计算机科学
COMPUTER SCIENCE

基于多分支注意力和深度下采样的医疗图像目标检测方法

顾成杰，孟义，朱东郡，张俊军

引用本文

顾成杰，孟义，朱东郡，张俊军. 基于多分支注意力和深度下采样的医疗图像目标检测方法[J]. 计算机科学，2025，52(11)：196-205.

GU Chengjie, MENG Yi, ZHU Dongjun, ZHANG Junjun. Medical Image Target Detection Method Based on Multi-branch Attention and Deep Down- sampling [J]. Computer Science, 2025, 52(11): 196-205.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240900088

基于多分支注意力和深度下采样的医疗图像目标检测方法

顾成杰[1]　孟　义[2]　朱东郡[1]　张俊军[1]
1 安徽理工大学公共安全与应急管理学院　合肥 231100
2 安徽理工大学计算机科学与工程学院　安徽 淮南 232000
(cjgu@aust.edu.cn)

摘　要　人工智能技术的发展,使得基于深度学习的医疗图像检测在临床实践中具有广泛的应用前景。然而,针对一些如肿瘤、斑块等医疗图像的目标检测,存在待标面积小、目标可提取特征少、提取难度大等问题。针对上述问题,提出了一种基于多分支注意力和深度下采样的医疗图像目标检测方法(MD-Det)。该方法引入特征提取模块(C2f-DWR),对多尺度特征进行提取,增强目标的特征表示。为了能够更有效地捕捉图像中的上下文信息,增强特征的提取能力,设计了一种深度下采样模块(D-down),其核心思想是通过融合多种采样方式,结合平均池化和最大池化的操作,充分利用它们各自的优势来提高特征提取的效果。在保持计算效率的同时,提高了目标检测精度。随后,提出了一种多分支注意力机制(Multi-branch Attention,MA),用于提取和加权融合不同维度的特征,每个分支提取输入张量的不同维度特征,包括空间和通道特征。通过生成注意力权重,强调重要特征并进行加权融合,增强了网络的特征提取能力,提升了模型的检测性能。最后,提出了一种新的联合优化策略,将*Wise-IoU*损失和*NWD*损失进行加权,形成一个联合回归损失函数,进一步提高了目标识别的准确率。实验表明,所提方法可以有效提高医疗图像目标的检测精度,在医疗数据集 Tumor 和 Liver 上的 $mAP_{0.5}$ 相较于基准模型 YOLOv8n,分别提高了 2.5 个百分点和 1.1 个百分点。

关键词:可变形卷积;目标检测;YOLO;注意力机制
中图分类号　TP391

Medical Image Target Detection Method Based on Multi-branch Attention and Deep Down-sampling

GU Chengjie[1],MENG Yi[2],ZHU Dongjun[1] and ZHANG Junjun[1]
1 School of Public Safety and Emergency Management,Anhui University of Science and Technology,Hefei 231100,China
2 School of Computer Science and Engineering,Anhui University of Science and Technology,Huainan,Anhui 232000,China

Abstract　With the development of artificial intelligence technology,medical image detection based on deep learning has a wide application prospect in clinical practice. However,for some medical image target detection such as tumor and plaque,there are some problems,such as small area to be labeled,few features to be extracted and difficult to extract. To solve these problems,this paper proposes a medical image target detection method(MD-Det) based on multi-branch attention and deep subsampling. The feature extraction module(C2f-DWR) is introduced to extract multi-scale features and enhance the feature representation of the target. This paper designes a deep down-sampling module(D-down) to capture the context information in the image more effectively and enhance the feature extraction capability. The core idea is to combine average pooling and maximum pooling operations to make full use of their respective advantages to improve the feature extraction effect by fusing multiple sampling methods. The accuracy of target detection is improved while maintaining the computational efficiency. Then,a multi-branch attention(MA) mechanism is proposed,which extracts and weights features of different dimensions,with each branch extracting features of different dimensions of the input tensor,including spatial and channel features. By generating attention weights,important features

