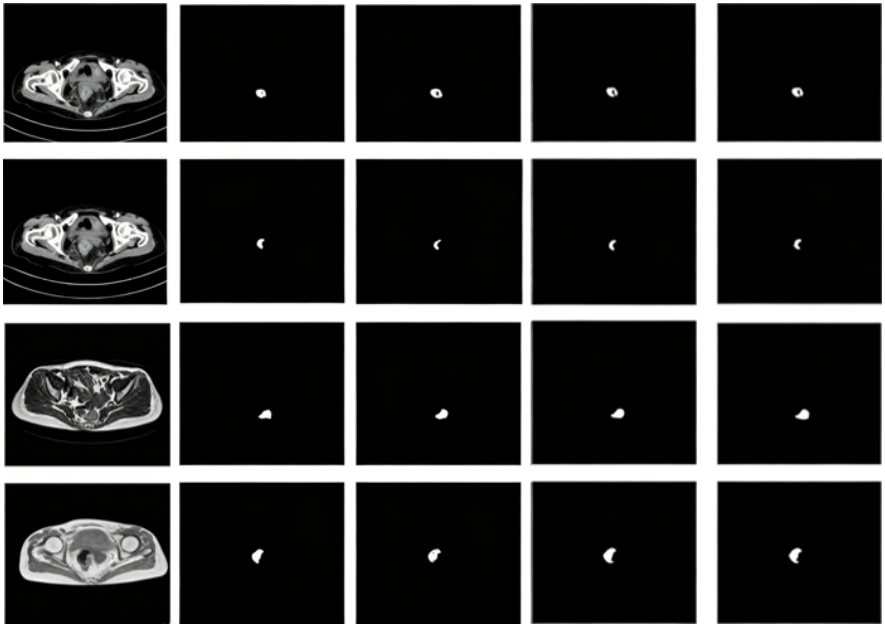


Рисунок 9. Визуализация результатов сегментации разными моделями: каждый ряд представляет исходное изображение, U-Net, GhostNet, MGA-UNet, эталон

второй и третий – результаты U-Net и GhostNet соответственно, четвёртый – результаты предложенной модели, пятый – эталонную разметку (Ground Truth). Визуальный анализ показывает, что сеть U-Net склонна к избыточной сегментации (over-segmentation) и менее точной локализации границ. В свою очередь, MGA-UNet обеспечивает прецизионную сегментацию контуров опухоли, демонстрируя высокую степень соответствия эталону.

Количественные показатели эксперимента приведены в таблице 2. На частном наборе данных метрика Dice предложенной модели превосходит аналогичные показатели U-Net, GhostNet, Attention U-Net, U-Net++, Mamba-UNet и U-SAM на 4.45%, 1.52%, 0.87%, 0.64%, 1.32% и 0.53% соответственно. По показателю IoU прирост составил 3.93%, 1.41%, 0.79%, 0.58%, 1.08% и 0.47% соответственно. Согласно результатам парного t-теста, различие в значениях Dice между MGA-UNet и базовой U-Net

Таблица 2. Сравнение производительности различных моделей (%)

| Модель | Dice | IoU | Params (M) | FLOPs (G) | Время (мс) |
|---|---|---|---|---|---|
| U-Net [9] | 91.76 | 89.65 | 31.0 | 48.2 | 12.5 |
| GhostNet [13] | 94.69 | 92.17 | 8.5 | 12.3 | 5.8 |
| Attention U-Net [26] | 95.34 | 92.79 | 34.9 | 52.6 | 13.2 |
| U-Net++ [10] | 95.57 | 93.00 | 36.6 | 55.1 | 14.0 |
| Mamba-UNet [29] | 94.89 | 92.50 | 18.5 | 28.4 | 9.6 |
| U-SAM [30] | 95.68 | 93.11 | 12.4 | 17.8 | 8.4 |
| **MGA-UNet** | **96.21** | **93.58** | **10.2** | **15.6** | **7.2** |

Примечание: U-SAM (Zhang et al., 2025) – модель на основе SAM, в таблице приведены параметры только адаптера.

является статистически значимым ($p < 0.01$). Все оценочные метрики подтверждают превосходство предложенного метода над современными аналогами.

