

Intelligent System for Analyzing MRI Images to Find the Main Elements of the Human Spine

Kanstantsin Kurachka

*Department of Information Technology
Pavel Sukhoi State Technical University of Gomel
Gomel, Republic of Belarus
kurochka@gstu.by*

Huanhai Ren

*Department of Information Technology
Pavel Sukhoi State Technical University of Gomel
Gomel, Republic of Belarus
rhh_cn@qq.com*

Abstract—This study aims to build an intelligent system based on MRI image analysis to find the main elements of the human spine, such as vertebrae and intervertebral discs. The core of this intelligent system is to design a neural network model that can accurately identify and find vertebrae and intervertebral discs. Therefore, this paper proposes an improved YOLOv8-seg image segmentation model, which introduces two lightweight convolution modules, GhostConv and C3Ghost, to reduce the computational complexity and parameter complexity of the model while maintaining high performance; at the same time, the CBAM attention mechanism is introduced to focus on specific features to improve model performance. The results show that the intelligent system built based on the improved YOLOv8-seg image segmentation model can automatically find and segment the vertebrae and intervertebral disc areas in lumbar MRI images, with high accuracy and recall, significantly improving segmentation efficiency and reducing human errors. In the future, we will further optimize the model structure and increase the diversity and scale of the data set to further improve the performance and stability of the intelligent system.

Keywords—intelligent system, YOLOv8-seg, the human lumbar spine, MRI, CBAM

I. Introduction

With the rapid development of social economy, people's pace of life is accelerating, life pressure and work intensity are significantly improved, and the incidence rate of lumbar spine disease is also rising year by year, which seriously affects the quality of life of many people. Epidemiological studies around the world show that lumbar disc herniation and other lumbar diseases occur at all ages, but the middle-aged people between 40-60 years old are the most common, and the incidence rate of men in this age group is about twice that of women [1] [2].

The human lumbar spine is typically composed of 5 vertebrae, but due to developmental abnormalities or trauma, some individuals may have 4 or 6 vertebrae. The lumbar spine plays an extremely important role in the human body, with functions mainly including regulating lower limb movement, protecting abdominal organs, and transmitting systemic loads. In the past, imaging techniques such as ultrasound and computed

tomography (CT) were widely used for the observation of lumbar spine diseases. However, these technologies have limitations in their ability to identify soft tissues, making it difficult to clearly display subtle lesions in structures such as intervertebral discs, ligaments, and nerve roots [3] [4]. In contrast, magnetic resonance imaging (MRI) technology stands out due to its significant advantages such as low radiation, high resolution, and multi planar imaging. MRI has extremely high resolution ability for soft tissue, which can clearly present details of lesions such as intervertebral disc herniation, nerve root compression, ligament damage, etc., providing important basis for medical personnel to accurately judge and evaluate the patient's condition [5]. Therefore, MRI imaging technology has become the preferred method for the examination of lumbar spine lesions, widely used in the evaluation and diagnosis of lumbar spine fractures, lumbar spondylolisthesis, lumbar disc herniation and other lesions. Medical personnel can obtain detailed soft tissue and lumbar spine bone structure images through MRI scans, accurately analyze the type, scope, and severity of lesions, and then develop personalized treatment plans. In addition, MRI can also be used to track treatment outcomes, monitor changes in the patient's condition, and provide comprehensive support for the patient's recovery.

Therefore, it is of great clinical significance to construct an intelligent system based on accurate identification and segmentation of the human lumbar vertebrae and intervertebral discs. In recent years, with the rapid development of artificial intelligence technology [6] [7], deep learning has been widely applied in computer vision fields such as image classification [8], image segmentation [9] [10], and object detection [11]. Deep learning methods, with their powerful feature extraction capabilities, are able to learn more complex and refined feature representations, demonstrating significant advantages in accuracy, robustness, and generality. This has led to its rapid rise in the field of medical image analysis and diagnosis, becoming the mainstream technology for current medical image segmentation and providing strong support for accurate diagnosis of lumbar

spine and intervertebral disc related lesions. In medical image analysis, deep learning algorithms can recognize complex patterns and subtle features that are difficult for the human eye to detect through large-scale data training. For example, models based on Convolutional Neural Networks (CNN) can efficiently segment medical images, accurately outline the contours of the lumbar spine and intervertebral discs, and provide clear anatomical information for clinical doctors. This high-precision image segmentation capability not only helps to detect lesions early, but also provides important basis for the development of personalized treatment plans.

