中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2025)04-0977-12

论文引用格式:Wan A, Gao H L, Zhou X, Xue Z and Mou X G. 2025. YOLO-SF-TV: transcranial ultrasound images of the third ventricle as adetection model. Journal of Image and Graphics, 30(4):0977-0988(万奥, 高红铃, 周晓, 薛峥, 牟新刚. 2025. YOLO-SF-TV: 经颅超声图像三脑室检测模型. 中国图象图形学报, 30(4):0977-0988)[DOI:10.11834/jig.240293]

YOLO-SF-TV: 经颅超声图像三脑室检测模型

万奥[1],高红铃[2],周晓[1*],薛峥[3],牟新刚[1]

1. 武汉理工大学机电工程学院,武汉　433070; 2. 南昌大学第一附属医院神经内科,南昌　330006;
3. 华中科技大学同济医学院附属同济医院神经内科,武汉　433074

摘　要:**目的**　经颅超声成像技术作为高效率、低成本且无创的诊断手段,已逐步应用于帕金森病患者认知功能障碍诊断。由于经颅超声图像信噪比低、成像质量差、目标组织复杂且相似度高,需要依赖专业医生手动检测。但是人工检测不仅费时费力,还可能因为操作者的主观因素影响,造成检测结果出现差异性。针对这一问题,提出了一种基于Swin Transformer和多尺度深度特征融合的YOLO-SF-TV(YOLO network based on Swin Transformer and multi-scale deep feature fusion for third ventricle)模型用于经颅超声图像三脑室检测,以提高临床检测准确率,辅助医生进行早期诊断。**方法**　YOLO-SF-TV模型在YOLOv8的基础上使用基于窗口注意力的Swin Transformer作为模型特征提取网络,并引入空间金字塔池化合模块SPP-FCM(spatial pyramid pooling fast incorporating CSPNet and multiple attention mechanisms)扩大网络感受野,并增强多尺度特征融合能力。在网络的多尺度特征融合部分结合深度可分离卷积和多头注意力机制,提出了PAFPN-DM(path aggregation and feature pyramid network with depthwise separable convolution)模块,并对主干特征输出层增加多头注意力机制,以提高网络对不同尺度特征图中全局和局部重要信息的理解能力。同时,将传统卷积替换为深度可分离卷积模块,通过对每个通道单独卷积提高网络对不同通道的敏感性,以保证模型准确度的同时降低训练参数和难度,增强模型的泛化能力。**结果**　在本文收集的经颅超声三脑室图像数据及对应标签的数据集上进行实验,并与典型的目标检测模型对比。实验结果表明,本文提出的YOLO-SF-TV在经颅超声三脑室目标上的平均精确度均值(mean average precision, mAP)达到98.69%,相比于YOLOv8提升了2.12%,与其他典型模型相比检测精度达到最优。**结论**　本文提出的YOLO-SF-TV模型在经颅超声图像三脑室检测问题上表现优秀,SPP-FCM模块和PAFPN-DM模块可以增强模型检测能力,提高模型泛化性和鲁棒性。同时,本文制作的数据集将有助于推动经颅超声三脑室图像检测问题的研究。

关键词:经颅超声成像;计算机辅助诊断(CAD);三脑室;深度学习;YOLOv8;Swin Transformer

YOLO-SF-TV: transcranial ultrasound images of the third ventricle as adetection model

Wan Ao[1], Gao Hongling[2], Zhou Xiao[1*], Xue Zheng[3], Mou Xingang[1]

1. School of Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan　433070, China;
2. Department of Neurology, First Affiliated Hospital of Nanchang University, Nanchang　330006, China; 3. Department of Neurology, Tongji Hospital Affiliated to Tongji Medical College, Huazhong University of Science and Technology, Wuhan　433074, China

Abstract: **Objective** Cognitive impairment is the most dangerous nonmotor symptom of Parkinson's disease (PD) and affects approximately 25%−30% of patients every year. This condition seriously affects their quality of life and increases

收稿日期:2024-06-04;**修回日期**:2024-10-09;**预印本日期**:2024-10-16

*通信作者:周晓　zhouxiao@whut.edu.cn

the risk of death. However, the accuracy of clinical diagnosis of PD cognitive impairment is still limited. The proportion of patients with PD diagnosed before the age of 50 years is less than 4%. Some scholars have proposed that transcranial ultrasound imaging of the third ventricle can assist doctors in the diagnosis of PD cognitive impairment. As a rapid, noninvasive, and low-cost detection method, transcranial ultrasound imaging has been gradually applied to the diagnosis of cognitive dysfunction in patients with PD, helping doctors find the disease in time and treat it as soon as possible. Owing to the low signal-to-noise ratio of transcranial ultrasound images and the poor imaging quality, complexity, and similarity of target tissues, specialized physicians rely on manual detection. However, this process is time consuming, labor intensive, and may result in variability among detection results due to the influence of subjective factors related to the operator. Deep learning has been increasingly integrated with the medical field, especially the computer diagnosis(CAD) system based on deep learning used to diagnose PD with good results. In this work, a YOLO-SF-TV network based on Swin Transformer and multiscale feature fusion is proposed for transcranial ultrasound image third ventricle detection to assist physicians in the early diagnosis. **Method** A total of 2 400 transcranial ultrasound images of the third ventricle and the corresponding labels are acquired to form a dataset, and the third ventricle region in each image was manually labeled by a professional. The YOLO-SF-TV network is designed to consist of backbone, neck, and head components, whose roles are used to extract image features, fuse image features, and detect and classify targets, respectively. This work uses an algorithm based on YOLOv8 and the window-attention based Swin Transformer to improve the model backbone network and strengthen its ability to model global information. SPP-FCM, a spatial pyramid pooling module, is connected to the Swin Transformer network to enhance the network sensibility and integrate multiscale information. The SPP-FCM structure combines the characteristics of the CSPC structure in YOLOv7 while targeting the introduction of a multihead attention mechanism (MHAM) in the multilevel pooling part, which reduces the sensitivity of the model to noise and outliers during the extraction of multidimensional features. In the multiscale feature fusion PAFPN part of the network, the PAFPN-DM module is proposed by combining depthwise separable convolution (DCOW) with the multihead attention mechanism added to the backbone feature output layer to improve the network's ability to understand the important global and local information in different scale feature maps. At the same time, traditional convolution is replaced with a depth-separable convolution module, which improves the sensitivity of the network to different channels by convolving each channel individually, to ensure the accuracy of the model while reducing the training parameters and difficulty and enhancing the generalization ability of the model. **Result** Fivefold cross-validation evaluation was performed on the dataset to validate the performance of the different networks. The dataset was randomly divided into equal quintuples, of which four at a time were used as the training set and the remaining one as the test set. The training input image was resized to 640 × 640 pixels, and the training dataset was expanded using data enhancement methods such as random flip, random angle rotation, and Mosaic. The initial learning rate for model training was set to 0.001, and the learning rate decayed to 0.1 times of the original every 50 epochs, with a momentum of 0.9, a decay coefficient of 0.000 5, and a batch size of 8. GeForce RTX 3090 was used as the GPU in the experiments, and the mean average precision (mAP) metrics were applied as a measure for detecting network performance under the Ubuntu 20.04 operating system and PyTorch framework. Experimental results show that the YOLO-SF-TV algorithm achieves 98.69% detection accuracy on transcranial ultrasound third ventricle targets, an improvement of 2.12% relative to that of the YOLOv8 model. Therefore, the detection accuracy is optimized compared with that of typical models. **Conclusion** The proposed YOLO-SF-TV model performs excellently in the detection of the third ventricle in transcranial ultrasound images. The SPP-FCM and PAFPN-DM modules enhance the model's detection capability, generalization, and robustness. The produced dataset helps promote the research on third ventricle detection in transcranial ultrasound images.

