

基于改进YOLOv7的脑部MRI影像肿瘤检测算法

白静毅^{1,2}, 吴义熔^{1,2,3}, 李小龙^{1,2}, 孙水发^{1,2,4}

1. 三峡大学计算机与信息学院/水电工程智能视觉监测湖北省重点实验室, 湖北 宜昌 443002; 2. 三峡大学计算机与信息学院/智慧医疗宜昌市重点实验室, 湖北 宜昌 443002; 3. 北京师范大学人文和社会科学高等研究院, 广东 珠海 519087; 4. 杭州师范大学信息科学与技术学院, 浙江 杭州 311121

【摘要】脑部MRI影像数据具有数据量大、易受噪声和伪影干扰等特点,这些特点对种类、形状、边界既相似又复杂多变的脑肿瘤检测和分析提出了速度和准确度提升的挑战。基于此,在YOLOv7算法基础上提出一系列改进方法来提高检测的精度和速度。首先在特征提取阶段使用部分卷积PConv,以降低模型的参数量,提高整体检测速度。其次针对脑肿瘤复杂多变的特点,在特征提取时引入三维空间注意力机制SimAM,以提高模型对重要影像特征的关注。最后用WIoU替换原IoU损失函数,在边界框回归时提高对普通质量锚框的关注,以进一步提高检测精度。通过在公开的两个脑肿瘤数据集Brain_Tumor和Glioma_of_test上进行实验,改进后的模型mAP检测精度分别为96.9%和92.8%,相比YOLOv7原模型提高1.4%和2.4%;FPS每秒处理的图像数分别为162.7和158.1,相比YOLOv7原模型提高6.4和18.2,可以更为有效地检测脑部MRI影像中的脑肿瘤。

【关键词】脑肿瘤;YOLOv7;部分卷积;三维空间注意力;动态注意力

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)03-0336-11

Algorithm for brain MRI tumor detection based on improved YOLOv7

BAIJingyi^{1,2}, WU Yirong^{1,2,3}, LI Xiaolong^{1,2}, SUN Shuifa^{1,2,4}

1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering/College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; 2. Yichang Key Laboratory of Intelligent Medicine/College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; 3. Institute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University, Zhuhai 519087, China; 4. School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

Abstract: Brain MRI data is characterized by large volumes and susceptibility to noise and artifacts, which pose significant challenges of improving the speed and accuracy of brain tumor detection and analysis due to the tumors' diverse types, shapes, and boundaries that are both similar and highly variable. Therefore, a series of improvements based on YOLOv7 algorithm are proposed for enhancing detection precision and speed: (1) employing partial convolution during feature extraction to reduce the model's parameters and improve overall detection speed; (2) in light of the complex variability of brain tumors, introducing a three-dimensional spatial attention mechanism during feature extraction to enhance the model's focus on critical image features; (3) replacing the original IoU loss function with WIoU to increase the attention to medium-quality anchor boxes during bounding box regression for further improving detection accuracy. Experiments conducted on two public brain tumor datasets, Brain_Tumor and Glioma_of_test, show that the improved model achieves mAP of 96.9% and 92.8%, which are 1.4% and 2.4% higher than the original YOLOv7 model, and the frames per second reach 162.7 and 158.1, showing improvements of 6.4 and 18.2, respectively. These enhancements enable more effective detection of brain tumors in MRI images.

Keywords: brain tumor; YOLOv7; partial convolution; three-dimensional spatial attention; dynamic attention

【收稿日期】2024-11-28

【基金项目】国家自然科学基金(20BTQ066)

【作者简介】白静毅,硕士研究生,研究方向:智慧医疗,E-mail: 1214518574@qq.com

【通信作者】吴义熔,教授,博士生导师,研究方向:医学大数据和精确医学,E-mail: yrwu@bnu.edu.cn

前言

脑肿瘤是由不受控制的细胞增殖导致的异常生长的组织组成的,这种组织在大脑内没有生理功能。肿瘤不仅会增加大脑的体积和压力,还会引起肿胀,

导致异常的神经症状。基于计算机视觉的脑肿瘤自动检测的意义在于提高脑肿瘤诊断的效率,优化资源配置,推动医学科研和临床实践的发展。在众多脑成像技术中,磁共振成像(MRI)可以提供关于人体组织的丰富信息,并且具有高软组织分辨率和多平面成像能力,有助于更准确地定位病变。脑肿瘤MRI数据具有复杂性、多样性和相似性等特点^[1]。胶质瘤(Glioma)、脑膜瘤(Meningioma)和垂体瘤(Pituitary)是大脑中常见的3类肿瘤,它们在MRI图像中的表现如图1所示。胶质瘤边界模糊,信号与正常脑组织相似,形状不规则,通常分布于大脑的白质中,容易与正常脑白质混淆,表现为不规则的低信号或高信号区域。相比之下,脑膜瘤多为良性,呈圆形

或椭圆形,常紧贴硬脑膜,推移周围脑组织而不浸润其中,边界清晰。由于脑膜瘤与硬脑膜关系密切,早期可能被误认为是正常脑膜增厚。此外,脑膜瘤可能压迫邻近的脑组织,导致非肿瘤性信号异常,进一步增加识别难度。垂体瘤通常位于蝶鞍内,表现为鞍区的肿块,较大的肿瘤可能压迫周围结构。由于鞍区结构复杂,垂体瘤可能与垂体囊肿等其他鞍区病变难以区分。小型垂体瘤,特别是微腺瘤,因为其体积小、信号较弱,可能会被漏检。脑肿瘤检测工作还往往受到噪声、伪影以及影像质量不佳等干扰,这些干扰可能会对图像的可读性和解译性产生负面影响,增加肿瘤检测的难度。