Для более детального обоснования архитектурных инноваций предложенной модели в таблице 3 приведён сравнительный анализ ключевых

Таблица 3. Сравнение с близкими архитектурами

| Модель | Ghost | Многомасштабность | CBAM | Вентиль внимания | Params (M) | Область |
|---|---|---|---|---|---|---|
| MSGU-Net [19] | ✓ | ✓ (разные ветви) | – | ✓ | ≈12 | Кожные поражения |
| GA-UNet [20] | – | – | ✓ | ✓ | ≈15 | Опухоли печени |
| MGTUNet [21] | ✓ | Transformer | – | ✓ | ≈25 | Мульти-органная |
| **MGA-UNet** | ✓ | ✓ (RFB + расшир.) | ✓ | ✓ | **10.2** | Колоректальный рак |

модулей MGA-UNet в сравнении с сетями MSGU-Net, GA-UNet и MGTUNet. Как следует из сопоставления, модель MGA-UNet является единственной архитектурой, в которой одновременно интегрированы многомасштабные свёртки Ghost+RFB, бимодальный механизм внимания CBAM (канальный и пространственный) и блоки управления вниманием (Attention Gates). При этом разработанная модель обладает минимальным количеством обучаемых параметров, что подтверждает её специализацию и эффективность при обработке КТ-изображений колоректальных опухолей.

### 2.4. Абляционные эксперименты

Для оценки индивидуального вклада каждого из предложенных компонентов в общую производительность системы был проведён ряд

абляционных экспериментов. В качестве базовой линии (baseline) использовалась стандартная архитектура U-Net, к которой последовательно добавлялись: усовершенствованный свёрточный блок (Ghost+RFB), механизм внимания CBAM и блоки управления вниманием (Attention Gate) в пропускных соединениях. Количественные результаты представлены в таблице 4.

Таблица 4. Результаты абляционных экспериментов (%)

| Конфигурация | Dice | IoU |
|---|---|---|
| Базовый U-Net | 91.76 | 89.65 |
| + улучшенный свёрточный модуль (Ghost+RFB) | 93.88 | 91.02 |
| + улучшенный модуль + CBAM | 95.14 | 92.46 |
| + улучшенный модуль + CBAM + вентили внимания (MGA-UNet) | **96.21** | **93.58** |

Согласно представленным данным, каждый из интегрированных модулей внёс положительный вклад в точность сегментации. Внедрение многомасштабного свёрточного блока Ghost+RFB обеспечило прирост метрики Dice на 2,12% и IoU на 1,37%. Добавление механизма CBAM позволило дополнительно увеличить Dice на 1,26% и IoU на 1,44%. Включение блоков Attention Gate способствовало дальнейшему росту показателей на 1,07% и 1,12% соответственно.

Результаты свидетельствуют о наличии синергетического эффекта между компонентами, при этом механизм Attention Gate оказал наиболее существенное влияние на детализацию контуров опухоли [**26**]. Проведённый парный t-тест подтвердил, что улучшение метрик Dice и IoU, обеспеченное каждым модулем, является статистически значимым ($p < 0,05$).

### 2.5. Анализ чувствительности гиперпараметров

Для определения оптимальной конфигурации предложенной модели был проведён анализ чувствительности к весовому коэффициенту $\beta$ в комбинированном функционале потерь. Как показано на рисунке 10, эксперименты проводились в диапазоне $\beta \in [0, 1]$. Результаты демонстрируют, что наивысшая производительность (Dice $\approx 96.2\%$) достигается при $\beta = 0.5$.

При малых значениях коэффициента ($\beta < 0.2$) недостаточный вклад потерь Dice затрудняет прецизионную сегментацию границ, что ведёт к снижению метрик. Напротив, при чрезмерно высоких значениях ($\beta > 0.8$) заниженный вес бинарной кросс-энтропии (BCE) также вызывает умеренное снижение точности.