Key words: transcranial ultrasound imaging; computer aided diagnosis(CAD); third ventricle; deep learning; YOLOv8; Swin Transformer

0 引 言

经颅超声(transcranial ultrasound，TCS)是一种能够通过完整骨窗显示大脑深部结构的神经成像技术，它不仅分辨率与磁共振成像(magnetic resonance imaging，MRI)相当，还能直观显示神经退行性疾病的特征变化(Gao等，2024)。目前帕金森病(Parkinson's disease，PD)认知功能障碍的临床诊断的准确性仍然有限，患者在50岁之前确诊的比例不到4%(Heravi等，2023)。根据柳叶刀委员会在报告中提出的控制危险因素手段，及时发现并合理治疗PD患者可以预防或延长近50%的痴呆症状者发病(Livingston等，2020)，因此，亟需寻找PD患者认知障碍的影响标志物。

在这种情况下，TCS作为一种快速、无创、低成本、可复用性高、结果内部一致性高且患者接受度高的检测手段(Monaco等，2018)，可以在基层医疗机构广泛应用，促进医疗资源的公平分配，提高广大患者的医疗保障水平。经颅超声检测不仅在评估脑萎缩方面的准确性高，而且临床禁忌症状较少，能够为PD患者认知功能评估提供客观的影响学体征和脑结构图的改变(Fu等，2021)。临床上使用TCS图像进行三脑室目标检测可以减轻医生的工作负担，减少人为误差，提高诊断的准确性和一致性。通过使用深度学习模型自动化检测能够提升神经系统疾病的早期诊断和治疗效果，改善患者的生活质量。因此，研究三脑室TCS图像自动检测方法具有很高的临床以及社会价值。如图1所示，TCS图像中三脑室呈现两条近似线状结构。在经颅超声临床检测上会根据其特有标志物松果体(虚线框内)来辅助判断。但TCS检查对超声医生的临床经验有很大依赖性，在一定程度上限制了其广泛应用(Vlaar等，2011)。

近年来，深度学习与医学图像检测结合已经产生了许多成功的应用。例如，乳腺癌检测分类(Lotter等，2021)、脑肿瘤检测(Özbay和Özbay，2023)、甲状腺结节检测(于典 等，2023)、帕金森病检测(Yang等，2022)和肺肿瘤检测(周涛 等，2022)等。在众多深度学习网络结构中，以YOLO(you only look once)(Redmon和Farhadi，2018；Bochkovskiy等，2020；Ge等，2021；Wang等，2023)为代表的单目标检测模型

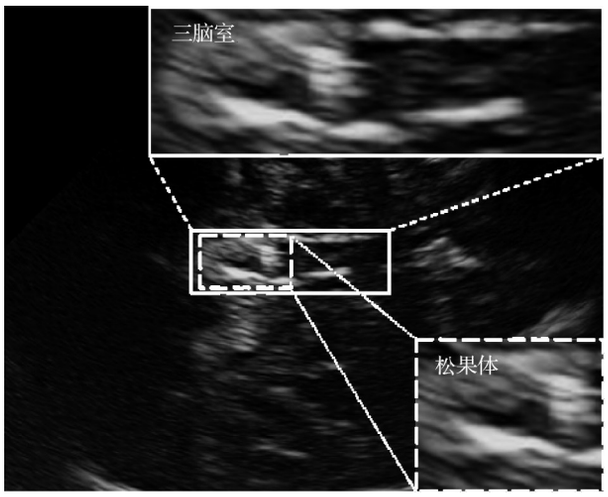


图1 三脑室TCS图像

Fig. 1 TCS images of the three ventricles

系列广泛应用于计算机辅助诊断(computer aided diagnosis，CAD)。Baccouche等人(2022)在早期乳腺癌诊断中将YOLO和生成对抗网络(generative adversarial networks，GAN)融合进行检测，解决检测中目标由于时间和纹理变化导致检测失衡的问题。Meng等人(2022)提出一种基于YOLOv5的CAD系统，实现对上消化道内窥镜中食管鳞状细胞癌(esophageal squamous cell carcinoma，ESCC)的检测和诊断，系统检测准确性与专家型类内窥镜医生检测准确率相当。于典等人(2023)将注意力机制融合到甲状腺超声图像中，实现了端到端的自动检测任务。

与此同时，基于自然语言处理(natural language processing，NLP)的Transofrmer(Ghazouani等，2023；Tarimo等，2024)架构和基于生成任务的Diffusion(Croitoru等，2023)架构都在目标检测任务中展现出不错的效果。Swin Transformer(Liu等，2021)使用类卷积神经网络中的层次化结构网络，通过窗口局部注意力机制和网络分层设计，克服了传统卷积操作的固有局限性，增强网络全局建模能力。Vit Transformer(Yin等，2022)基于图像分块思想，通过对图像进行分块处理实现位置编码和序列特征输出。基于特征金字塔结构的Transofrmer (pyramid vision Transformer，PVT)(Wang等，2022)能处理不同尺度图像信息，并且通过金字塔分解和跨尺度融合有效地捕获了图像的全局特征。具体地，Swin Unet(Cao等，2022)将Swin Transformer中的滑动窗口思想引入U-Net，在多器官和心脏分割任务中取得了良好的

效果。郝文月等人(2024)通过结合卷积神经网络(convolutional neural network, CNN)和Transformer,提出一种双分支算法,通过构建局部—全局信息实现血管超声图像分割任务。周雪等人(2024)针对结肠息肉检测困难的问题,提出了PVTA(pyramid vision Transformer and axial attention network)网络。该网络首先通过PVT实现网络特征提取,然后基于特征金字塔和空间金字塔进行特征融合,实现目标分割任务。Chen等人(2023)将基于Diffusion的架构首次引入目标检测任务中,提出了DiffusionDet的模型。实现了将目标检测建模任务变成从噪声框到目标框的去噪扩散过程。

目前对于TCS图像中三脑室检测问题的研究还较少,特别是尚未发现深度学习运用于三脑室检测的相关工作。在此背景下,本文针对TCS图像复杂背景下展开的基于深度学习的三脑室目标检测算法研究,为TCS图像的三脑室检测提供了有效的解决方案,能够显著提升检测的精度和效率,减轻医务人员的工作负担。本文主要贡献包括:1)制作了用于TCS图像三脑室检测问题研究的数据集,包括2 400幅TCS图像及其相应的三脑室图像标注标签;2)针对TCS图像背景复杂识别定位困难的问题,使用Swin Transformer作为特征提取网络,并在SPPF(spatial pyramid pooling fast)基础上融合CSPNet(cross stage partial network)和多头注意力机制(multi-head attention mechanism, MHAM),提出全新的SPP-FCM(spatial pyramid pooling fast incorporating CSPCNet and multiple attention mechanisms)模块,扩大网络感受野的同时提高网络对多维度尺寸特征的处理能力;3)针对网络在加强特征提取模块部分的训练过程中计算复杂度高、信息冗余等问题,结合深度可分离卷积和多头注意力模块设计PAFPN-DM(path aggregation and feature pyramid network with depth-wise separable convolution)模块,在减少网络特征融合部分冗余度的同时保证检测精度。