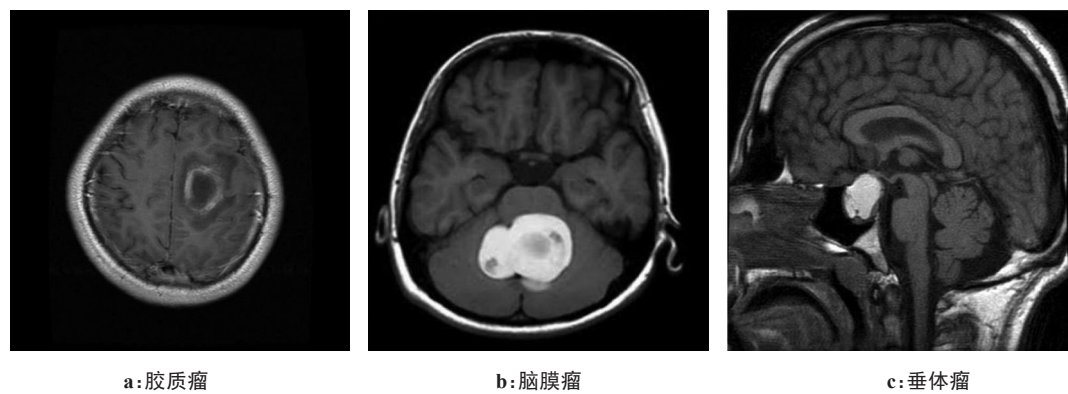


图1 脑肿瘤示例
Figure 1 Examples of brain tumors

传统的脑肿瘤检测方法是先进行特征提取再用分类器区分。Kalbkhani等^[2]使用二维离散小波变换和广义自回归条件异方差对脑肿瘤的细节进行建模,并用线性判别分析器LDA提取特征向量,最后用k邻近算法(KNN)和支持向量机(SVM)对其进行分类预测。Machhale等^[3]先对脑肿瘤影像进行中值滤波和颅骨掩蔽预处理,再提取其对称性、灰度和纹理特征,最后使用KNN、SVM和SVM-KNN进行检测。Parveen等^[4]从图像增强头骨区域分割开始,首先采用FCM进行图像分割,然后利用GLRLM进行特征提取,最后使用SVM对异常图像进行识别。这些方法都预先对图像进行处理,提高脑肿瘤图像的分辨率和对比度,保持异常区域的边缘,提升检测的精度。但传统方法依赖于手工设计图像的特征,会影响对于纹理、边缘、内部结构等复杂特征的提取,导致在精度和速度上都有一定的局限性。随着深度学习技术的快速发展,深度学习方法目前成为肿瘤检测较为常见的方法。深度学习检测模型如RCNN^[5]、YOLO^[6]、SSD^[7],能够自动从影像数据中学习特征表

示而不需要手工设计特征。Abd-Ellah等^[8]利用双阶段自动检测脑肿瘤的方法,在第一阶段使用卷积神经网络(CNN)进行特征提取并使用纠错输出编码支持向量机ECOC-SVM对特征分类,将影像分为正常和异常来检测脑肿瘤,在第二阶段使用基于5层区域的卷积神经网络R-CNN来定位异常MRI中的肿瘤。Noreen等^[9]提出使用两个预先训练好的深度学习模型对脑肿瘤多层次特征进行提取与拼接,并传递给softmax分类器对脑肿瘤分类。Abd-Ellah等^[8]和Noreen等^[9]提出的深度学习方法在脑肿瘤检测结果的精度方面有更好的表现,但在速度方面还有待提升。为了提升检测的速度,在其他类型疾病影像检测的方向,米吾尔依提等^[10]和刘涌涛等^[11]提出使用改进YOLOv7单阶段快速目标检测算法来检测肝囊型包虫病超声图像小病灶和肺结节,通过引入轻量级特征提取主干网络、注意力机制和更优的边界框回归(BBR)损失函数,提高检测的精度和速度,满足快速检测复杂疾病的要求。

针对脑肿瘤种类、形状、位置、边缘的复杂多样

性,为了更好地对复杂特征进行处理并提高速度,本文采用基于YOLOv7^[12]单阶段快速深度学习算法。YOLOv7算法引入了重参数化思想、多尺度跨层聚合以及辅助头训练,更容易在轻量化参数的基础上提取脑肿瘤的细微特征,以提升推理时的速度和精度。为了进一步提高检测的速度和精度,本文在YOLOv7模型的基础上提出以下改进方法:(1)在特征提取时使用部分卷积替换原来的卷积。针对脑肿瘤之间的相似性,不是对输入的所有数据进行统一卷积,而是根据数据的有效性调整卷积核的作用区域,以此提高检测的精度和速度。(2)将三维空间注意力机制加入YOLOv7模型中,同时关注 feature map 的空间域和通道域注意力。针对脑肿瘤的复杂性和多样性,在特征提取阶段,找到信息更丰富的神经元,进而提取到更具重要性的特征,从而提升模型整体检测的精确度。(3)在BBR时使用动态注意力损失函数,以解决脑肿瘤训练实例的不平衡问题。将小梯度增益分配给离群度小的高质量预测回归边界框和离群度高的低质量预测回归边界框,使其专注于普通质量的边界框并减少低质量边界框对BBR的影响。

1 YOLOv7 介绍

YOLOv7是一个端到端的目标检测网络,结构如图2所示,它通过一系列的卷积和池化操作,从输入图像中提取特征,再通过锚框来生成候选边界框,最后以非极大值抑制的方法选取最后的检测结果。其

中,特征提取和边界框回归的方式,对结果的精度和速度有较大的影响。脑肿瘤的形态和大小具有很大的变异性,YOLOv7使用SPPCSPC提取不同尺度下的特征,并进行拼接和通道维度的处理,增加模型对不同尺度目标的感知能力,能更有效地捕捉形状和边界不规则的脑肿瘤的特征。此外,YOLOv7中的ELAN-H可以利用分组卷积来增加特征的基数,通过特征融合和跨层连接,增强模型对复杂细微特征的学习能力,这对于检测小而不明显的脑肿瘤特别重要。YOLOv7还使用引导头预测作为指导来生成从粗到细的分层标签,这些标签分别用于辅助头和引导头学习。通过让较浅的辅助头直接学习引导头已学到的信息,引导头能够专注于学习尚未学习的剩余信息。带辅助头的训练方法通过增强中间层特征的学习,使得模型能够捕捉到更多细微特征,可以显著提高检测的准确性。但是,YOLOv7在特征融合和连接的过程中没有根据特征的重要性来调整融合和连接的方法,这可能会导致检测准确率的下降,将在后面的章节基于此对YOLOv7提出改进。在速度方面,YOLOv7使用重参数化REP在训练阶段增强模型的学习能力,以更好地提取高分辨率MRI影像中脑肿瘤的细微特征,并在推理阶段简化模型结构,以减少模型的计算复杂度,提高推理速度,从而满足临床实时检测大量肿瘤影像的需求。基于此,考虑在重参数化的基础上进一步降低模型的参数量,以进一步提高检测速度。