This intelligent system is based on the PyTorch deep learning framework and uses an improved YOLOv8-seg lumbar image segmentation model. [12] [13]. It significantly improves the segmentation efficiency and accuracy by introducing two lightweight convolution modules, GhostConv and C3Ghost, as well as the CBAM attention module. These improved modules not only optimize the computational complexity of the model, but also enhance its ability to extract features of the lumbar spine and intervertebral discs, enabling it to quickly and accurately identify the contours of each segment of the lumbar spine and intervertebral disc. Through this efficient image segmentation technology, medical personnel can clearly understand the scope and degree of lesions, providing important imaging evidence for clinical diagnosis and treatment plan formulation.

II. Introduction to YOLOv8-seg

The YOLOv8-seg network architecture mainly consists of three core components: Backbone, Neck, and Head. The overall architecture is shown in Figure 6. Compared to YOLOv5, YOLOv8 introduces an innovative C2f module in the backbone network to replace the original C3 module. The C2f module significantly enhances the feature extraction capability by optimizing the gradient flow path, thereby improving the overall performance of the model.

In the neck network section, YOLOv8-seg adopts an advanced multi-scale feature fusion architecture combining FPN (Feature Pyramid Network) and PAN (Path Aggregation Network). This design effectively enhances the model's perception of target detail features and contextual information through multi-level feature aggregation and information transmission mechanisms, thereby improving the accuracy of object detection and instance segmentation.

The head network design embodies the architectural innovation of YOLOv8-seg, mainly including two key improvements: firstly, adopting a decoupled head structure to separate classification tasks from detection tasks. This design not only improves the professionalism of tasks, but also achieves a paradigm shift from anchor based to anchor free, significantly reducing the number of bounding box predictions and optimizing the

computational efficiency of non maximum suppression (NMS); Secondly, in terms of loss function design, a task aligned allocator was introduced for positive sample allocation, and a distribution focal loss was adopted. These improvements effectively improved the training stability and detection accuracy of the model.

In terms of implementing instance segmentation functionality, YOLOv8-seg has added a segment segmentation branch in the head network, and its mask generation mechanism includes two key steps.

A. Prototype Mask Generation

Generate a fixed number of prototype masks (Prototype Masks) through the Protonet network in the P3 feature layer. These prototype masks are universal mask templates obtained through network learning, with global sharing characteristics, that is, all instance segmentation tasks are based on the same set of prototypes for mask reconstruction.

B. Mask coefficient generation

For each detected target instance, the model outputs a corresponding set of weight coefficients through the segment mask terminals of the P3, P4, and P5 feature layers. These coefficients determine the weighted combination of prototype masks, and ultimately generate the predicted masks (Masks) of the target instance through linear combination, as expressed mathematically in equation (1).

$$Masks = M_c P \quad (1)$$

In equation (1), P is the prototype mask with dimensions $B \times K \times H \times W$, and B is the batch size; K is the number of prototype masks generated, default is 32; $H \times W$ is the size of the prototype mask; M_c is the mask coefficient. By using mask prototypes and mask coefficient weighting, high-quality masks can be generated using simple linear calculations, which is more efficient than traditional two-stage instance segmentation methods and can achieve a better balance between speed and accuracy.

III. YOLOv8-seg Improvement

A. Lightweight Convolution Module

GhostConv and C3Ghost are two lightweight convolution modules introduced in YOLOv8, aimed at reducing the computational and parameter complexity of the model while maintaining high performance.

a) *GhostConv*: It is a lightweight convolution operation based on the Ghost Module concept. It reduces computational complexity by generating "phantom feature maps". The Ghost Module first uses a small number of convolution kernels to generate a portion of feature maps, and then performs simple linear transformations on these feature maps to generate more "phantom feature maps". Finally, concatenate the intrinsic feature map and the phantom feature map to obtain the final output.

b) *C3Ghost*: It is a lightweight module based on the GhostConv concept, which improves the efficiency of the model by reducing computation and parameter count while maintaining high performance.

B. CBAM attention mechanism

CBAM (Convolutional Block Attention Module) is an innovative attention mechanism designed specifically for convolutional neural networks, aimed at adaptively enhancing key information in feature maps. CBAM combines Channel Attention and Spatial Attention mechanisms to weight feature maps from both channel and spatial dimensions, significantly improving the model's representation ability. Its unique dual attention mechanism not only captures the dependencies between channels, but also recognizes the importance of spatial positions, making the model more robust in complex scenes. In addition, CBAM has extremely low computational overhead and hardly increases the number of parameters and computational complexity of the model, making it seamlessly integrated into existing CNN architectures and significantly improving model performance.