Дополнительно было исследовано влияние методов пула и начальной скорости обучения на эффективность модели. Установлено, что замена

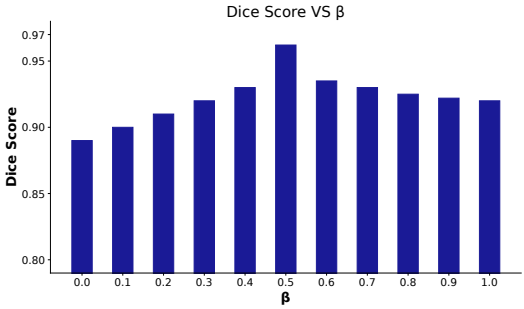



Рисунок 10. Зависимость Dice от коэффициента ($\beta$)

максимального пула (Max Pooling) на усредняющий (Average Pooling) при субдискретизации приводит к снижению коэффициента Dice примерно на 0.8%. При увеличении начальной скорости обучения до $10^{-3}$ наблюдается дестабилизация (осцилляция) процесса обучения, в результате чего значение Dice падает на 1.2%. Таким образом, в данной работе обосновано использование максимального пула с шагом 2 и скорости обучения $10^{-4}$.

В таблице 5 представлены результаты расширенного анализа гиперпараметров, включая коэффициент затухания весов (weight decay), вероятность Dropout и стратегии планирования скорости обучения. Эксперименты подтверждают, что при отклонении от выбранной конфигурации (lr=1e-4, $s = 4$, WD=1e-5, Dropout=0.5, косинусный отжиг) точность сегментации снижается в пределах 0.5%–1.5%. Это свидетельствует о допустимой чувствительности модели к настройкам гиперпараметров при сохранении стабильно высоких показателей в окрестности оптимума.

## Заключение

В данной работе для решения проблемы недостаточной точности существующих моделей при сегментации колоректальных опухолей предложена усовершенствованная архитектура MGA-UNet на базе U-Net. В разработанной модели стандартные свёрточные слои заменены комбинацией модулей Ghost и RFB, что позволяет сети извлекать более диверсифицированные признаки в процессе дискретизации. Интеграция механизма внимания CBAM между энкодером и декодером способствует концентрации модели на ключевых признаках, а внедрение блоков Attention Gate в пропускные соединения усиливает релевантные локальные характеристики.

Экспериментальные результаты на частном наборе KT-данных колоректального рака показали, что предложенный метод превосходит существующие аналоги, достигая значения коэффициента Dice 96.21%. Следует отметить, что столь высокий показатель Dice получен на выборке

Таблица 5. Расширенный анализ чувствительности к гиперпараметрам (Dice %)

| Гиперпараметр | Значение | Dice (%) | FLOPs (G) | Время (мс) |
|---|---|---|---|---|
| Скорость обучения | $10^{-5}$ | 94.82 | 15.6 | 7.2 |
| | $5 \cdot 10^{-5}$ | 95.62 | 15.6 | 7.2 |
| | $\mathbf{10^{-4}}$ | **96.21** | 15.6 | 7.2 |
| | $5 \cdot 10^{-4}$ | 95.35 | 15.6 | 7.2 |
| Коэффициент сжатия $s$ | 2 | 95.67 | 19.2 | 8.5 |
| | 3 | 96.01 | 17.1 | 7.8 |
| | **4** | **96.21** | 15.6 | 7.2 |
| | 5 | 95.98 | 14.2 | 6.7 |
| Вес распада | 0 | 95.83 | 15.6 | 7.2 |
| | $\mathbf{10^{-5}}$ | **96.21** | 15.6 | 7.2 |
| | $10^{-4}$ | 95.92 | 15.6 | 7.2 |
| Dropout | 0.0 | 96.24 | 15.6 | 7.2 |
| | 0.3 | 96.18 | 15.6 | 7.2 |
| | **0.5** | **96.21** | 15.6 | 7.2 |
| | 0.7 | 95.86 | 15.6 | 7.2 |
| Планировщик | Фиксированный | 95.44 | 15.6 | 7.2 |
| | Ступенчатый | 95.91 | 15.6 | 7.2 |
| | **Косинусное затухание** | **96.21** | 15.6 | 7.2 |

ограниченного объёма (108 пациентов) после строгой предварительной обработки (центрирование опухоли, настройка окна интенсивности), в связи с чем данные результаты могут потребовать дополнительной верификации при экстраполяции на более вариативные клинические данные.

При количестве параметров всего 10.2M и времени инференса 7.2 мс на изображение, модель перспективна для использования в сценариях клинической диагностики в реальном времени.

Направления дальнейших исследований включают:

- расширение предложенной модели до сегментации 3D КТ-объёмов [31] с разработкой 3D-версий модулей Ghost и CBAM;
- внедрение стратегий полуавтоматического обучения (semi-supervised learning) [32] для использования неразмеченных данных и повышения обобщающей способности сети;
- апробацию модели на открытых наборах данных для подтверждения её робастности.