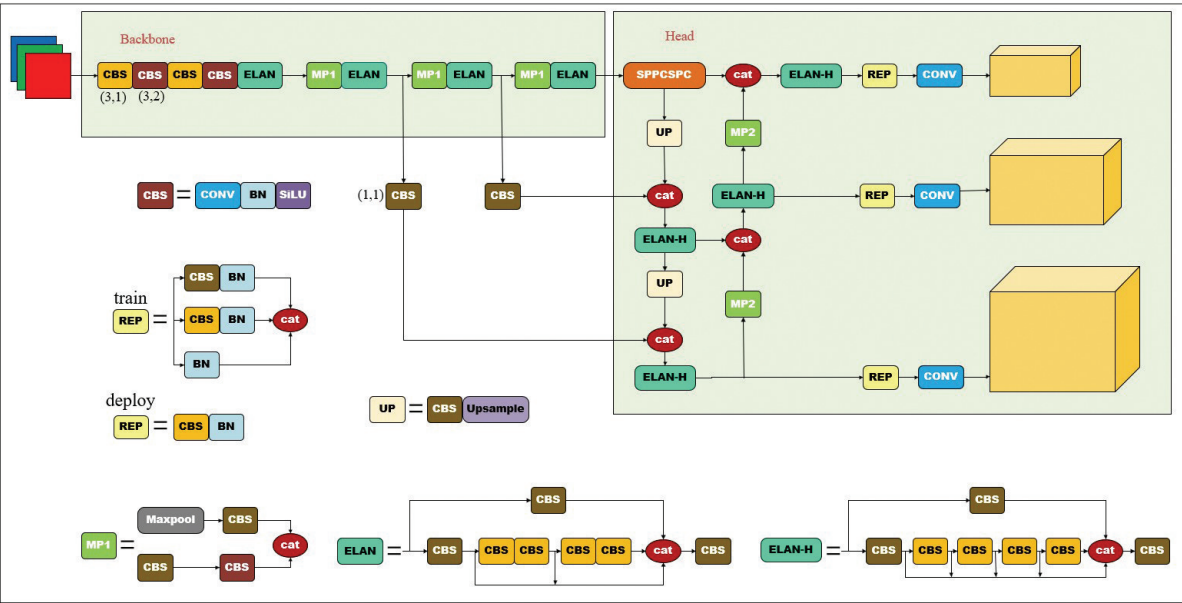


图2 YOLOv7整体结构
Figure 2 Overall structure of YOLOv7

2 改进方法

YOLOv7虽然在脑肿瘤检测任务中有一定精度和速度的提升,但由于YOLOv7在特征提取和BBR阶段没有考虑到特征之间的相对重要性,并且由于脑肿瘤结构、形状、边缘、位置的各异性,不同肿瘤甚至相同肿瘤之间区别较大,解译性良莠不齐,YOLOv7算法本身并不能很好地对其进行特征甄别和处理融合,为此提出一系列改进方法来提高脑肿瘤检测的准确度和速度。

2.1 部分卷积

本文提出使用简单的PConv部分卷积,部分卷积根据数据的有效性调整卷积核的作用区域,只对每个卷积窗口的有效部分进行卷积^[13]。对于冗余度和相似度高的脑肿瘤来说,这样可以减少冗余的计算和内存访问,提高特征处理速度和检测精度。Chen等^[13]提出feature map的特征冗余性:feature map在不同通道上具有高度相似性。对于连续或规则的内存访问,将第一个或最后一个连续的通道视为整个特征图的代表进行计算。本研究提出只在一部分输入特征上应用常规的Conv进行空间特征提取,而对其他的通道保持不变,如图3所示。部分卷积可以减少卷积的计算时间,降低模型整体检测时间,增大速度。卷积的计算延迟时间Delay由计算量FLOPs^[14]和计算速度FLOPs^[15]共同决定,如式(1)所示:

$$\text{Delay} = \frac{\text{FLOPs}}{\text{FLOPS}}$$

(1)

FLOPs是浮点计算次数,包括乘法和加法,只与模型本身有关。FLOPS是每秒钟浮点运算的次数即计算速度,受访问内存的次数影响。之前的研究提出的深度卷积DWConv^[16]和组卷积GConv^[17]都是通过减少计算量FLOPs来使模型速度升高,后面的很多网络MobileNets^[18]、ShuffleNets^[19]、GhostNet^[20]等都应用深度卷积和组卷积的方法。这些方法利用卷积核的冗余性减少计算参数,实现FLOPs的下降,但模型

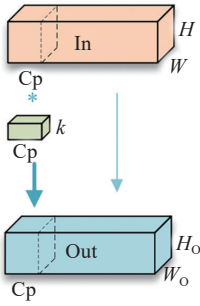
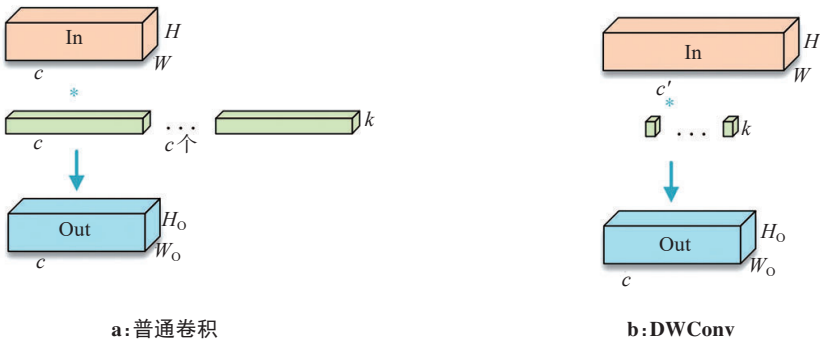


图3 部分卷积
Figure 3 Partial convolution

整体速度提升并不明显。其中一个共同的原因是它们以增加内存访问量为代价减少计算量FLOPs,例如:DWConv在减少计算量FLOPs的过程中,为了补偿下降的精度,相应地增加输入通道数,提高对内存访问的次数,这会降低计算速度FLOPS。为了解决这一问题,本文提出使用部分卷积PConv^[13],在不降低计算速度FLOPS的同时降低计算量FLOPs,从而降低计算时间Delay以提高整个模型的检测速度。

由上述分析可以得出解决问题的关键是在不增加内存访问量的前提下减少总计算次数。这里的内存访问量主要来源有两个:对feature map访问和对卷积核访问。对于输入 $In \in R^{c \times H \times W}$, 输出 $Out \in R^{c \times H_o \times W_o}$, 假设输入输出通道数相同,如图4所示,以一次卷积过程为例。普通卷积的卷积核对输入的每个通道都进行卷积求和,所以普通卷积计算次数FLOPs为 $H \times W \times k^2 \times c^2$, 考虑到输入和输出要访问两次,则其内存访问量为 $H \times W \times 2c + k^2 \times c^2 \approx H \times W \times 2c$ 。而DWConv的每个卷积核只在对应的一个输入通道上进行空间滑动,并将结果贡献于对应的一个输出通道,因此DWConv计算次数为 $H \times W \times k^2 \times c$ 。DWConv虽然有效减少FLOPs,但会导致比较严重的精度下降。因此在实际应用时,要拓



a: 普通卷积 b: DWConv

图4 普通卷积和DWConv
Figure 4 Standard convolution and DWConv

展其通道数补偿精度下降。假设通道数拓展到了 c' ，则DWConv的FLOPs增大为 $H \times W \times k^2 \times c'$ 。拓展通道还会造成对存储器访问量的提升，所以其内存访问量为 $H \times W \times 2c' + k^2 \times c' \approx H \times W \times 2c'$ 。这会导致DWConv的FLOPs相比与普通卷积有所下降，从而导致整个系统的速度下降。