Through this dual attention mechanism, CBAM not only enhances the feature extraction capability of the model, but also achieves an excellent balance between computational efficiency and model performance, becoming an indispensable component in modern convolutional neural networks.

C. Improved YOLOv8-seg model

In the YOLOv8 seg model, GhostConv and C3Ghost lightweight modules are introduced to optimize the computational efficiency and parameter count of the model. Specific improvements include replacing most of the traditional Conv modules with GhostConv modules, and replacing the C2f module in the YOLOv8 seg model with the C3Ghost module. The GhostConv module reduces computational complexity by generating "phantom feature maps", while the C3Ghost module further reduces model complexity while maintaining feature expression ability. These improvements significantly enhance the inference speed of the model while maintaining high detection and segmentation accuracy. The improved YOLOv8-seg model structure is shown in "Figure 1", which illustrates the overall architecture after module replacement.

In the YOLOv8 framework, targeted architecture optimization was performed on the yolov8-seg.yaml configuration file. Specifically, two important structural improvements were made to the backbone module: firstly, all standard convolutional layers (Conv) except for the initial layer (Layer 0) were replaced with lightweight Ghost Conv modules (GhostConv), which significantly reduced the computational complexity and number of parameters of the model while maintaining feature extraction capabilities; Secondly, the Convolutional Attention Mechanism module (CBAM) is integrated in level

10, which enhances the model's ability to extract key features through a dual mechanism of channel attention and spatial attention.

In the process of optimizing the model architecture, the configuration file of YOLOv8 (yolov8-seg.yaml) was adjusted. Specifically, the C2f structure in the original head module was replaced with a more efficient lightweight convolution module C3Ghost. This improvement aims to maintain the feature extraction capability while reducing the computational complexity of the model.

IV. Experiment and result analysis

A. Dataset preparation

The experimental dataset used in this study contains a variety of lumbar spine MRI images, covering different morphological features (such as disc herniation, spinal stenosis, etc.) and various sizes. All images are stored in JPG format. To verify the effectiveness of the model, we selected 512 representative sagittal MRI images of the lumbar spine from the public dataset. The dataset is divided into training set, test set and validation set in a ratio of 7:2:1. These sagittal images clearly show the anatomical structure of the lumbar spine, including key tissues such as vertebral bodies, and intervertebral discs, providing rich morphological information for model training. "Figure 2" shows samples of lumbar spine MRI images with typical pathological features in the dataset, which fully demonstrates the diversity and representativeness of the dataset.

In this study, the labeling of human lumbar sagittal MRI images was completed using the advanced labelme labeling tool to ensure the accuracy and consistency of the labeling. During the labeling process, we subdivided the lumbar structure into 10 categories, including 5 lumbar vertebral categories (L1, L2, L3, L4, L5) and 5 intervertebral disc categories (D1, D2, D3, D4, D5).

Since the number of original images is not too abundant, the original images are expanded by means of data augmentation. However, if the validation set is data augmented, the samples in the validation set will no longer represent the data distribution in the real scene, which may make the model too optimistic on the validation set and fail to accurately evaluate the generalization ability of the model. Therefore, in order to maintain the authenticity and representativeness of the validation set, data augmentation is only performed on the training set and the test set. Based on the Albumentations image enhancement library, the image is enhanced by vertical flipping, horizontal flipping, rotation, random scaling, random adjustment of image brightness and contrast, and addition of Gaussian noise to form an enhancement pipeline. Each enhancement operation has a certain probability of being applied, which ensures that the input image will be different each time training, enhancing