## Благодарности

Авторы выражают благодарность сотрудникам Первой университетской больницы Хэнаньского университета науки и технологий (Первая

больница Хэнаньского университета науки и технологий, г. Лоян, Китай) за предоставленные данные.

## Список использованных источников

[1] Bray F., Laversanne M., Sung H., Ferlay J., Siegel R. L., Soerjomataram I., Jemal A. *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries* // CA: A Cancer Journal for Clinicians.– 2024.– Vol. **74**.– No. 3.– Pp. 229–263. doi ↑170

[2] Zhao S., Wang S., Pan P., Xia T., Chang X., Yang X., Guo L., Meng Q., Yang F., Qian W., Xu Z., Wang Y., Wang Z., Gu L., Wang R., Jia F., Yao J., Li Z., Bai Y. *Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis* // Gastroenterology.– 2019.– Vol. **156**.– No. 6.– Pp. 1661–1674.e11. doi ↑170

[3] Long J., Shelhamer E., Darrell T. *Fully convolutional networks for semantic segmentation* // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA, USA, June 07–12, 2015).– IEEE.– 2015.– ISBN 978-1-4673-6964-0.– Pp. 3431–3440. doi arXiv 1411.4038 ↑170

[4] Ben-Cohen A., Diamant I., Klang E., Amitai M., Greenspan H. *Fully convolutional network for liver segmentation and lesions detection* // *International Workshop on Deep Learning in Medical Image Analysis*, DLMIA 2016, LABELS 2016 (Athens, Greece, October 21, 2016), Lecture Notes in Computer Science (LNIP).– vol. **10008**, Cham: Springer.– 2016.– ISBN 978-3-319-46975-1.– Pp. 77–85. doi arXiv 1606.06650 ↑170

[5] Isensee F., Petersen J., Klein A., Zimmerer D., Jaeger P. F., Kohl S., Wasserthal J., Koehler G., Norajitra T., Wirkert S., Maier-Hein K. H. *nnU-Net: Self-adapting framework for U-Net-based medical image segmentation.*– 2018.– 11 pp. doi arXiv 1809.10486 ↑170

[6] Chen T., Son Y., Park A., Baek S.-J. *Baseline correction using a deep-learning model combining ResNet and UNet* // Analyst.– 2022.– Vol. **147**.– No. 19.– Pp. 4285–4292. doi ↑170

[7] Chlebus G., Schenk A., Moltz J. H., Ginneken B. van, Hahn H. K., Meine H. *Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing* // Scientific Reports.– 2018.– Vol. **8**.– No. 1.– id. 15497.– 7 pp. doi ↑170

[8] Iqbal A., Sharif M. *UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network* // Expert Systems with Applications.– 2023.– Vol. **221**.– No. 1.– id. 119718. doi URL ↑171

[9] Ronneberger O., Fischer P., Brox T. *U-Net: Convolutional networks for biomedical image segmentation* // *International Conference on Medical Image Computing and Computer-Assisted Intervention.*– V. III, MICCAI (Munich, Germany, October 5–9, 2015), Lecture Notes in Computer Science (LNIP).– vol. **9351**, Cham: Springer.– 2015.– ISBN 978-3-319-24573-7.– Pp. 234–241. doi arXiv 1505.04597 ↑171, 183

[10] Zhou Z., Siddiquee M. M. R., Tajbakhsh N., Liang J. *UNet++: A nested U-Net architecture for medical image segmentation* // *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, DLMIA 2018, ML-CDS 2018 (Granada, Spain, September 20, 2018), Lecture Notes in Computer

Science (LNIP).– vol. **11045**, Cham: Springer.– 2018.– ISBN 978-3-030-00888-8.– Pp. 3–11. doi arXiv 1807.10165 ↑171, 183

[11] Seo H., Huang C., Bassenne M., Xiao R., Xing L. *Modified U-Net (mU-Net) with incorporation of object-dependent high-level features for improved liver and liver-tumor segmentation in CT images* // IEEE Transactions on Medical Imaging.– 2020.– Vol. **39**.– No. 5.– Pp. 1316–1325. doi ↑171