对部分卷积来说，只在几个输入通道 c_p 上应用卷积核，而在其余的通道保持不变。部分卷积的FLOPs为 $H \times W \times (k^2 \times c_p^2 + c^2)$ ，内存访问量为 $H \times W \times 2c_p + k^2 \times c_p^2$ ，实现了在不增加内存访问量的同时降低总计算次数。将普通卷积替换为部分卷积加 1×1 的卷积，如图5所示，加入residual防止梯度爆炸，促使模型收敛，并在2个 1×1 卷积之间添加BN层和激活函数以保持feature map的特征多样性，同时降低复杂度。

2.2 三维空间注意力机制

在特征提取时，采用三维空间注意力机制，同时关注feature map的空间域和通道域注意力，在形态、大小、位置多样复杂的肿瘤中迅速找到需要重点关注的特征。传统的注意力模块大多只关注通道注意

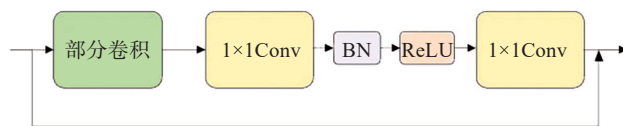


图5 部分卷积使用优化
Figure 5 Optimization with partial convolution

力或者空间注意力，如Hu等^[21]提出的SENet和Jaderberg等^[22]提出的STN网络，它们只在feature map的一维或者二维空间进行注意力标注，如图6所示，这使得特征的细化局限于通道或者空间维度，限制权重随通道和空间变化的灵活性。为了将通道注意力和二维空间注意力机制结合，Woo等^[23]提出CBAM，它包含通道注意力模块和空间注意力模块两部分。通道注意力模块通过全局池化操作获取每个通道的重要性，并通过全连接层来学习通道间的关系，从而生成通道注意力图。空间注意力模块则通过通道维度上的平均池化和最大池化操作来捕捉不同空间位置的重要性，并结合通道注意力图生成最终的空间注意力图。

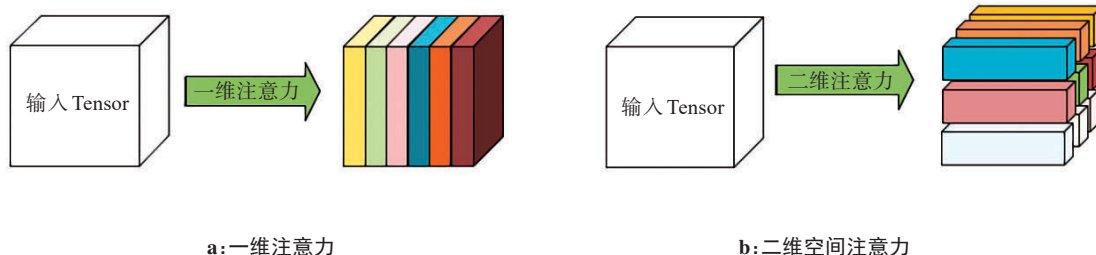


图6 一维注意力和二维空间注意力
Figure 6 One-dimensional attention and two-dimensional spatial attention

CBAM虽然融合通道注意力和空间注意力，但这种方法不能直接生成真正的三维权重，并且这样的结构需要经过两步计算，这会导致计算时间过长。两个模块的相对独立性也导致该方法在通道和空间同时变化的区域缺乏灵活性。为了同时关注通道和空间的变化注意力并降低结构复杂度，提出在不增加额外参数数量的前提下对每个神经元进行三维注意力权重赋值。神经学空间抑制原理^[24]表明信息最丰富的神经元通常是那些对周围神经元表现出独特放电模式的神经元，并且活跃的神经元还可能抑制周围神经元的活动。根据此原理，Yang等^[25]提出使用三维空间注意力机制SimAM，如图7所示，通过最小化每个神经元的能量函数，找到目标神经元和其他

神经元的线性可分离性，将高权重注意力赋值于表现出明显空间抑制效应的神经元。

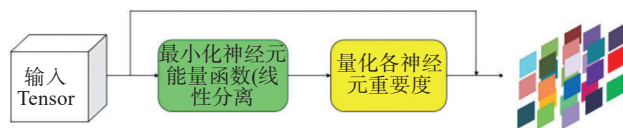


图7 三维空间注意力
Figure 7 Three-dimensional spatial attention

基于神经学和数学理论，定义每个神经元的能量函数为式(2)，以此估计每个神经元的重要性。其

中 $M = H \times W$ 是通道上神经元的个数, t 表示输入特征 $X \in R^{C \times H \times W}$ 单个通道中的目标神经元, x_i 表示输入特征 $X \in R^{C \times H \times W}$ 单个通道中的其他神经元, i 是空间维度索引。 \hat{t} 和 \hat{x}_i 是 t 和 x_i 的线性变换如式(3)所示, w_i 和 b_i 是变换的权重和偏置。对 y_o 和 y_i 采用二进制标签, 分别为 -1 和 1。可以通过最小化式(2)找到同一通道中 t 和 x_i 之间的线性可分离性: 当 $\hat{t} = y_i$, 所有 $\hat{x}_i = y_o$ 时, 对应目标神经元的能量 e_i 取到最小值。

$$e_i(w_i, b_i, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \quad (2)$$

$$\hat{t} = w_i t + b_i, \hat{x}_i = w_i x_i + b_i \quad (3)$$

为了防止过拟合, 对式(2)添加正则化表达式得到式(4), 并通过 w_i 和 b_i 的快速解析式(5)对式(4)进行简化。其中 μ_i 和 σ_i^2 的表达式为式(6), 它们为同一通道上除目标神经元外所有神经元的均值和方差。而由于式(5)的计算是在单个通道上进行的, 可以假设同通道的神经元是独立同分布的。由此可以对所有神经元进行均值 $\hat{\mu}$ 和方差 $\hat{\sigma}^2$ 的计算得到式(7):

$$e_i(w_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} \left(-1 - (w_i x_i + b_i) \right)^2 + \left(1 - (w_i t + b_i) \right)^2 + \lambda w_i^2 \quad (4)$$

$$w_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda} \quad (5)$$

$$b_i = -\frac{1}{2} (t + \mu_i) w_i$$

$$\mu_i = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i \quad (6)$$

$$\sigma_i^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - \mu_i)^2$$

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i \quad (7)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$$