1 数据集

1.1 数据集介绍

TCS使用超声成像技术,在检测过程中会受到包括检测角度、仪器固有噪声等多种因素干扰,导致图像成像质量低、散斑噪声明显和边缘信息弱等问题。同时,图像也会受到病人年龄、性别和病情程度等影响,这都为后续的目标检测产生影响。如图2所示,根据三脑室的结构特点将其分为4个类别:图2(a)中三脑室两侧非常接近,边界模糊;图2(b)中三脑室与松果体都呈离散状态;图2(c)中三脑室一侧与松果体接近,甚至与其融为一体;图2(d)中三脑室与松果体接近或者融为一体。

1.2 数据预处理

本文在经颅超声图像上使用LabelImg进行三脑室目标位置信息标注,标注后会生成相对应目标信息的xml文件,在标注中尽量使标注框与三脑室边缘贴合。

实验使用5倍交叉验证评估的方法进行模型性能测试。具体地,将实验数据集分为5份,每份240幅图像,选择其中的1份作为测试集合,另外4份为训练集。训练时统一将图像调整为640×640像素,并对输入图像采用随机旋转、缩放和Mosaic等数据增强手段丰富模型训练样本。

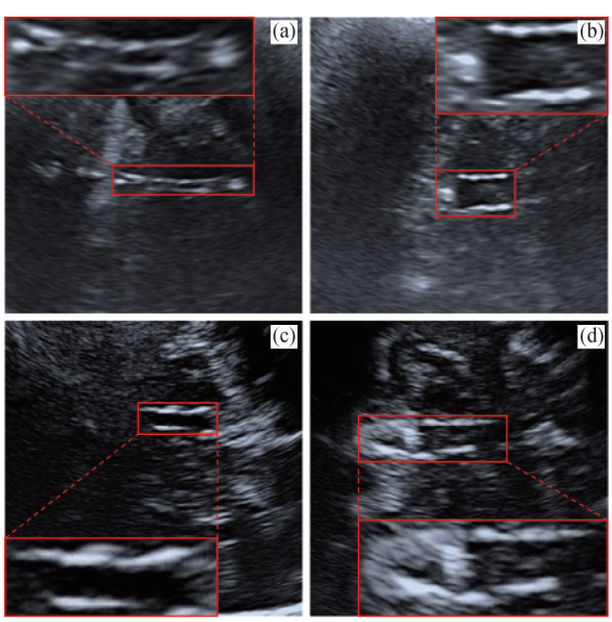


图2 不同类型的三脑室

Fig. 2 Different types of third ventricles

2 本文方法

YOLO-SF-TV (YOLO network base on Swin Transformer and multiscale deep feature fusion for third ventricle)模型整体结构如图3所示。模型包括主干特征提取网络(backbone)、特征融合网络

(neck)以及分类和回归操作的检测头(YOLO head)3个部分,其作用分别是特征提取、特征融合和目标检测与分类。YOLO-SF-TV模型在backbone中使用Swin Transformer替换YOLOv8中的CSPDarknet作为TCS图像特征提取器。为了充分利用模型网络上的多尺度信息,将Swin Transformer的3个分块合并模块输出特征作为主干网络的多级特征输出层,并对最高的特征输出层后面增加SPP-FCM模块,加强网络对多维通道特征的提取能力。最后,将深度可分离卷积和多头注意力机制引入neck中,用于加强特征融合的PAFPN模块,以此设计出全新的PAFPN-DM模块。PAFPN-DM模块中首先对主干网络的3个特征输出层结合multi-head attention模块,并在原始CSP基础上融合深度可分离卷积得到全新的DWCSP(depthwise separable convolution in CSP)结构,实现对每个通道单独卷积,提高网络的通道敏感性。