[12] Wang Y., Lombardo E., Huang L., Belka C., Riboldi M., Kurz Ch., Landry G. *Head and neck cancer localization with Retina UNet for automated segmentation and time-to-event prognosis from PET/CT images* // *Head and Neck Tumor Segmentation and Outcome Prediction*, HECKTOR 2022 (Singapore, September 22, 2022), Lecture Notes in Computer Science.– vol. **13626**, Cham: Springer.– 2023.– ISBN 978-3-031-27419-0.– Pp. 202–211. doi ↑171

[13] Han K., Wang Y., Tian Q., Guo J., Xu Chu., Xu Cha. *GhostNet: More features from cheap operations* // *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR (Seattle, WA, USA, June 13–19, 2020).– IEEE.– 2020.– ISBN 978-1-7281-9360-1.– Pp. 1580–1589. doi arXiv 1911.11907 ↑171, 172, 174, 183

[14] Ma N., Zhang X., Zheng H.-T., Sun J. *ShuffleNet V2: Practical guidelines for efficient CNN architecture design* // *Computer Vision — ECCV 2018*, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP).– vol. **11218**, Cham: Springer.– 2018.– ISBN 978-3-030-01263-2.– Pp. 122–138. doi% doi arXiv 1807.11164 ↑171

[15] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner Th., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. *An image is worth* 16 × 16 *words: Transformers for image recognition at scale* // *9th International Conference on Learning Representations*, ICLR 2021 (Virtual Event, Austria, May 3–7, 2021).– 2021.– ISBN 979-8-3313-0008-1.– Pp. 611.– 21 pp. doi arXiv 2010.11929 URL ↑172

[16] Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A. L., Zhou Y. *TransUNet: Transformers make strong encoders for medical image segmentation.*– 2021.– 13 pp. doi arXiv 2102.04306 ↑172

[17] Hatamizadeh A., Tang Y., Nath V., Yang D., Myronenko A., Landman B., Roth H., Xu D. *UNETR: Transformers for 3D medical image segmentation* // *2022 IEEE/CVF Winter Conference on Applications of Computer Vision*, WACV (Waikoloa, HI, USA, January 03–08, 2022).– IEEE.– 2022.– ISBN 978-1-6654-0916-2.– Pp. 1748–1758. doi arXiv 2103.10504 ↑172

[18] Сухов П. А., Данилюк С. С. *Применение трансформеров для сегментации медицинских изображений* // Вестник науки.– 2024.– Т. **1**.– № 6 (75).– С. 1539–1546.– ид. 230. URL ↑172

[19] Cheng H., Zhang Y., Xu H., Li D., Zhong Z., Zhao Y., Yan Zh. *MSGU-Net: A lightweight multi-scale ghost U-Net for image segmentation* // Frontiers in Neurorobotics.– 2025.– Vol. **18**.– id. 1480055. doi ↑172, 183

[20] Pang B., Chen L., Tao Q., Wang E., Yu Y. *GA-UNet: A lightweight ghost and attention U-Net for medical image segmentation* // Journal of Imaging Informatics in Medicine.– 2024.– Vol. **37**.– No. 4.– Pp. 1874–1888. doi ↑172, 183

[21] Pan L., Wang L., Feng Zh., Xu Zh., Xu L., Peng Sh. *MGTUNet: A new UNet for colon nuclei instance segmentation and quantification.*– 2022.– 5 pp. doi arXiv 2210.10981 ↑172, 183

[22] Liu S., Huang D., Wang Y. *Receptive field block net for accurate and fast object detection* // *Computer Vision — ECCV 2018.*– V. XI, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP).– vol. **11215**.– 2018.– ISBN 978-3-030-01251-9.– Pp. 404–419. doi arXiv 1711.07767 ↑173

[23] Chen L. C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. *DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs* // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2018.– Vol. **40**.– No. 4.– Pp. 834–848. doi arXiv 1606.00915 ↑174

[24] He K., Zhang X., Ren S., Sun J. *Deep residual learning for image recognition* // *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR (Las Vegas, NV, USA, June 27–30, 2016).– IEEE.– 2016.– ISBN 978-1-4673-8851-1.– Pp. 770–778. doi% doi arXiv 1512.03385 ↑175