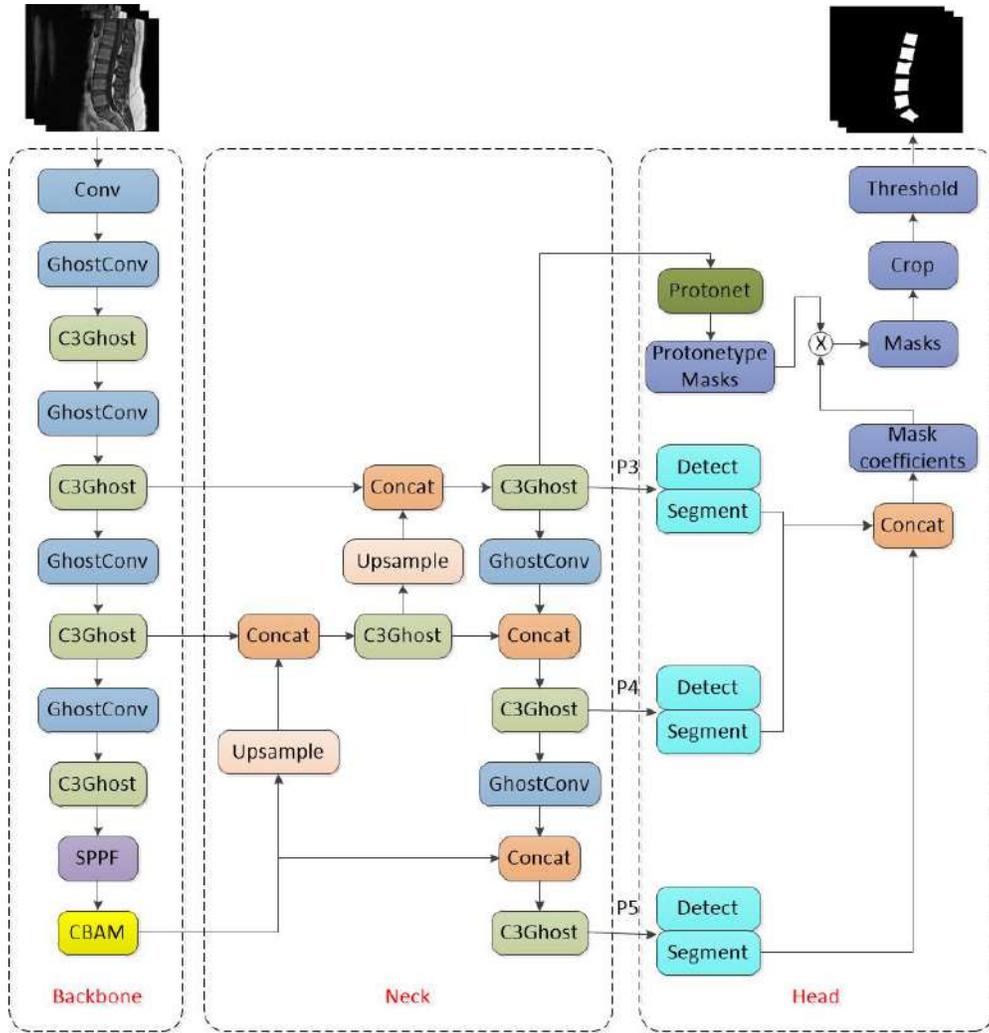


Figure 1. Improved YOLOv8-seg model.

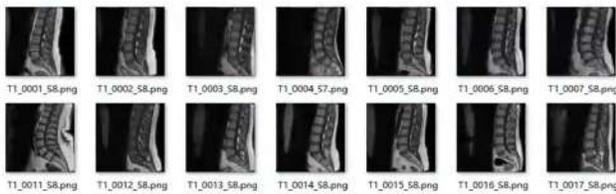


Figure 2. The partial images in the dataset.

the robustness and generalization of the model while reducing overfitting.

B. Experimental environment

The experimental environment for segmenting MRI human lumbar vertebrae and intervertebral discs using the improved YOLOv8-seg segmentation model is shown in “Table I”.

Table I
Experimental environment

Software and hardware names	Specific parameters
CPU	Intel(R) Xeon(R) Gold 6226R
Memory	64GB
GPU	Quadro RTX 5000
Operating system	Window 10 Professional Edition
Software environment	Python3.12.7+torch2.6.0+cu118

C. Experimental process

In this experiment, the model training was performed for 200 rounds, and the batch size of each round of training was set to 16, with an initial learning rate of 0.01. The resolution of the input image was uniformly adjusted to 640×640 pixels to ensure that the model can fully capture the detailed features of the lumbar spine and intervertebral disc.

To comprehensively evaluate the segmentation perfor-

mance of the model, the mean average precision (mAP) was used as the core evaluation indicator, including mAP@0.5 and mAP@0.5:0.95. Among them, mAP@0.5 measures the segmentation accuracy of the model when the intersection over union (IoU) threshold is 0.5, while mAP@0.5:0.95 is the average value of multiple IoU thresholds (from 0.5 to 0.95, with a step size of 0.05), which comprehensively reflects the performance of the model under different accuracy requirements. The mAP indicator not only comprehensively considers the accuracy and recall of the model, but also can effectively evaluate the generalization ability and robustness of the model. Through the above multi-dimensional performance evaluation, the performance of the model in the lumbar spine and intervertebral disc segmentation tasks can be comprehensively measured.