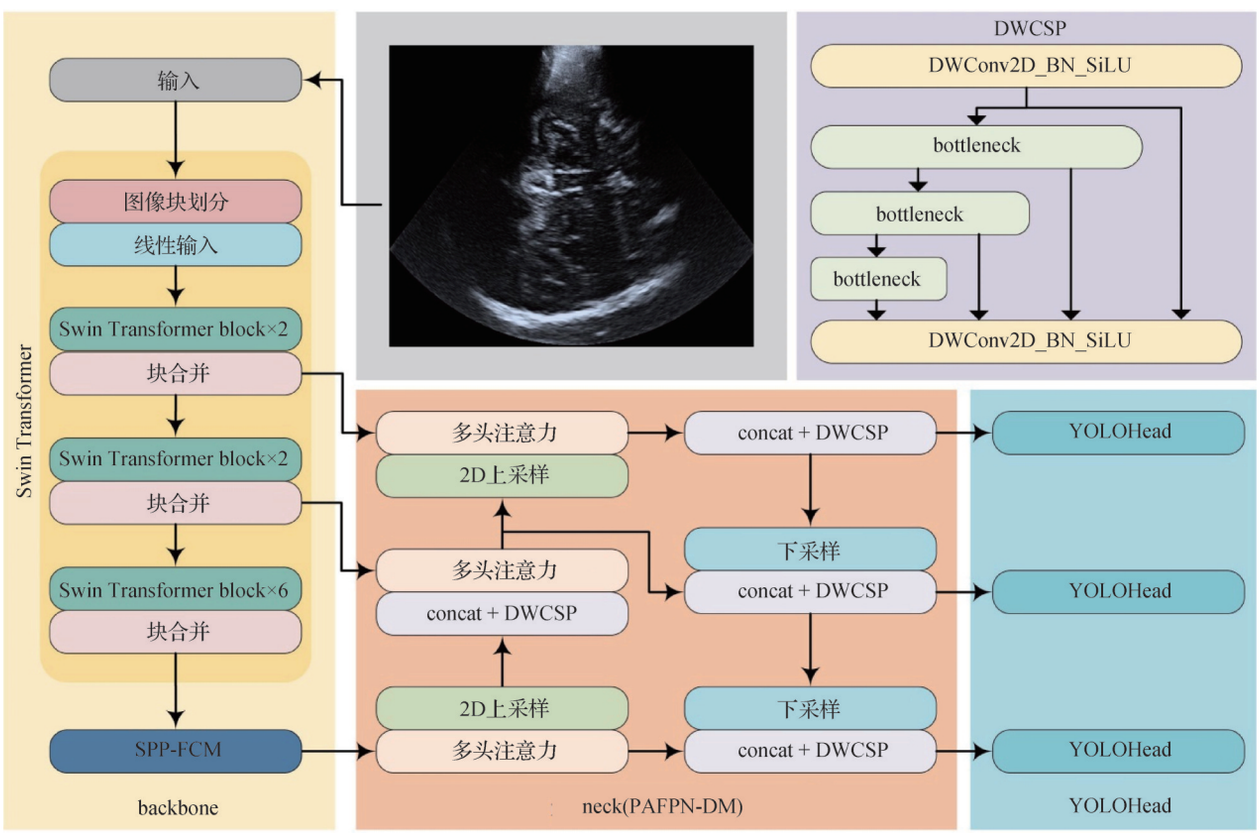


图3　YOLO-SF-TV网络结构

Fig. 3　The network structure of YOLO-SF-TV

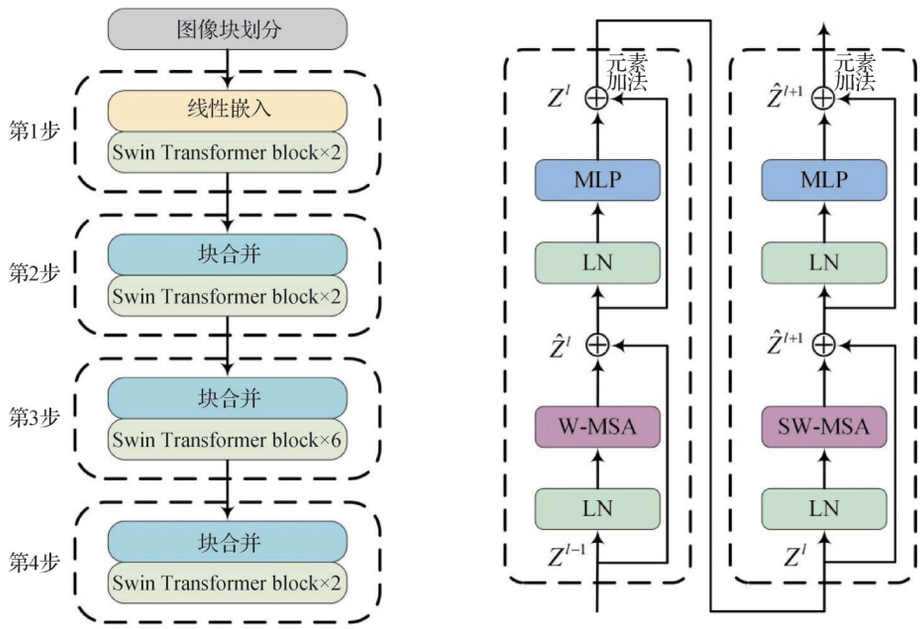
2.1　Swin Transformer模块

基于CNN的CSPDarknet主干网络对复杂背景的TCS图像中目标的全局信息建模能力不足,尤其是在捕捉长距离依赖方面,本文通过引入基于Swin Transformer架构的主干特征提取网络,合理地整合三脑室目标中不同尺度上的有效特征,帮助网络更好地处理多尺度目标检测任务。Swin Tranformer通过分层结构,对每层特征图进行下采样,形成输入图像逐级下采样特征图,通过渐进式下采样避免一次性下采样导致的信息损失问题(Liu等,2021)。Swin Transformer结构如图4所示,图4(a)是模块总体结构,Swin Transformer模块由图像分块模块(patch partition)、线性嵌入模块(liner embedding)、分块合并模块(patch merging)和Swin Transformer block这4个部分组成。

基于Swin Transformer的主干网络首先将输入的经颅超声三脑室图像通过图像分块模块划分为4×4大小的不重叠图像块。并通过线性嵌入模块分别对以上16个图像块序列化处理。随后对每个序列化后的窗口通过Swin Transformer block进行特征提取,在提取特征完成后对以上信息进行分块合并操作。通过重复进行Swin Transformer block和分块合并操作,实

现在降低图像尺度的同时进行图像特征维度的堆叠。

Swin Transformer block 如图4(b)所示，其主要由归一化模块(layer normalization，LN)、窗口多头自注意力模块(windowed multi-head self-attention，W-MSA)、移动窗口多头自注意力模块(shifted window multi-head self-attention，SW-MSA)和多层感知模块(multi-layer perceptron，MLP)这4个部分组成。Swin Transformer block通过层归一化处理、窗口自注意力计算、多层感知处理和移动窗口的多头自注意力计算实现对输入信息的特征计算。



(a) Swin Transformer整体结构　　(b) Swin Transformer block结构

图4　Swin Transformer结构

Fig. 4　The structure of Swin Transformer((a)the overall structure of swin Transformer;(b)the structure of Swin Transformer block)

2.2　多头注意力机制

注意力机制能通过自适应地调整不同特征的重要性来提高模型的表达能力。针对模型训练过程中由于特征尺度和维度跨度过大导致信息丢失的问题，本文在模型中引入多头注意力机制增强特征层外部有效融合多尺度信息。通过这种策略实现模型能够同时关注不同尺度的特征，增强对大、中、小目标的细节和全局信息。

多头注意力机制如图5所示，在输入的特征图中进行基于查询$\boldsymbol{Q}$(query)、键$\boldsymbol{K}$(key)和值$\boldsymbol{V}$(value)的特征映射操作，具体为

$$f_{\text{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=f_{\text{softmax}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

式中，d_k为各键的特征维度，用于权重缩放，经过softmax归一化至[0,1]区间。设计通过将多头注意力分为8个特征子空间进行特征计算，并将子空间的结果进行注意力输出，得到8个注意力头$Z_0\sim Z_7$，接着将所有注意力头拼接并与另一个可学习矩阵$\boldsymbol{W}$线性变换得到多头MulHead。具体为

$$Z_i=f_{\text{Attention}}(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},\boldsymbol{V}\boldsymbol{W}_i^{V}) \tag{2}$$

式中，$\boldsymbol{W}_i^{Q}$、$\boldsymbol{W}_i^{K}$、$\boldsymbol{W}_i^{V}$分别表示$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$的权重矩阵。

$$f_{\text{MulHead}}=f_{\text{Concat}}(Z_1,\cdots,Z_8)\boldsymbol{W} \tag{3}$$

式中，$\boldsymbol{W}$为线性变换的权重，Z_i为多头注意力中的第i个注意力头，f_{Concat}是将多个特征头进行拼接，f_{MulHead}为最后输出结果。通过在特征融合部分引入多头注意力机制，使网络能够从不同空间中学习到更多的特征信息，提高模型的特征表达能力。

2.3　SPP-FCM模块

在YOLOv8网络中SPPF模块借鉴SPP(spatial pyramid pooling)的空间金字塔思想，通过采用不同尺度池化层的方法对不同维度的信息进行特征提取，实现局部和全局信息的融合。但是SPPF特征提取过程主要依赖连续池化操作，在特征提取过程中对于噪声和异常值较为敏感，会影响模型的稳定性和准确性。为此，本文提出了一种改进的SPP-FCM模块。如图6所示，SPP-FCM结构上结合了YOLOv7