对上述公式化简可以得到神经元能量的最简形式 e'_i 如式(8)所示。由于每个神经元与其他神经元的线性分离性与该神经元能量负相关, 因此认为每个神经元的重要性为 $1/e'_i$ 。由此给出三维空间注意力机制单个神经元注意力表达式(9), 其中 X_{in} 为输入特征, X_{out} 为输出特征。通过式(9)可以快速准确得到三维空间注意力进而在复杂的脑肿瘤影像中提取到更具重要性的特征, 从而提升模型整体检测的精确度。

$$e'_i = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (8)$$

$$X_{out} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X_{in} \quad (9)$$

2.3 动态注意力BBR损失函数

在目标检测任务中, BBR损失函数对最后的结

果准确率影响极大。理想的损失函数应当满足: 如图8所示锚框和实际目标框趋于重叠时, 梯度的幅度收敛为零。传统的YOLO是基于二者交并比IoU来量化它们的重叠程度^[22], 并以 $L_{IoU} = 1 - \text{IoU}$ 为BBR损失。但是在两个边界框没有重叠时, 这种方法会导致IoU恒为0, 从而造成损失函数反向梯度消失, 无法进行回归任务。后来DIoU^[26]、CIoU^[26]、SIoU^[27]等方法在损失函数中引入与边界框box相关的几何因子并通过在原损失函数的基础上添加惩罚项 R_i 来解决损失函数反向梯度消失的问题如式(10)所示:

$$L_i = L_{IoU} + R_i \quad (10)$$

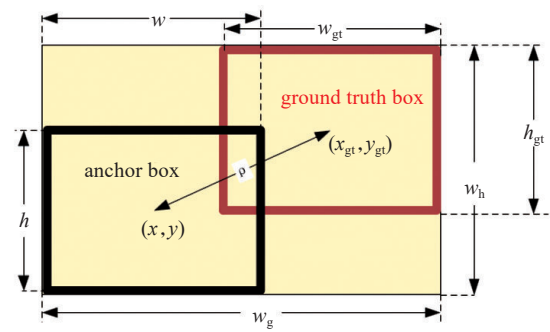


图8 锚框和实际目标框

Figure 8 Anchor box and ground truth box

DIoU是一种基于距离的IoU扩展, 考虑检测框之间的中心点距离, 更好地捕捉目标检测框之间的空间关系。DIoU相比传统的IoU能够更准确地度量两个检测框的重叠程度。CIoU是对DIoU的改进, 除了考虑中心点距离外, 还考虑宽高比和长宽比的差异, 可以更准确地描述目标检测框之间的重叠情况, 在处理不同形状和尺寸的目标时表现更好。SIoU是一种复杂的边界框回归方法, 通过融合角度考虑和规模敏感性, 解决传统损失函数的限制, 能更全面地考虑目标的位置、角度和形状等因素, 从而在预测准确性方面取得显著的提升。但是在边界框回归任务中, 由于目标的稀疏性, 存在着训练实例不平衡的问题。前面几种方法引入的几何因素例如距离和高宽比, 会加重对低质量数据的惩罚, 产生过大的梯度, 对网络的训练过程具有害处。为了不影响网络的整体训练结果, 解决训练实例不平衡的问题, 应当给予普通质量的例子较大的关注度并对其分配较大梯度, 而对离群度高的异常值例子以及离群度小的高质量例子分配小梯度增益。Lin等^[28]提出Focal-EIoU来解决这个问题, 指出anchor box的梯度增益随 L_{IoU} 变化的曲线, 其梯度变化趋势应与图9相似, 并给出 $L_{Focal-IoU}$ 如式(13)所示:

$$R_{\text{EIoU}} = \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2} + \frac{(x - x_{gt})^2}{W_g^2} + \frac{(y - y_{gt})^2}{H_g^2} \quad (11)$$

$$L_{\text{EIoU}} = L_{\text{IoU}} + R_{\text{EIoU}} \quad (12)$$

$$L_{\text{Focal-IoU}} = \text{IoU}^\gamma L_{\text{EIoU}} \quad (13)$$

其中, γ 是控制离群值抑制程度的参数; L_{EIoU} 包含3个部分: IoU 损失、距离损失和高宽损失。这种方法虽然能调整对不同质量例子的关注度, 但它是静态的关注, 忽略了 anchor box 质量的变化, 即锚框的质量分布是在相互比较中动态改变的。

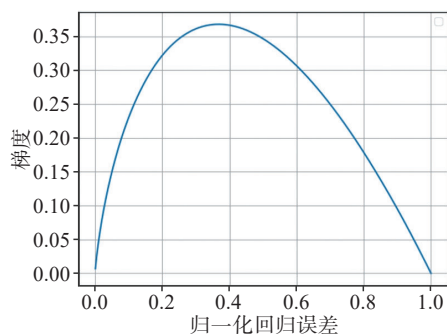


图9 $L_{\text{Focal-IoU}}$ 梯度分布趋势

Figure 9 Trend of gradient distribution for $L_{\text{Focal-IoU}}$

为了量化 anchor box 变化的质量分布, 并对不同质量的 anchor box 给予不同的梯度, 提出使用动态注意力 $\text{WIoU}^{[29]}$ 使模型专注于普通质量的 anchor box。为了降低距离和长宽比等几何因素对低质量数据的回归惩罚, 并且在 anchor box 与目标框重合良好时削弱几何因素的惩罚, 选择构建惩罚项 R_{WIoU} 如式(14)所示, 以及 R_{WIoU} 与 L_{IoU} 相乘的距离注意力机制损失函数 L_{WIoU} 如式(15)所示。在 R_{WIoU} 中, W_g 和 H_g 是最小封闭边界的宽和高, (x, y) 和 (x_{gt}, y_{gt}) 分别为 anchor box 与目标框的中心坐标, 如图8所示, 为了加快模型的整体收敛速度, 降低几何因素对低质量例子的惩罚, 这里没有引入宽高比等新的度量。在 L_{WIoU} 中, $L_{\text{IoU}} \in [0, 1]$, 可以降低高质量 anchor box 的 R_{WIoU} , 减少对其关注度; $R_{\text{WIoU}} \in [1, e)$, 可以显著放大普通质量 anchor box 的 L_{IoU} , 提高对其关注度。为了动态表示 anchor box 的离群度, 构建 $\beta = L_{\text{IoU}} / \overline{L_{\text{IoU}}}$, 并引入梯度增益参数 r 如式(17)所示, 其中 anchor box 的离群度-梯度增益变化曲线如图10所示。对于离群度高的低质量 anchor box 和离群度小的高质量 anchor box, 为其分配较小的梯度增益, 而对处于中间水平的普通质量 anchor box 分配较大的梯度增益, 以此来把边界框回归的工作重心集中在普通质量 anchor box 上, 最后将

r 与 L_{WIoU} 相乘得到 L_{WIoU} 如式(16)所示。 L_{WIoU} 在对形态多样、结构复杂的脑肿瘤检测时能动态地分配注意力, 降低低质量数据对结果的影响, 有效提升检测结果的准确度。