到稿日期:2024-09-13　返修日期:2024-12-19
基金项目:国家自然科学基金重大科研仪器研制项目(52227901);科技部国家重点研发计划(长三角科技创新共同体联合攻关专项)(2023CSJGG1103);安徽省高等学校科学研究项目(2023AH051197);安徽理工大学引进人才科研启动基金(2023yjrc33)
This work was supported by the National Natural Science Foundation of Special Fund for Research on National Major Research Instrument (52227901),National Key Research and Development Plan of the Ministry of Science and Technology (Yangtze River Delta Science and Technology Innovation Community Joint Research Special Fund)(2023CSJGG1103),Natural Science Research Project of Colleges and Universities in Anhui Province (2023AH051197) and Scientific Research Foundation for High-level Talents of Anhui University of Science and Technology (2023yjrc33).
通信作者:朱东郡(djzhu_aust@163.com)

are emphasized and weighted together. The feature extraction capability of the network is enhanced, and the detection performance of the model is improved. Finally, a new joint optimization strategy is proposed, which weights *Wise-IoU* loss and *NWD* loss to form a joint regression loss function to further improve the accuracy of target recognition. Experiments show that the proposed method can effectively improve the detection accuracy of the model in medical image targets, and the $mAP_{0.5}$ of the medical data sets Tumor and Liver are increased by 2.5 percentage points and 1.1 percentage points, respectively.

Keywords Deformable convolution, Target detection, YOLO, Attention mechanism

1 引言

随着计算机视觉技术的发展，目标检测在医疗图像临床诊断中的辅助作用日益突出。传统诊断中，放射科医生通过CT或MRI获取图像并观察病灶，由于每日审阅大量图像，医生容易产生视觉疲劳，从而增加了误诊和漏诊的概率。计算机辅助诊断技术可以自动检测和识别病变组织，为医生提供支持，有效减轻医生的负担并减少诊断错误。然而，尽管计算机辅助检测在医疗图像领域具有重要作用，但医疗图像目标检测依然存在一些挑战。例如，医疗图像中的目标通常是小目标(图像中像素占比低于32×32的区域就属于小目标[1])，它们通常难以观察且检测精度较低，这给计算机辅助检测带来了很大的困难。在医学领域，这种情况非常常见，如在早期肿瘤检测过程中，由于肿瘤在此时体积较小并且病变组织和周围组织差异不大，因此计算机很难区分出图像中的病变[2]。在肺结节筛查中，绝大部分结节的直径都很小，想要精确定位结节非常困难[3]。此外，临床获取的医学图像大多清晰度不高，尤其是在CT成像时，为减少造影剂对患者的影响，通常使用低剂量造影剂，这导致了图像质量较差且噪声较多，使得原本就复杂的医疗图像目标检测变得更加困难。与自然图像检测相比，医疗图像存在目标较小、清晰度低、画质差等问题。因此，现有的自然图像检测算法难以直接应用于医疗图像领域，需要结合其独特特征进行针对性的改进。卷积神经网络(Convolutional Neural Network，CNN)的出现，促进了目标检测领域的进步和发展。2012年，Krizhevsky等[4]提出更深层次、更多参数量的Alexnet CNN并取得成功。2014年，Simonyan等[5]提出了VGG(Visual Geometry Group)网络，使用小型池化层和卷积核替换Alexnet中的大型池化层和卷积层，通过加深网络层数来获取更好的拟合特征。2015年，He等[6]提出的残差网络(Residual Network，ResNet)，解决了随着网络层次加深模型出现性能下降的问题。此外，针对小目标的检测，主要采用基于深度学习的目标检测算法，在平衡检测精度的同时保持了较快的检测速度。但是，当前小目标检测仍面临着诸多挑战。例如，下采样过程中的小目标特征信息容易丢失，影响有效识别，可以采用多尺度特征提取网络，保留低层空间特征并结合高层语义信息，以增强小目标的表示能力。同时，小目标常常被复杂背景掩盖，容易与背景混淆，引入注意力机制(如自注意力或区域注意力)可以帮助模型更关注小目标周围的区域，降低背景干扰。此外，小目标在图像中缺乏上下文信息，影响模型的理解能力，可以通过局部上下文模块将小目标周围区域的特征与其自身特征相结合，并引入自注意力机制，使得模型能够在特征图上关注重要的上下文区域，进一步增强对小目标的理解。Song等[7]在浅层特征图中融入了深层语义信息，并提出中心点无锚框方法，使得网络通过中心点回归定位目标，从而对小目标的定位更加灵活，但是网络的精度有所损失。如今，YOLO系列算法已获得空前的发展，不仅检测速度快，对大、中型目标的检测效果也优于主流的双阶段目标检测算法。但是，这些算法在处理小目标时仍会遇到困难，会造成小目标的漏检和识别精度下降，进而导致模型在密集目标场景中的性能下降。这种性能下降影响了模型在复杂环境中的应用，尤其是在医疗图像这类需要高精度检测的任务中。针对上述问题，本文提出了一种改进YOLOv8n算法，通过改进特征提取网络，使网络能够提取更加丰富的特征信息，更好地适应小目标复杂多变的情形。本文的创新点与贡献如下：