For the model training data, the bounding box detection performance and mask segmentation performance are analyzed.

a) *bounding box detection performance*: The training data of bounding box detection performance is shown in “Table II”. The accuracy of all categories is above 0.938, indicating that the model has high accuracy in detecting targets; the recall rate is above 0.965, indicating that the model can effectively detect most targets; the mAP50 value is above 0.983, indicating that the average accuracy of the model is very high when the IoU threshold is 0.5; the mAP50-95 value is between 0.684 and 0.847, indicating that the average accuracy of the model at different IoU thresholds has decreased.

Table II
The bounding box detection performance

Class	Box(P)	Box(R)	Box(mAP50)	Box(mAP50-95)
a11	0.96	0.984	0.989	0.758
D1	0.938	0.985	0.986	0.697
D2	0.956	0.984	0.983	0.703
D3	0.975	0.985	0.992	0.694
D4	0.96	0.985	0.983	0.684
D5	0.962	0.985	0.994	0.768
L1	0.942	0.965	0.985	0.792
L2	0.944	0.985	0.994	0.78
L3	0.971	0.995	0.994	0.847
L4	0.974	0.985	0.994	0.823
L5	0.974	0.985	0.985	0.794

b) *mask segmentation performance*: The training data of mask segmentation performance are shown in “Table III”. The accuracy of mask segmentation is above 0.938, indicating that the model also has high accuracy in the segmentation task; the recall rate is above 0.965, indicating that the model can effectively detect most targets in the segmentation task; the mAP50 value is above 0.983, indicating that the model has a high average precision in the segmentation task; the mAP50-95 value is between 0.663 and 0.839, indicating that the average precision

of the model in the segmentation task at different IoU thresholds has decreased.

Table III
The training data of mask segmentation

Class	Mask(P)	Mask(R)	Mask(mAP50)	Mask(mAP50-95)
a11	0.96	0.984	0.989	0.752
D1	0.938	0.985	0.986	0.687
D2	0.956	0.984	0.983	0.692
D3	0.975	0.985	0.992	0.677
D4	0.96	0.985	0.983	0.663
D5	0.962	0.985	0.994	0.706
L1	0.942	0.965	0.985	0.806
L2	0.944	0.985	0.994	0.838
L3	0.971	0.995	0.994	0.839
L4	0.974	0.985	0.994	0.814
L5	0.974	0.985	0.985	0.802

D. Image prediction

Use a trained human lumbar vertebrae and intervertebral disc segmentation model to perform object detection and segmentation on multiple images, and output detection results, including category labels and accurate bounding boxes of human lumbar vertebrae and intervertebral discs.

The image prediction in this study selected two types of images for prediction. The first type is the test image in the public dataset, and the second type is the real unprocessed MRI human lumbar spine image collected from the hospital imaging center. Firstly, multiple test images in the public dataset were predicted, and the prediction results are shown in “Figure 3”.

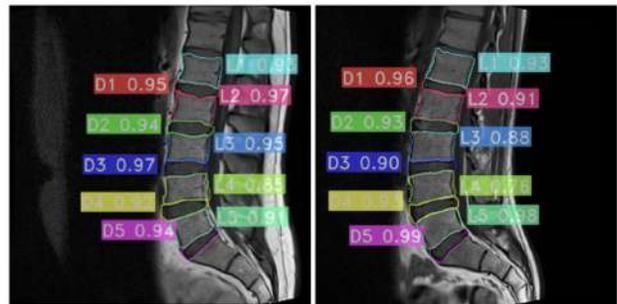


Figure 3. Public dataset image prediction.

Secondly, multiple real unprocessed MRI images of the human lumbar spine were predicted, and the prediction results are shown in “Figure 4”.

V. Conclusion

In order to enable the intelligent system to more accurately and quickly find and segment the main elements of the human spine, such as the lumbar vertebrae and intervertebral discs, this paper proposes an improved instance segmentation model based on YOLOv8-seg. By introducing the CBAM attention mechanism



Figure 4. Real unprocessed image Prediction.

and two lightweight convolution modules, GhostConv and C3Ghost, the segmentation accuracy and speed are improved. Experiments were conducted on the public human lumbar sagittal dataset. The overall accuracy, recall rate, and mAP50 of the model training were 0.96, 0.984, and 0.989, respectively. Good prediction results were obtained on a public human lumbar sagittal image dataset, proving the effectiveness and feasibility of the intelligent system.