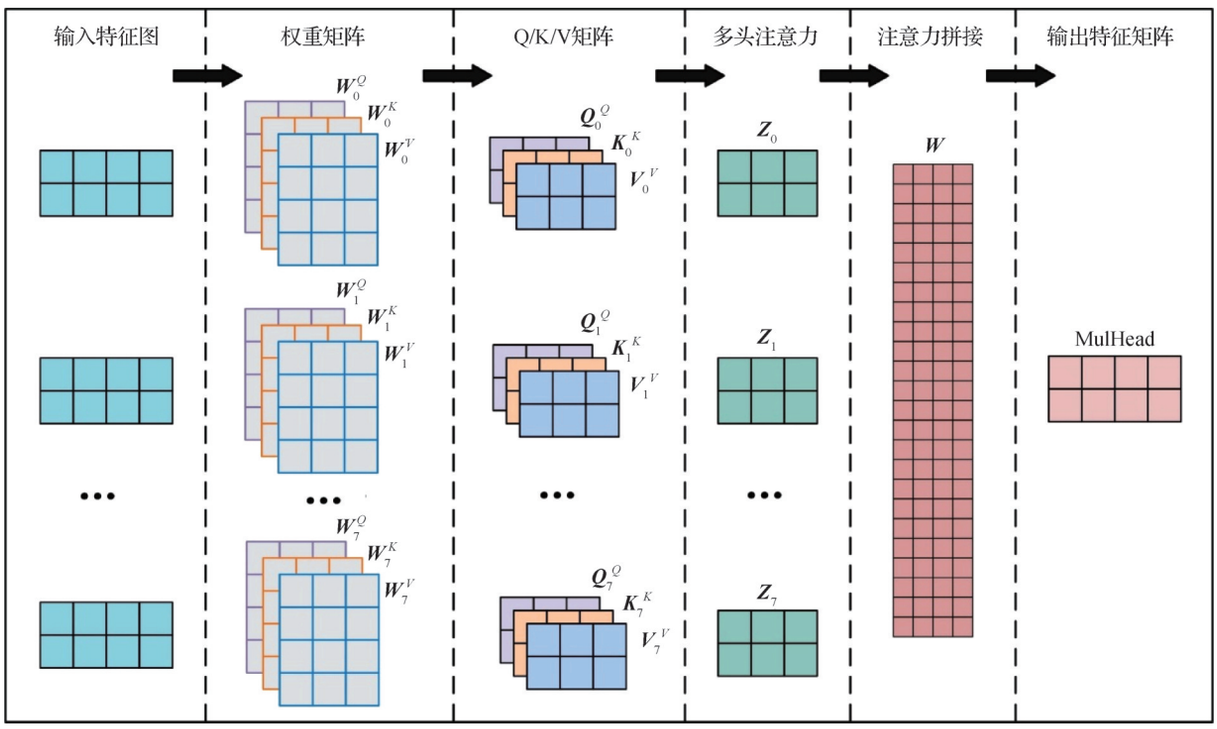


图5　多头注意力结构

Fig. 5　The structure of multi-head attention mechanism

中SPPCSPC的分层并行设计的思路，通过交叉阶段导向不同的拼接连接，实现在保留更多原始信息的情况下增强特征融合能力，并减少了信息的丢失。同时，针对性地在多级池化部分引入多头注意力机制，降低模型在提取多维度特征的过程中对噪声和异常值的敏感性。

SPP-FCM模块连接到Swin Transformer特征提取网络最高维度特征输出层后，用来对多维度的信息进行特征重组。SPP-FCM首先对输入的特征图进行分组处理，分别进行连续的1×1、5×5、9×9和13×13大小的三级堆叠池化操作，实现保证网络多维度感受野的同时加强特征融合能力。接下来对5个并列的特征维度连接多头注意力模块，实现多层级特征提取。最后通过拼接等操作连接各个维度特征序列得到输出特征。

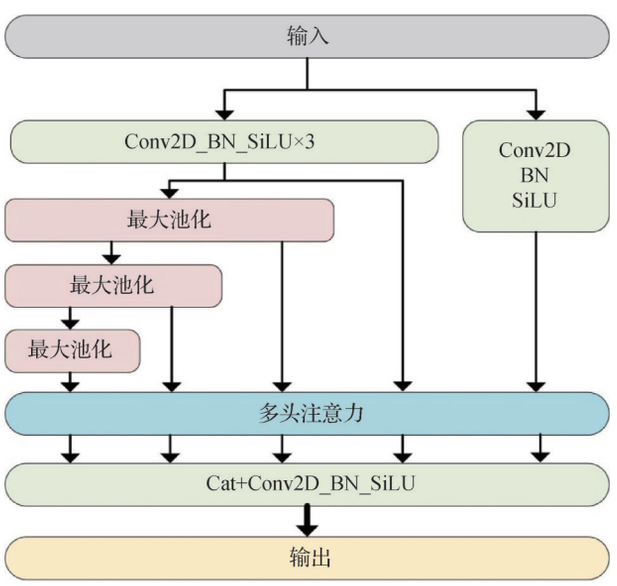


图6　SPP-FCM模块结构

Fig. 6　The structure of SPP-FCM module

2.4　PAFPN-DM模块

在TCS图像三脑室检测中，通常包含不同尺度和复杂结构的解剖学目标，这使得多尺度目标检测成为一项具有挑战性的任务。而YOLOv8网络中主干网络输出的3个特征层进行加强特征融合的结构存在计算复杂度高、信息冗余以及位置一致性差等缺点，这也限制了模型在医学设备资源受限、对准确性和解释性要求较高的情境下的实际应用。为此，本文将多头注意力机制和深度可分离卷积引入网络加强特征提取neck部分，设计出全新的PAFPN-DM模块用于解决以上问题。如图7所示，PAFPN-DM模块通过自顶向下的路径和自底向上的路径，实现了在不同尺度和维度的特征融合。

PAFPN-DM分别对backbone输出的3个特征层

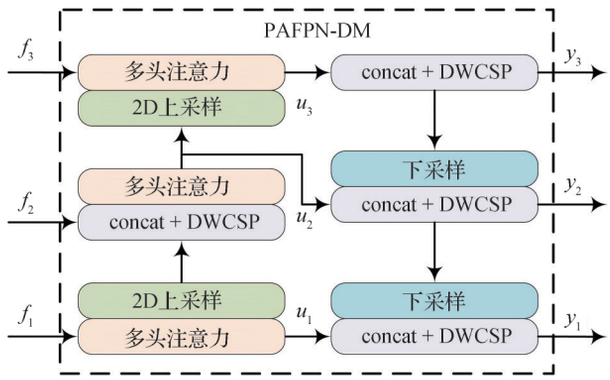


图7　PAFPN-DM模块结构

Fig. 7　The structure of PAFPN-DM module

f_1, f_2, f_3自底下向上地路径增强，生成3个输出特征层y_1，y_2，y_3到YOLOHead中的预测结果。PAFPN-DM模块分为4个步骤：

1）主干网络经过SPP-FCM模块输出的特征层f_1经过多头注意力模块后，生成特征层u_1，接着进行两倍上采样，特征层尺度从20 × 20变成40 × 40，并与主干网络输出的特征层f_2堆叠后经过多头注意力模块处理，生成特征层u_2；

2）特征层u_2经过上采样处理，特征层尺寸变成80 × 80，并与主干网络输出的特征层f_3堆叠后经过多头注意力模块处理堆叠生成特征层u_3。对u_3通过DWCSP模块处理生成PAFPN-DM输出特征层y_3；