1)构建了基于可变形卷积的C2f-DWR模块，该模块的引入可以增强对多尺度特征的提取，提高了小目标的检测精度及检测能力。

2)提出的D-down模块能够增加每个像素点的感受野，使得模型不仅能够捕捉到更广泛的上下文信息，还能提取到更高级别、更抽象的特征，提高模型检测效果的同时，使模型更加轻量化。

3)提出了多分支注意力机制(Multi-branch Attention，MA)，该模块提升了模型收敛速度以及精度，增强了模型对小目标的检测性能。

4)将*Wise-IoU*损失函数和*NWD*损失函数进行加权组合，形成了一个综合的回归损失函数。这进一步提高了模型对医疗图像目标的检测性能。

2 相关工作

2.1 目标检测

目标检测技术在深度学习的推动下得以迅速发展，使得目标检测算法被广泛应用于医学图像领域。当前，目标检测算法主要分为单阶段YOLO(You Only Look Once)和SSD以及双阶段Fast R-CNN[8]和Faster R-CNN[9]两类。YOLO系列算法由Redmond等[10]于2015年提出，是首个基于单阶段的深度学习目标检测网络。SSD[11]是一种基于深度学习的目标检测算法，结合了Faster R-CNN的锚机制和YOLO的端到端架构。Fast R-CNN通过将SPP层简化为单尺度的ROI Pooling层，统一了候选区域特征的大小。Lee等[12]将一种名为分组卷积的方法集成到SSD网络中，以提升肝脏病变检测的准确性。此改进显著提高了检测精度，取得了良好的效果。Albahli等[13]针对黑色素瘤检测，提出了网络设计的改进方案，选择YOLOv4[14]作为基线网络。由于原始结构在识别黑色素瘤区域时表现有限，因此他们对网络架构进行了调整，提升了其区分能力和检测精度。Zhang等[15]将多尺度特征提

取与多分辨率候选框策略相结合，并应用于网络样本训练中。实验结果显示，这一策略显著提升了对小型乳腺肿瘤区域的检测和定位精度。Li 等[16]设计了一种新颖的 3D 卷积神经网络用于肺结节检测，该网络的基础架构是由 19 层 CNN 组成的密集连接模块，无需使用区域建议网络。在分类器中，他们引入了 maxout 单元，以便有效处理肺结节的变化，提升了对小结节的召回率。Zhao 等[17]提出了一种结合多尺度特征融合的肺结节检测方法，在候选结节检测阶段，该方法将多尺度特征与基于深度学习的 Faster R-CNN 框架相结合，显著提高了对小尺寸肺结节的检测能力。Ding 等[18]提出了一种将反卷积层与 Faster R-CNN 网络相结合的方法，以提取纵轴切片图像中的候选区域。这一改进对于肺结节的自动检测非常有效，并显著提升了结节候选区域的分类和定位精度。Faster R-CNN 引入区域提议网络(RPN)，将目标检测任务整合为端到端的训练过程。RPN 生成候选区域后进行分类和定位，但需要训练两个不同的网络，导致时间较长，因此检测速度较慢。同时，小目标特征在其池化和卷积过程中易被忽略，从而造成检测效果不佳。Carion 等[19]提出的 Detection Transformer(DETR)方法，通过 Transformer[20]网络实现端到端的训练过程，基于编码器-解码器框架，并结合集合预测的直观方式来解决目标检测任务。DETR 的计算复杂度较高，尤其在小目标检测的实时应用中，可能无法满足性能需求。此外，DETR 使用全局自注意力机制，这可能导致下采样后部分小目标特征丢失，难以有效捕捉细节特征。尽管目标检测算法不断发展，但在医学领域，这类具有复杂场景以及小目标居多的图像检测仍存在挑战。