[25] Woo S., Park J., Lee J.-Y., Kweon I.-S. *CBAM: Convolutional block attention module* // *Computer Vision — ECCV 2018*, ECCV 2018 (Munich, Germany, September 8–14, 2018), Lecture Notes in Computer Science (LNIP).– vol. **11211**, Cham: Springer.– 2018.– ISBN 978-3-030-01233-5.– Pp. 3–19. doi% doi arXiv 1807.06521 ↑175

[26] Oktay O., Schlemper J., Folgoc L. L., Lee M., Heinrich M., Misawa K., Mori K., McDonagh S., Hammerla N. Y., Kainz B., Glocker B., Rueckert D. *Attention U-Net: Learning where to look for the pancreas.*– 2018.– 10 pp. doi arXiv 1804.03999 ↑176, 183, 184

[27] Fei G., Yan B., Chen J., Qiao K., Ning P., Shi D. *Liver tumor segmentation based on dilated convolution of stacked tree aggregation structure* // Acta Optica Sinica.– 2021.– Vol. **41**.– No. 18.– id. 1810002. doi URL ↑178

[28] Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L., Desmaison A., Köpf A., Yang E. Z., DeVito Z., Raison M., Tejani A., Chilamkurthy S., Steiner B., Fang L., Bai J., Chintala S. *PyTorch: An imperative style, high-performance deep learning library* // *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*, NeurIPS 2019 (Vancouver, BC, Canada, December 8–14, 2019).– 2019.– ISBN 9781713807933.– Pp. 8024–8035. arXiv 1912.01703 doi URL ↑179

[29] Wang Z., Zheng J. Q., Zhang Y., Cui G., Li L. *Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation.*– 2024.– 12 pp. doi arXiv 2402.05079 ↑183

[30] Zhang H., Guo W., Wan Sh., Zou B., Wang W., Qiu C., Liu K., Jin P., Yang J. *Tuning vision foundation models for rectal cancer segmentation from CT scans* // Communications Medicine.– 2025.– Vol. **5**.– No. 1.– id. 256.– 11 pp. doi ↑183

[31] Çiçek Ö., Abdulkadir A., Lienkamp S. S., Brox T., Ronneberger O. *3D U-Net: Learning dense volumetric segmentation from sparse annotation* // *International Conference on Medical Image Computing and Computer-Assisted Intervention.*– V. II, MICCAI 2016 (Athens, Greece, October 17–21, 2016), Lecture Notes

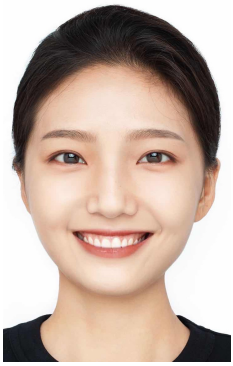
in Computer Science (LNIP).– vol. **9901**, Cham: Springer.– 2016.– ISBN 978-3-319-46722-1.– Pp. 424–432. doi % doi arXiv 1606.06650 ↑186

[32] Zhou Y., Wang Y., Tang P., Bai S., Shen W., Fishman E. K., Yuille A. L. *Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training // 2019 IEEE Winter Conference on Applications of Computer Vision*, WACV (Waikoloa, HI, USA, January 07–11, 2019).– IEEE.– 2019.– ISBN 978-1-7281-1975-5.– Pp. 121–140. doi arXiv 1804.02586 ↑186

*Поступила в редакцию* 19.04.2026;
*одобрена после рецензирования* 28.04.2026;
*принята к публикации* 29.05.2026;
*опубликована онлайн* 20.06.2026.

Рекомендовал к публикации *к.ф.-м.н. А. Н. Виноградов*

**Информация об авторах:**





Юйцянь Ван

Аспирант Томского политехнического университета (ТПУ). Область научных интересов: системный анализ, управление и обработка информации, статистика, медицинская информатика, глубокое обучение, сегментация изображений

0000-0003-3755-0586
*e-mail:* wangyuqian3333@gmail.com

Сергей Владимирович Аксёнов

Доцент отделения информационных технологий (ИШИТР) Томского политехнического университета, кандидат технических наук. Научные интересы связаны с системным анализом, управлением и обработкой информации, программными системами и медицинской информатикой

0000-0002-1251-7133
*e-mail:* axyonov@tpu.ru

*Авторы внесли равный вклад в подготовку публикации.*

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*