3）特征层y_3经过下采样处理，特征层变回40 × 40，然后与特征层u_2堆叠并经过DWCSP模块处理，生成PAFPN-DM输出特征层y_2；

4）特征层y_2经下采样处理，特征层尺寸变回20 × 20，与特征层u_1堆叠并经过DWCSP模块处理，生成PAFPN-DM输出特征层y_1。

3　实验与分析

3.1　实验环境与模型训练

为验证本文算法的有效性，在相同验证集下设计实验并进行分析。实验硬件选择、环境配置和相关参数设置如表1所示。实验中Epoch设置150次，实验初始学习率设置为0.001，并且每完成50次迭代就将学习率衰减为原本的0.1倍，学习动量设置为0.9，衰减系数设置为0.000 5，每次迭代批量大小设置为8。

表1　模型训练环境

Table 1　Model training environment

参数	设置
CPU	Intel Xeon Gold 5218
GPU	RTX3090
内存大小	63 G
显存大小	24 GB
Python版本	3.8
CUDA版本	11.8
深度学习框架	PyTorch 2.0.0
Batch Size	8
初始学习率	0.001

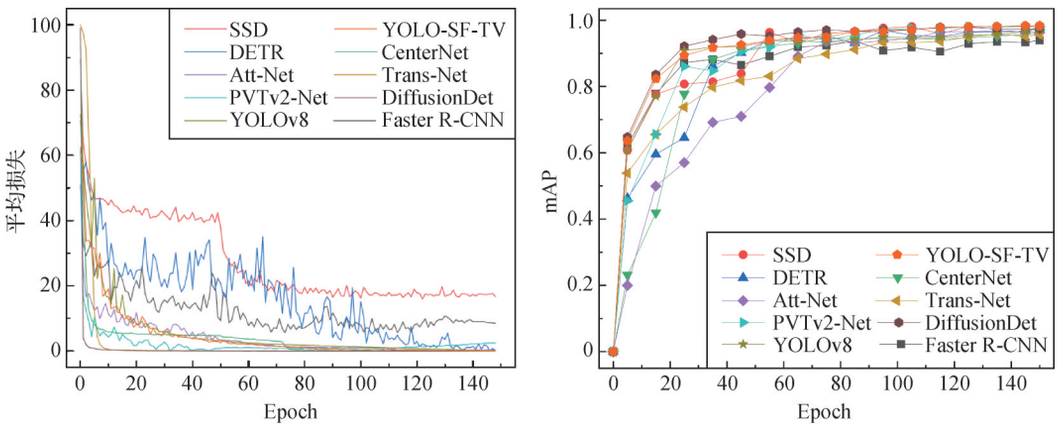
3.2　评价指标

本文使用精确度（precision）、召回率（recall）和平均精确度均值（mean average precision，mAP）作为三脑室目标检测性能评价指标。

3.3　实验结果与分析

在相同的平台以及数据集上对本文提出的YOLO-SF-TV算法以及9种对比实验算法进行训练和测试，并将各个算法得到的最优模型在测试集中进行性能测试。各个模型训练过程中的损失曲线以及mAP曲线如图8所示。实验表明，在训练过程中所有模型在前60个Epoch损失函数曲线下降速度最快。同时，与之相对应的mAP曲线上升速度最快。本文改进算法在训练过程损失函数下降稳定性和波动都较小。同时，在训练过程中的mAP上升曲线也更为平稳，出现的波动较小。

在相同实验环境以及验证集下完成对各网络模型的性能测试，具体实验结果见表2。其中FLOPs为模型的计算量大小，Params为模型参数量大小，二者均在输入图像尺寸为640 × 640像素下计算得出。在测试集下，本文算法在精确率、召回率和mAP上均为最优，其中最重要的指标mAP相比于Faster R-CNN、SSD（single shot multiBox detector）、DETR（detection Transformer）（Mohamed等，2021）、Center-Net（Guo等，2021）、Att-Net（于典 等，2023）、Trans-Net（郝文月 等，2024）、PVTv2-Net（周雪 等，2024）、DiffusionDet（Chen等，2023）和YOLOv8等典型检测算法分别提高了9.77%、3.25%、5.96%、2.22%、3.80%、2.95%、3.19%、2.35%和2.12%。



(a) loss曲线图　　(b) mAP曲线图

图 8　各模型 Loss 曲线和 mAP 曲线

Fig. 8　Loss curve and mAP curve for each model ((a) loss curves; (b) mAP curves)

表 2　不同模型效果对比

Table 2　Comparison of different model effects

模型	准确度/%	召回率/%	mAP/%	Params/M	FLOPs/G
Faster R-CNN	52.97	95.22	88.92	28.29	174.06
SSD	96.30	87.64	95.44	**23.88**	137.03
DETR	91.87	94.17	92.73	55.68	67.11
CenterNet	96.34	90.00	96.47	32.66	**54.83**
Att-Net	94.29	92.70	94.89	70.82	94.45
Trans-Net	93.56	93.82	95.74	43.63	82.70
PVTv2-Net	95.65	92.70	95.50	138.96	126.75
DiffusionDet	94.41	94.94	96.34	188.13	146.57
YOLOv8	96.53	93.82	96.57	68.15	129.06
YOLO-SF-TV(本文)	**98.60**	**96.98**	**98.69**	92.71	131.94

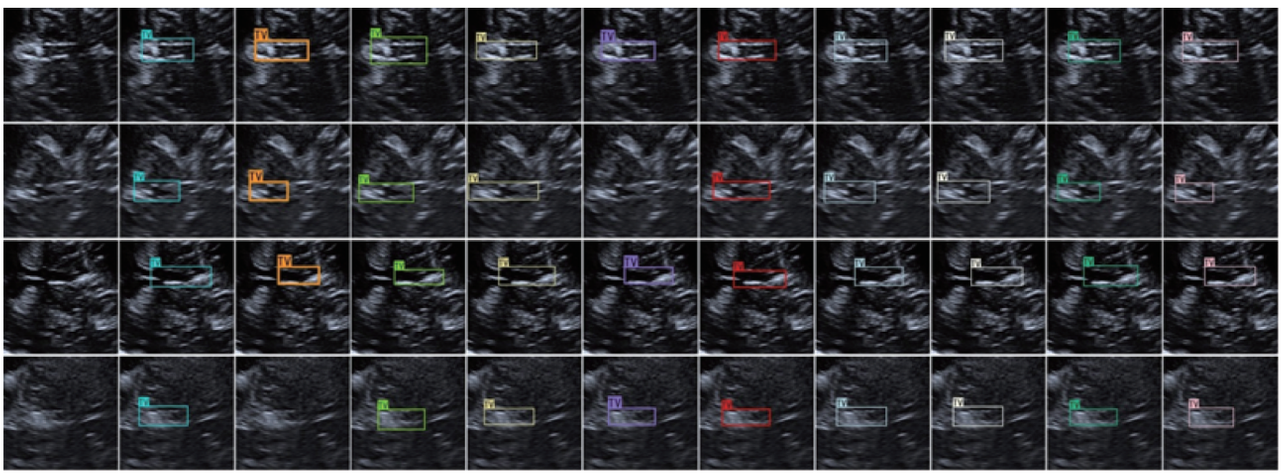
注：加粗字体表示各列最优结果。

为了更充分验证各个模型在TCS图像中三脑室检测效果，本文在验证集中对每个算法实际检测效果进行测试，测试结果如图9所示。图9第1行图像中，当TCS图像中三脑室清晰可见时，所有算法都能够准确且恰当地检测出来；图9第2—4行图像中，当三脑室边界出现模糊时，SSD、Faster R-CNN、DETR和CenterNet算法检测目标存在边缘遗漏的情况，特别是SSD在第3行、CenterNet在第2行图像中未能检测到目标。从图9(g)—图9(k)的对比实验中可以看出来，Att-Net、Trans-Net、PVTv2-Net、DiffusionDet和YOLOv8都能够检测出目标，但是检测框框选目标的精确度相比于YOLO-SF-TV算法表现更差，这也表明在检测过程中本文模型定位准确性更好。