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)}\right) \quad (14)$$

$$L_{\text{WIoU}} = R_{\text{WIoU}} L_{\text{IoU}} \quad (15)$$

$$L_{\text{WIoU}} = r L_{\text{WIoU}} \quad (16)$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (17)$$

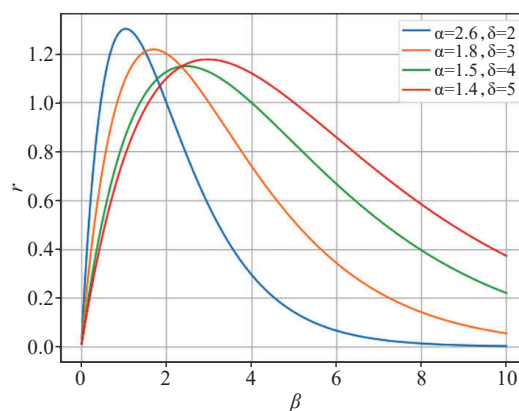


图10 L_{EIoU} 梯度增益分布

Figure 10 Gradient gain distribution of L_{EIoU}

3 实验与分析

实验基于 Pytorch 的深度学习框架和 12.4 版本的 CUDA。GPU 使用显存容量为 24 G 的 NVIDIA RTX3090。实验各参数的初始化设置如下: 学习率为 0.01, 迭代次数为 300, 权重衰减为 0.000 5, 动量因子为 0.8。

3.1 数据集

为了在检测模型速度与精度的同时检验模型的鲁棒性, 实验的数据集选用两个公开数据集: 来自 Insure Chain 的 Brain_Tumor 和来自 Roboflow 的 Glioma_of_test。Brain_Tumor 含有 2 161 张分辨率为 640×640 像素的脑部核磁共振图像, 可分为“胶质瘤”、“脑膜瘤”、“垂体瘤”和“无肿瘤”4类, 对应的每类图像数量分别为 492、483、587、599 张。Glioma of test 含有 2 338 张分辨率为 640×640 像素的脑部核磁共振图像, 类别和 Brain_Tumor 相同, 对应每类图像数量分别为 760、555、405、618 张。为了在 Pytorch 框架下训练以及验证改进后的模型, 对标签信息以 txt 格式标注, 并将 Brain_Tumor 划分为含有 1 878 张影像的训练集和 188 张影像的验证集, 将 Glioma_of_

test 划分为含有 1 392 张影像的训练集和 478 张影像的验证集。

3.2 评价指标

本文主要从精度和速度两个方面来评价模型检测的效率,精度表现主要以各类别平均精度 AP 的均值(mAP)来衡量,而速度则以每秒钟处理的图像数(FPS)来衡量。mAP 表达式如式(18)所示, n 表示类别的总数, AP_i 表示第 i 个类别的平均精度,它取决于不同阈值下的精确率(Precision)和召回率(Recall),如式(19)所示。Precision 表示模型预测为正例的样本中有多少是真正的正例,是模型预测的边界框中真正包含物体的比例,如式(20)所示。其中,TP(True Positives)为正确预测为正例的样本数量,FP(False Positives)为错误预测为正例的样本数量。Recall 表示模型能够正确检测出的正例样本的比例,是模型成功检测到真实物体的比例,如式(21)所示。FN(False Negatives)表示未能正确预测为正例的样本数量。mAP 综合考虑模型在不同类别上的检测性能,并给出模型在整个数据集上的综合表现,它可以综合评估模型在目标检测任务中的整体性能,而不仅是关注单个类别的性能表现。推理速度指标衡量模型在给定硬件上的处理速度,通常以每秒处理的图像数 FPS 来衡量。在实际应用中,更高的 FPS 意味着模型能够更快地处理图像,并实现实时目标检测。改进后各类脑瘤的混淆矩阵如图 11 所示。

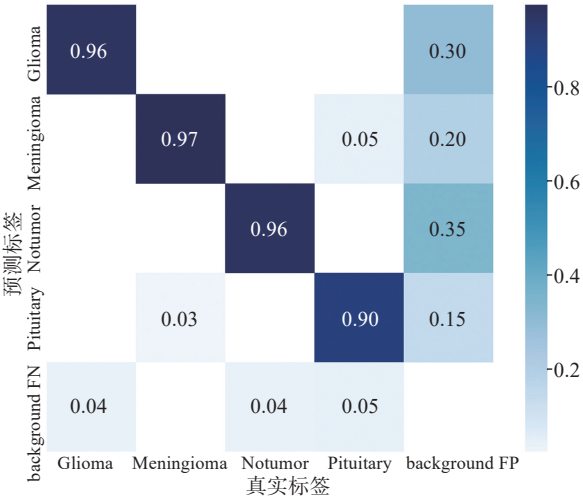


图 11 改进后各类脑瘤的混淆矩阵

Figure 11 Confusion matrix for various types of brain tumors after improvement

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i$$
 (18)

$$AP = \int_0^1 P(r)dr$$
 (19)

$$Precision = \frac{TP}{TP + FP}$$
 (20)

$$Recall = \frac{TP}{TP + FN}$$
 (21)

3.3 总体实验

经过在数据集上验证,得到本文模型与其他主流模型的 mAP 值和 FPS 值,如表 1 所示。本文的模型在 Brain_Tumor 上的 mAP 值和 FPS 值分别为 96.9%、162.7 帧/s;在 Glioma_of_test 上的 mAP 值和 FPS 值分别为 92.8%、158.1 帧/s,在脑肿瘤检测的精度和速度上较其他模型都有一定的提升。

表 1 本文模型与其他模型对比实验结果
Table 1 Comparative experimental results of the proposed model with other models

数据集	实验模型	mAP/%	FPS/ 帧·s ⁻¹
Brain_Tumor	Faster-RCNN	95.8	148.6
	YOLOv5	93.5	142.4
	YOLOv7	95.5	156.3
	YOLOv8	95.9	158.6
	改进模型	96.9	162.7
Glioma_of_test	Faster-RCNN	89.3	135.4
	YOLOv5	88.7	132.6
	YOLOv7	90.4	139.9
	YOLOv8	90.8	144.6
	改进模型	92.8	158.1