2.2 注意力机制

在计算机视觉中，注意力机制通过动态调整输入图像特征的权重，提升网络对重要区域的关注能力。Hu 等[21]首次提出了通道注意力概念并引入 SE-Net(Squeeze-and-Excitation Network)，通过自适应地调整通道间的权重，增强深度神经网络的表示能力。其核心是挤压和激励模块，用于收集全局信息，捕获通道关系，提高特征表达能力。为了降低 CNN 的计算成本并将计算资源集中于重要区域，Jaderberg 等[22]提出了空间注意力网络(Spatial Transformer Networks，STN)，通过可学习的空间转换模块对特征图进行空间变换，增强了表示能力。空间注意力网络利用注意力机制增强了图像特征表达，通过与原始特征图融合进一步提升了特征效果。通道注意力与空间注意力可串联或并联，从而形成混合注意力机制。Woo 等[23]提出的卷积注意力模块(CBAM)在通道和空间维度上进行注意力操作，CBAM 的引入不仅提升了模型的整体识别能力，还增强了对小目标、遮挡目标等难检测区域的关注。Zhao 等[24]提出的 Wasserstein 自我注意和定向交叉注意模块，使得基于 Transformer 的定向目标检测器能够更好地处理具有方向和旋转特性的目标，从而显著提升检测性能。Chen 等[25]提出了 NAM 注意力机制，通过 BN 缩放因子来计算注意力权重，有效抑制了不显著特征，提高了小目标检测的准确率。注意力机制的优势在于能重点关注和增强图像中的关键区域，从而提高小目标和被遮挡目标的检测能力。尽管注意力机制能显著提升目标检测的准确率，但是其计算开销相对比较大，这种额外的计算需求可能降低处理速度。

3 MD-Det

图 1 展示了本文所提出的 MD-Det 方法的网络结构图。MD-Det 引入特征提取模块(C2f-DWR)来替换骨干网络中的部分 C2f 模块，以提高模型对小目标物体的识别能力，并在骨干网络尾部增加多分支注意力机制 MA 以进一步提升模型的速度、精度以及泛化性能；同时，将部分 Conv 模块替换为 D-down 模块，增强网络对输入图像的整体感知能力，有助于捕获更丰富的上下文信息，还能缩小特征图尺寸，减少后续层的计算量，加快训练和推理速度，从而在不牺牲性能的情况下实现更高效的计算；然后，将 Wise-IoU 损失函数和 NWD 损失函数加权组合成一个新的综合回归损失函数，该损失函数有效地解决了小目标检测中的误差、类别不平衡和锚框匹配问题，显著提升了模型的检测精度和鲁棒性。

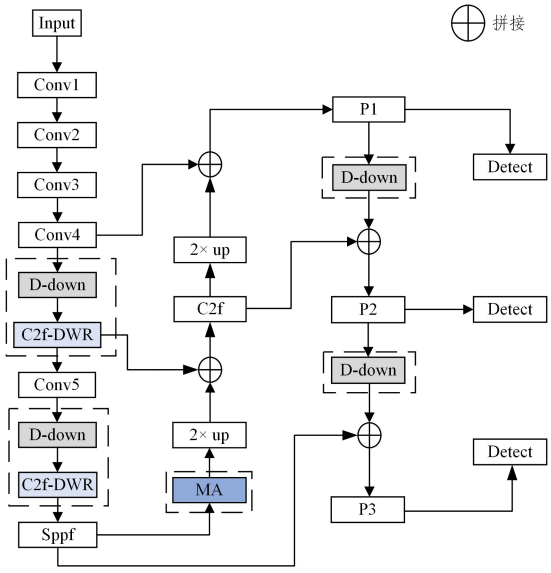

图 1 MD-Det 结构

Fig. 1 Structure of MD-Det

3.1 特征提取模块

由于目标的大小、形状、位置、方向等都具有一定的变化性，因此使用传统的卷积操作往往难以准确地获取目标所在位置，甚至可能会出现目标漏检或误检的问题。为了解决上述问题，引入了 C2f-DWR 模块。C2f-DWR 是在 YOLOv8 网络架构的基础上，将可扩张残差模块 DWR 加入主干网络的 C2f 层中形成的一个新的特征提取模块。DWR 是一个可扩张残差注意力模块，是一种高效的两步获取多尺度上下文信息方法的模块，在残差内部，采用两步法高效绘制多尺度上下文信息，并融合多尺度感受野生成特征图。第一步：从输入特征中生成相关残差特征，称为区域残差化。在这一步中，生成一批不同大小区域的简洁特征图，并将其作为第二步形态滤波的输入。第二步：分别采用多速率扩展深度卷积对不同大小的区域特征进行形态学滤波，该操作被称为语义残差化。每个通道特征只应用一个期望的接受域，以防止可能的冗余接受域。DWR 模块通过两个不同的步骤表现(区域残留和语义残留)获取多尺度上下文信息的独特方法使得特征提取过