3.4　消融实验

为了进一步验证算法的有效性，在TCS图像数据集中进行消融实验分析，结果见表3。

第1组实验将YOLOv8主干网络替换为Swin Transformer进行特征提取；第2—4组实验将分别在第1组实验的基础上增加SPPF、SPP-FCM和PAFPN-DM模块，以此来检测各个模块对网络性能的影响；第5和6组实验分别是在主干网络特征提取后增加的SPPF和SPP-FCM模块与PAFPN-DM模块之间组合对比验证检测性能。在第2和3组实验中，对比SPP-FCM模块和SPPF模块对网络的影响。实验结果表明，两者对网络的检测性能分别提高了0.63%和1.63%，相比之下SPP-FCM性能更优。第4组实



(a) 原图 (b) YOLO-SF-TV (c) SSD (d) FasterR-CNN (e) DETR (f) CenterNet (g) Att-Net (h) Trans-Net (i) PVTv2-Net (j) DiffusionDet (k) YOLOv8

图9 不同算法的检测结果

Fig. 9 Detection results of different algorithms ((a) original images; (b) YOLO-SF-TV; (c) SSD; (d) Faster R-CNN; (e) DETR; (f) CenterNet; (g) Att-Net; (h) Trans-Net; (i) PVTv2-Net; (j) DiffusionDet; (k) YOLOv8)

表3 消融实验

Table 3 Ablation experiment

ID	SPPF	SPP-FCM	PAFPN-DM	mAP/%
1	×	×	×	95.02
2	√	×	×	95.65
3	×	√	×	96.65
4	×	×	√	95.71
5	√	×	√	97.00
6	×	√	√	**98.69**

注：加粗字体表示最优结果。“√”和“×”分别表示使用和未使用对应模块。

验使用PAFPN-DM模块替换第1组实验中的PAFPN模块，实验结果表明，融合了多头注意力机制和深度可分离卷积的PAFPN-DM模块对网络检测性能有所提升。第5和6组实验组合了SPPF、SPP-FCM和PAFPN-DM这3个模块，结果表明网络检测性能分别增加了1.98%和3.67%。

4 结 论

在TCS三脑室图像研究中，通常需要专业医生手动筛选病人样本，不仅费时费力，而且检测准确率和效率低。为此，本文展开了基于深度学习的经颅超声三脑室图像自动检测方法研究。实验制作了一个包含2 400幅经颅超声图像及其相应标注的数据集。以此数据集为基础，针对TCS三脑室图像检测困难的问题，本文提出了YOLO-SF-TV检测算法，并与典型的目标检测算法进行了对比实验且取得了最优效果。YOLO-SF-TV算法在YOLOv8的基础上结合Swin Transformer对全局特征提取能力的优势，并针对Swin Transformer特征提取的高维度特征层设计全新的SPPD-FCM模块，增强模型对多维度特征的融合能力。同时在PAFPN结构的基础上提出融合深度可分离卷积和多头注意力机制的PAFPN-DM模块，使网络加强特征提取部分在增加特征融合部分稀疏性的同时保证了检测精度。实验结果表明，本文提出的YOLO-SF-TV模型在三脑室自动检测任务中取得了良好的效果，同时本文收集的数据集将有助于推动经颅超声图像三脑室检测问题的研究。

YOLO-SF-TV检测算法在召回率方面表现出较高的性能，然而在精确度方面存在较低的表现，这可能导致检测结果中存在较多的假阳性。尽管在实际应用中通常会有医生进行复核验证，但这仍然可能增加额外的工作量。因此，后续工作需要进一步提高算法的精确度，以提升检测结果的准确性和可靠性。同时，在实际使用中模型对硬件要求高，具体使用中耗时相对于实时任务而言仍然较大。在后续的研究中，应该探索更加轻量化的网络设计和处理方法，以便将模型部署在硬件需求更低的设备上进行实时检测。

本文提出的检测算法仅针对经颅超声图像三脑室目标，但是经颅超声图像中还存在例如中脑、中缝核和黑质等重要区域。在后续工作中检测PD患者

TCS图像中所有重要区域，这对辅助医生临床诊断更具有研究价值和意义。

参考文献（References）

Baccouche A，Garcia-Zapirain B，Zheng Y F and Elmaghraby A S. 2022. Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques. Computer Methods and Programs in Biomedicine，221：#106884［DOI：10.1016/j.cmpb.2022.106884］

Bochkovskiy A，Wang C Y and Liao H Y M. 2020. YOLOv4：optimal speed and accuracy of object detection［EB/OL］.［2024-05-20］. https://arxiv.org/pdf/2004.10934.pdf

Cao H，Wang Y Y，Chen J，Jiang D S，Zhang X P，Tian Q and Wang M N. 2022. Swin-Unet：Unet-like pure Transformer for medical image segmentation//Proceedings of 2022 ECCV：European Conference on Computer Vision. Tel Aviv，Israel：Springer：205-218［DOI：10.1007/978-3-031-25066-8_9］

Chen S F，Sun P Z，Song Y B and Luo P. 2023. DiffusionDet：diffusion model for object detection//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris，France：IEEE：19773-19786［DOI：10.1109/ICCV51070.2023.01816］

Croitoru F A，Hondru V，Ionescu R T and Shah M. 2023. Diffusion models in vision：a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence，45(9)：10850-10869［DOI：10.1109/TPAMI.2023.3261988］

Fu X Y，Zhang Y C，Ding C W，Yang M，Song X，Wang C S，Chen X F，Zhang Y，Sheng Y J，Mao P，Mao C J and Liu C F. 2021. Association between homocysteine and third ventricle dilatation，mesencephalic area atrophy in Parkinson's disease with cognitive Impairment. Journal of Clinical Neuroscience，90：273-278［DOI：10.1016/j.jocn.2021.06.006］

Gao H L，Qu Y，Chen S C，Yang Q M，Li J Y，Tao A Y，Mao Z J and Xue Z. 2024. Third ventricular width by transcranial sonography is associated with cognitive impairment in Parkinson's disease. CNS Neuroscience and Therapeutics，30(2)：#e14360［DOI：10.1111/cns.14360］

Ge Z，Liu S，Wang F，Li Z and Sun J. 2021. YOLOX：exceeding YOLO series in 2021［EB/OL］.［2024-05-20］. https://arxiv.org/pdf/2107.08430.pdf

Ghazouani F，Vera P and Ruan S. 2023. Efficient brain tumor segmentation using Swin Transformer and enhanced local self-attention. International Journal of Computer Assisted Radiology and Surgery，19(2)：273-281［DOI：10.1007/s11548-023-03024-8］

Guo H Y，Yang X，Wang N N and Gao X B. 2021. A CenterNet++ model for ship detection in SAR images. Pattern Recognition，112：#107787［DOI：10.1016/j.patcog.2020.107787］