从图 12 随机选取的原图以及检测结果图中可以更为直观地看出检测后的效果。其中,图 12 第 1 行为各类型肿瘤原图,第 2 行为 YOLOv7 原模型的检测结果,第 3 行为改进后模型的检测结果,检测结果图中目标框第一行的数字为检测置信度。图 12 前 4 列显示改进后的模型能以较高的置信度检测出胶质瘤、脑膜瘤、垂体瘤以及无肿瘤(整个脑部区域)。同时图 12 后两列显示,改进后的模型能排除错误干扰脑膜瘤、胶质瘤并检测出垂体瘤结果。图 13 显示改进后模型的 mAP 平均检测精度明显高于原模型。综上所述,改进模型在脑肿瘤检测的精确度和速度方面都有明显的提升,具有综合优势。

3.4 消融实验

为了进一步体现各个改进点对整体模型效率提升的有效性,分别只对原 YOLOv7 模型添加单个改进点 PConv、SimAM、Wiou 以及 3 个改进点中任意两个的组合,并在 Brain_Tumor 上进行实验,结果如表 2 所示。从表 2 可知,每种改进方式都有不同程度的提升。添加 PConv 可以显著提升 FPS 指标,对模型的速

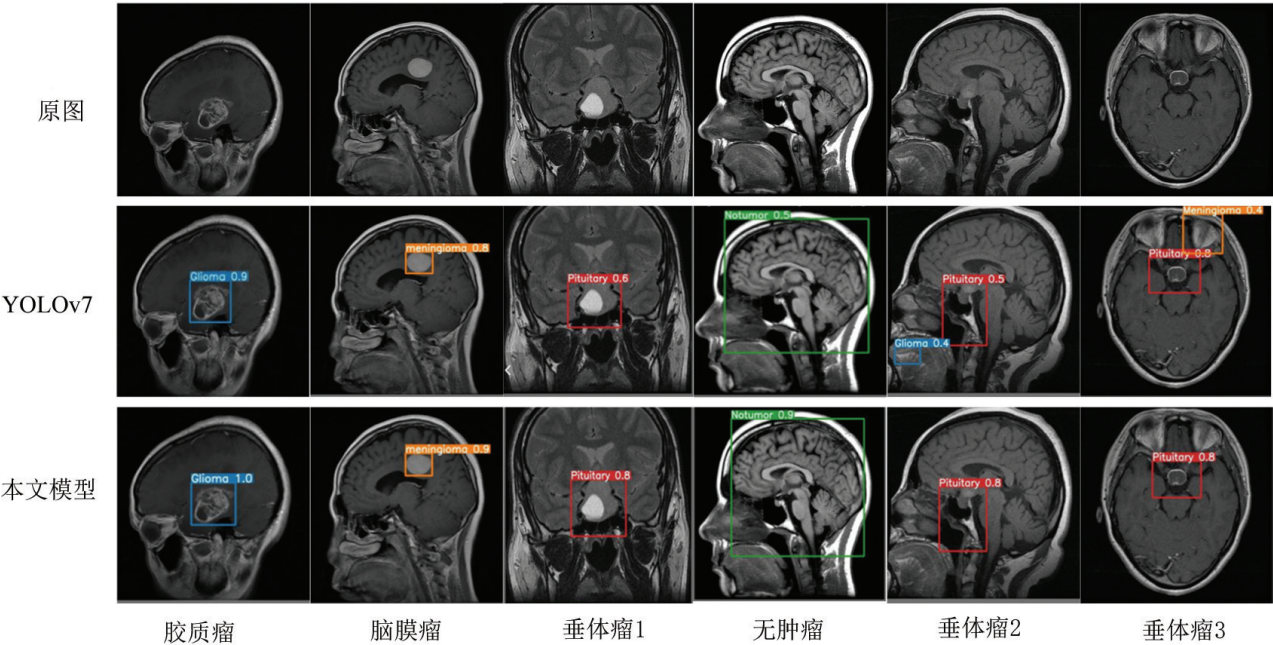


图 12 不同模型对各类脑肿瘤的检测效果

Figure 12 Detection performances of different models for various types of brain tumors