Although the intelligent system based on this model performs well in the search and segmentation of the human lumbar spine and intervertebral disc, there are still some accuracy issues in the segmentation of the curved boundaries on both sides of the image. Therefore, in future research, it is possible to consider combining the deep learning model with the traditional edge extraction algorithm to further optimize the segmentation model and algorithm and enhance the practicality and accuracy of the intelligent system.

References

- [1] Andersson G B J. Epidemiological features of chronic low-back pain[J]. The lancet, 1999,354(9178): 581-585
- [2] RAVINDRA V M,SENGLAUB S S,RATTANI A,et al.Degenerative lumbar spine disease:estimating global incidence and worldwide volume[J].Global Spine Journal,2018,8(8):784-794
- [3] Hides J A, Richardson C A, Jull G A. Magnetic resonance imaging and ultrasonography of the lumbar multifidus muscle: comparison of two different modalities[J]. Spine, 1995, 20(1): 54-58.
- [4] Danneels L A, Vanderstraeten G G, Cambier D C, et al. CT imaging of trunk muscles in chronic low back pain patients and healthy control subjects[J]. European spine journal, 2000, 9: 266-272.
- [5] Holmes J, Sacchi L, Bellazzi R. Artificial intelligence in medicine[J]. Ann R Coll Surg Engl, 2004, 86: 334-8.
- [6] Surden H. Artificial intelligence and law: An overview[J]. Georgia State University Law Review,2019, 35: 19-22.
- [7] Kurachka K., Kamrakou U., Masalitina N. The automated classification system for lumbar spine anatomic elements //Nonlinear Dynamics and Applications, 2017, Vol. 23, P. 127-134.
- [8] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. San Francisco: AAAI Press,2017: 4278-4284.
- [9] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv,2014:1412.7062.

- [10] ZHAO H.S., SHI J.P, QI X.J., et al. Pyramid scene parsing network [C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition.New York: IEEE Press, 2017: 2881-2890.
- [11] Zhao S., Wu X., Chen B., et al. Automatic vertebrae recognition from arbitrary spine MRI images by a category-Consistent self-calibration detection framework[J]. Medical Image Analysis, 2021,67: 101826-101826.
- [12] Hurtado-Avilés J. et al. Validity and reliability of a computer-assisted system method to measure axial vertebral rotation //Quantitative imaging in medicine and surgery, 2022, Vol. 12, № 3, P. 1706.
- [13] Yang Xiaotian, TanJinlin, Yu Xin, etal. Ship Target Tracking Basedon GAM YOLOv8 RemoteSensingImages. Journal of Jilin University (Earth Science Edition), 2025, 55(1) : 328 339.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИЗА МРТ ИЗОБРАЖЕНИЙ ДЛЯ ПОИСКА ОСНОВНЫХ ЭЛЕМЕНТОВ ПОЗВОНОЧНИКА ЧЕЛОВЕКА

Курочка К. С., Рен Хуанхай

Целью данного исследования является создание интеллектуальной системы для анализа МРТ изображений с целью локализации и определения геометрических характеристик основных элементов позвоночника человека, таких как позвонки и межпозвоночные диски. Основой данной интеллектуальной системы является нейросетевая модель, которая может с достаточной для практического применения точностью идентифицировать и находить позвонки и межпозвоночные диски. В данной статье предлагается усовершенствованная модель сегментации изображений YOLOv8-seg, которая вводит два облегченных модуля свертки, GhostConv и C3Ghost, для снижения вычислительной сложности и уменьшения параметров модели при сохранении высокой производительности; в то же время вводится механизм внимания СВМ для фокусировки на определенных функциях для повышения производительности модели. Результаты показывают, что интеллектуальная система, созданная на основе усовершенствованной модели сегментации изображений YOLOv8-seg, может автоматически находить и сегментировать области позвонков и межпозвоночных дисков на изображениях МРТ поясничного отдела позвоночника с высокой точностью и полнотой, что значительно повышает эффективность сегментации и снижает количество человеческих ошибок. В будущем мы продолжим оптимизировать структуру модели, а также увеличим разнообразие и масштаб набора данных для дальнейшего повышения производительности и стабильности интеллектуальной системы.

Received 21.03.2025