Hao W Y，Cai H Y，Zuo T T，Jia Z W，Wang Y and Chen X D. 2024. Intravascular ultrasound image segmentation fusing Transformer branch and topology enforcement. Laser and Optoelectronics Progress，61(12)：#1237008（郝文月，蔡怀宇，左廷涛，贾忠伟，汪毅，陈晓冬. 2024. 融合Transformer分支和拓扑强制的血管内超声图像分割方法. 激光与光电子学进展，61(12)：#1237008）［DOI：10.3788/LOP231918］

Heravi F S，Naseri K and Hu H H. 2023. Gut microbiota composition in patients with neurodegenerative disorders（Parkinson's and Alzheimer's）and healthy controls：a systematic review. Nutrients，15(20)：#4365［DOI：10.3390/nu15204365］

Liu Z，Lin Y T，Cao Y，Hu H，Wei Y X，Zhang Z，Lin S and Guo B N. 2021. Swin Transformer：hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：9992-10002［DOI：10.1109/ICCV48922.2021.00986］

Livingston G，Huntley J，Sommerlad A，Ames D，Ballard C，Banerjee S，Brayne C，Burns A，Cohen-Mansfield J，Cooper C，Costafreda S G，Dias A，Fox N，Gitlin L N，Howard R，Kales H C，Kivimäki M，Larson E B，Ogunniyi A，Orgeta V，Ritchie K，Rockwood K，Sampson E L，Samus Q，Schneider L S，Selbæk G，Teri L and Mukadam N. 2020. Dementia prevention，intervention，and care：2020 report of the Lancet Commission. The Lancet，396(10248)：413-446［DOI：10.1016/S0140-6736(20)30367-6］

Lotter W，Diab A R，Haslam B，Kim J G，Grisot G，Wu E，Wu K，Onieva J O，Boyer Y，Boxerman J L，Wang M Y，Bandler M，Vijayaraghavan G R and Gregory Sorensen A. 2021. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nature Medicine，27(2)：244-249［DOI：10.1038/s4159 1-020-01174-9］

Meng Q Q，Gao Y，Lin H，Wang T J，Zhang Y R，Feng J，Li Z S，Xin L and Wang L W. 2022. Application of an artificial intelligence system for endoscopic diagnosis of superficial esophageal squamous cell carcinoma. World Journal of Gastroenterology，28(37)：5483-5493［DOI：10.3748/wjg.v28.i37.5483］

Mohamed C，Peter Z，Ross B and Stephen H. 2021. DEFT：detection embeddings for tracking［EB/OL］.［2024-05-20］. https://arxiv.org/pdf/2102.02267.pdf

Monaco D，Berg D，Thomas A，Di Stefano V，Barbone F，Vitale M，Ferrante C，Bonanni L，Di Nicola M，Garzarella T，Marchionno L P，Malferrari G，Di Mascio R，Onofrj M and Franciotti R. 2018. The predictive power of transcranial sonography in movement disorders：a longitudinal cohort study. Neurological Sciences，39(11)：1887-1894［DOI：10.1007/s10072-018-3514-z］

Özbay E and Özbay F A. 2023. Interpretable features fusion with precision MRI images deep hashing for brain tumor detection. Computer Methods and Programs in Biomedicine，231：#107387［DOI：10.1016/j.cmpb.2023.107387］

Redmon J and Farhadi A. 2018. YOLOv3：an incremental improvement［EB/OL］.［2024-05-20］. https://arxiv.org/pdf/1804.02767.pdf

Tarimo S A, Jang M A, Ngasa E E, Shin H B, Shin H and Woo J. 2024. WBC YOLO-ViT: 2 way-2 stage white blood cell detection and classification with a combination of YOLOv5 and vision transformer. Computers in Biology and Medicine, 169: #107875 [DOI: 10.1016/j.compbiomed.2023.107875]

Vlaar A, Tromp S C, Weber W E, Hustinx R M and Mess W H. 2011. The reliability of transcranial duplex scanning in parkinsonian patients: comparison of different observers and ultrasound systems. Ultraschall in der Medizin-European Journal of Ultrasound, 32(S1): 83-88 [DOI: 10.1055/s-0028-1109945]

Wang C Y, Bochkovskiy A and Liao H Y M. 2023, YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]

Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, Lu T, Luo P and Shao L. 2022. PVT v2: improved baselines with pyramid vision Transformer. Computational Visual Media, 8(3): 415-424 [DOI: 10.1007/s41095-022-0274-8]

Yang Y Z, Yuan Y, Zhang G, Wang H, Chen Y C, Liu Y C, Tarolli C G, Crepeau D, Bukartyk J, Junna M R, Videnovic A, Ellis T D, Lipford M C, Dorsey R and Katabi D. 2022. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. Nature Medicine, 28(10): 2207-2215 [DOI: 10.1038/s41591-022-01932-x]

Yin H, Vahdat A, Alvarez J M, Mallya A, Kautz J and Molchanov P. 2022. A-ViT: adaptive tokens for efficient vision Transformer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10799-10808 [DOI: 10.1109/CVPR52688.2022.01054]

Yu D, Peng Y J and Guo Y F. 2023. Ultrasonic image segmentation of thyroid nodules-relevant multi-scale feature based h-shape network. Journal of Image and Graphics, 28(7): 2195-2207 (于典, 彭延军, 郭燕飞. 2023. 面向甲状腺结节超声图像分割的多尺度特征融合"h"形网络. 中国图象图形学报, 28(7): 2195-2207) [DOI: 10.11834/jig.220078]

Zhou T, Liu S, Dong Y L, Bai J and Lu H L. 2023. Parallel decomposition adaptive fusion model: cross-modal image fusion of lung tumors. Journal of Image and Graphics, 28(1): 221-233 (周涛, 刘珊, 董雅丽, 白静, 陆惠玲. 2023. 肺部肿瘤跨模态图像融合的并行分解自适应融合模型. 中国图象图形学报, 28(1): 221-233) [DOI: 10.11834/jig.210988]

Zhou X, Bai Z Y, Lu Q J and Fan S L. 2023. Colorectal polyp segmentation combining pyramid vision Transformer and axial attention. Computer Engineering and Applications, 59(11): 222-230 (周雪, 柏正尧, 陆倩杰, 樊圣澜. 2023. 融合视觉Transformer和轴向注意的结直肠息肉分割. 计算机工程与应用, 59(11): 222-230) [DOI: 10.3778/j.issn.1002-8331.2203-0110]

作者简介

万奥,男,硕士研究生,主要研究方向为计算机视觉和深度学习。E-mail:782674190@qq.com

周晓,通信作者,男,教授,主要研究方向为机计算机视觉、智能感知和高性能嵌入式系统。E-mail:zhouxiao@whut.edu.cn

高红铃,女,主治医生,主要研究方向为帕金森病诊断与治疗。E-mail:780116631@qq.com

薛峥,女,教授,主治医生,主要研究方向为帕金森病、脑血管疾病和神经重症诊断与治疗。E-mail:xuezheng@hust.edu.cn

牟新刚,男,副教授,主要研究方向为计算机视觉、图像处理和嵌入式系统设计。E-mail:sunnymou@whut.edu.cn