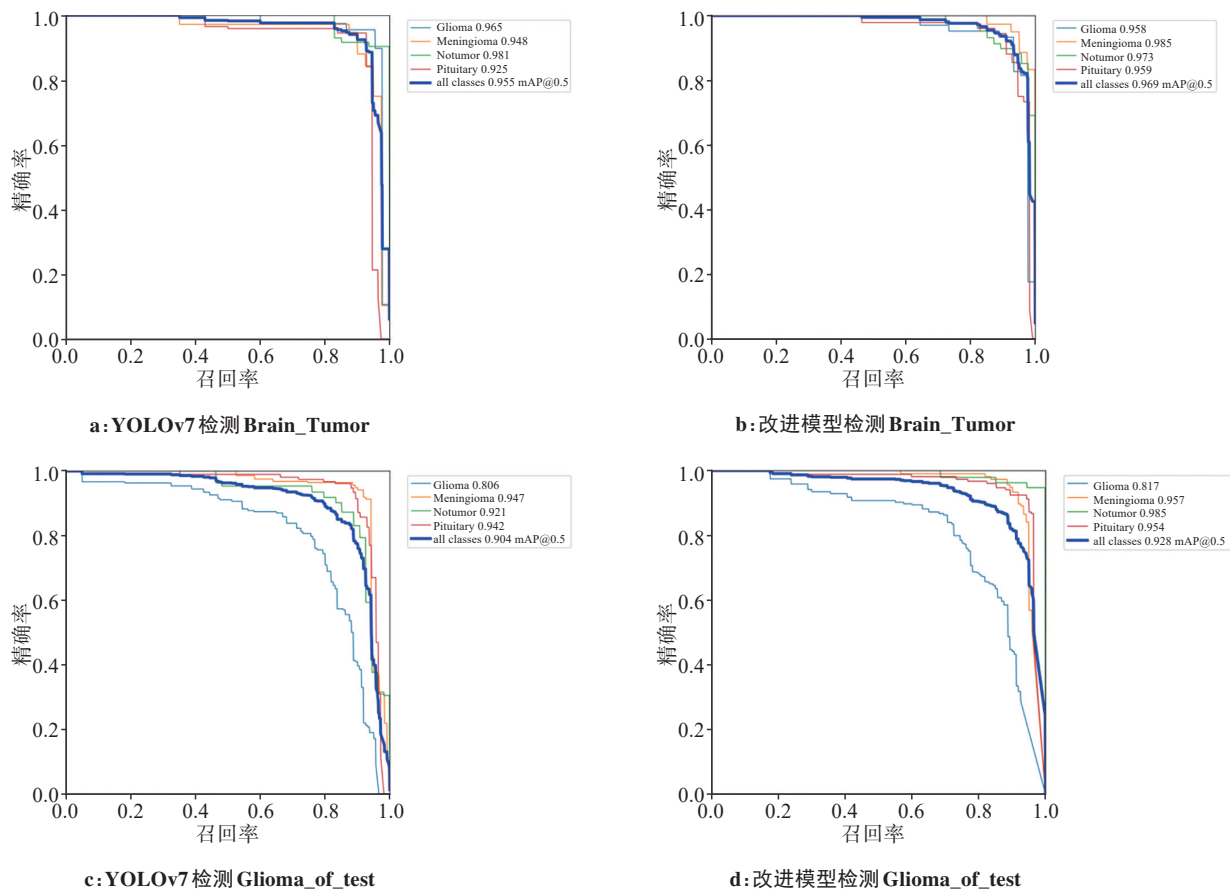


图 13 精确率-召回率曲线

Figure 13 Precision-recall curves

度有较大提升。在 mAP 指标方面,添加 SimAM 和 WIoU 的模型表现较好。同时将 PConv 与其他改进点添加到原模型时,模型的检测速度和精度都有所提升,这体现了每个改进点的有效性。

表 2 各改进点对模型整体性能的影响
Table 2 Effect of each improvement on the overall performance of the model

实验模型	PConv	SimAM	WIoU	mAP/%	精确率/%	召回率/%	FPS/ 帧·s ⁻¹
YOLOv7	×	×	×	95.5	92.6	91.1	156.3
改进模型 1	√	×	×	95.6	92.8	92.6	172.2
改进模型 2	×	√	×	96.4	91.3	91.7	157.1
改进模型 3	×	×	√	96.5	95.4	93.6	149.7
改进模型 4	√	√	×	96.7	92.7	91.2	169.3
改进模型 5	√	×	√	96.5	92.7	93.5	167.6
改进模型 6	×	√	√	96.7	92.5	93.3	143.3
改进模型	√	√	√	96.9	93.4	93.7	162.7

4 结 论

本文针对脑部 MRI 影像肿瘤检测的需求,在 YOLOv7 模型的基础上进行多项改进。首先,针对脑肿瘤的相似性,提出在特征提取时使用 PConv,以此来消除冗余计算,提高检测的速度。其次,针对脑肿瘤的复杂多样性,引入 SimAM 三维空间注意力机制到特征提取过程中,以增强对重要特征的关注。最后,针对脑肿瘤数据的不平衡问题,在 BBR 阶段使用 WIoU 来提升对普通质量 anchor box 的关注。通过实验发现,对比原 YOLOv7 模型,改进后模型在两个公开数据集上的 mAP 值和 FPS 值均有所提升,这证明本文的改进方法能有效提高脑肿瘤检测的精度和速度。通过消融实验发现,在同时加入 PConv、SimAM 和 WIoU 改进时,虽然模型检测准确度和速度整体提高,但速度却比只加入 PConv 要低,下一步将在提高准确率的基础上进一步减少改进融合带来的速度影响,优化和改进脑肿瘤影像检测的算法和技术。

【参考文献】

[1] Chong AW, McAdory LE, Low DC, et al. Primary intraventricular tumors-Imaging characteristics, post-treatment changes and relapses[J]. Clin Imaging, 2022, 82: 38-52.

[2] Kalbkhani H, Shayesteh MG, Zali-Vargahan B. Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series[J]. Biomed Signal Process Control, 2013, 8(6): 909-919.

[3] Machhale K, Nandpuru HB, Kapur V, et al. MRI brain cancer classification using hybrid classifier (SVM-KNN) [C]//2015 International Conference on Industrial Instrumentation and Control (ICIC). Piscataway, NJ, USA: IEEE, 2015: 60-65.

[4] Parveen, Singh A. Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM[C]//2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN). Piscataway, NJ, USA: IEEE, 2015: 98-102.

[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014

IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 580-587.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[8] Abd-Ellah MK, Awad AI, Khalaf AA, et al. Two-phase multi-model automatic brain tumour diagnosis system from magnetic resonance images using convolutional neural networks[J]. EURASIP J Image Video Process, 2018, 2018(1): 97.

[9] Noreen N, Palaniappan S, Qayyum A, et al. A deep learning model based on concatenation approach for the diagnosis of brain tumor [J]. IEEE Access, 2020, 8: 55135-55144.

[10] 米吾尔依提·海拉提, 热娜古丽·艾合麦提尼亚孜, 卡迪力亚·库尔班, 等. 基于改进YOLOv7的肝囊型包虫病超声图像小病灶检测[J]. 中国医学物理学杂志, 2024, 41(3): 299-308.

Hailati M, Aihemaitiniyazi R, Kuerban K, et al. Small lesion detection in ultrasound images of hepatic cystic echinococcosis based on improved YOLOv7[J]. Chinese Journal of Medical Physics, 2024, 41(3): 299-308.

[11] 刘涌涛, 王宝珠, 郭志涛. 基于改进YOLOv7网络模型的肺结节检测算法[J]. 中国医学物理学杂志, 2023, 40(12): 1509-1517.

Liu YT, Wang BZ, Guo ZT. Lung nodule detection algorithm using improved YOLOv7 network model[J]. Chinese Journal of Medical Physics, 2023, 40(12): 1509-1517.

[12] Wang CY, Bochkovskiy A, Liao HY. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 7464-7475.

[13] Chen JR, Kao SH, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 12021-12031.

[14] Paria B, Yeh CK, Yen IE, et al. Minimizing FLOPs to learn efficient sparse representations[C]//ICLR 2020. New York, USA: ICLR, 2019: 1-20.

[15] Dolbeau R. Theoretical peak FLOPS per instruction set: a tutorial [J]. J Supercomput, 2018, 74(3): 1341-1377.

[16] Howard AG, Zhu ML, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17). <https://arxiv.org/abs/1704.04861>.

[17] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[J]. Commun ACM, 2017, 60(6): 84-90.

- [18] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [19] Ma NN, Zhang XY, Zheng HT, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 122-138.
- [20] Han K, Wang YH, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 1577-1586.
- [21] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [22] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2015: 2017-2025.
- [23] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [24] Zhao CG, Kong YJ, Li DW, et al. Suppression of distracting inputs by visual-spatial cues is driven by anticipatory alpha activity[J]. PLoS Biol, 2023, 21(3): e3002014.
- [25] Yang LX, Zhang RY, Li LD, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2021: 11863-11874.
- [26] Zheng ZH, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2020: 12993-13000.
- [27] Gevorgyan Z. SIOU loss: more powerful learning for bounding box regression[EB/OL]. (2022-05-25). <https://arxiv.org/abs/2205.12740>.
- [28] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2999-3007.
- [29] Tong ZJ, Chen YH, Xu ZW, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-04-08). <https://arxiv.org/abs/2301.10051>.

(编辑:陈丽霞)