程更加高效和集中。区域剩余化简化了各种大小的特征图，为随后的语义剩余化进行了优化。其中，基于语义的形态学过滤使用了专门选择的感受野来进行应用。这种方法不但大大简化了学习过程，还确保了相关特征的有序和有效提取，这对于准确的医疗诊断至关重要。同时，区域形式的特征映射使深度扩张卷积的作用简化为形态过滤，从而使学习过程有序。区域化特征映射通过空间约束与先验引导，使深度扩张卷积的空洞模式退化为可控的形态学滤波操作。这种结构化设计将复杂特征学习转化为多尺度形态规则的渐进式整合，从而实现了高效、稳定的学习过程。在绘制多尺度上下文后，对多个输出进行聚合。接下来，在特征图上执行 BN，采用逐点卷积对特征进行合并，通过 BN 消除多源特征间的分布差异，抑制噪声干扰；再利用逐点卷积进行特征融合，该操作通过压缩冗余信息聚焦关键区域，并形成最终残差。最后，将最终残差添加到输入特征映射，以构建更强、更全面的特征表示。

D-Conv 表示深度卷积，D-n 表示具有 n 的膨胀率的膨胀卷积。过大的感受野，如扩张率超过 7，可能会导致细节特征丢失，因为过大的感受野会导致目标的边缘和纹理特征被忽略。大感受野虽能引入更多的上下文信息，但过大的扩张率容易丢失部分目标特征，降低目标识别的准确性。因此，选择合适的扩张率对于保留细节特征至关重要。为此，选择了不同扩张率进行对比，结果如表 1 所列。由表 1 可知，将扩张率设置为 1，3，5 时检测效果较为合适。扩张率为 5 时，感受野较小，更能捕捉局部特征并保留目标的边缘和纹理特征。相比之下，扩张率为 7 时，感受野增大，虽然引入了更多的上下文信息，但可能忽略目标局部特征信息，使得最终目标特征判别性降低。因此，设置具有 1，3，5 的扩张率的 3 个扩张卷积分支。

表 1 扩张率结果对比

Table 1 Comparison of expansion rate results

扩张率	$mAP_{0.5}$
1，3，7	0.791
1，5，7	0.793
1，3，9	0.782
1，5，9	0.785
1，3，5	**0.802**

注：加粗数据为最优值。

此外，DWR 模块独特的区域重构和语义重构方法不仅有效地捕捉了多尺度的上下文信息，也保证了相关特征的有序高效提取。C2f-DWR 模块结构图如图 2 所示。

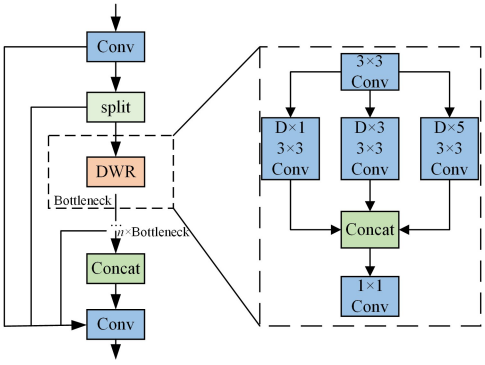

图 2 C2f-DWR 结构

Fig. 2 Structure of C2f-DWR

3.2 深度下采样模块

医疗图像通常具有高分辨率，处理时需要大量计算资源，导致计算量和内存消耗巨大。高分辨率图像中的局部特征可能包含大量冗余信息，难以有效表达全局结构和高级语义特征。此外，医疗图像中常见的噪声、伪影以及不规则背景干扰也会对目标检测造成挑战。因此，模型要能聚焦于重要特征，同时忽略次要细节。本文提出了一种引入多尺度特征融合策略的方法，该策略通过有效结合不同尺度的特征，融合高层语义特征与低层空间特征，能够更好地保留小目标特征，避免过度下采样导致的特征丢失，增强了对小目标的检测能力。首先，对输入进行平均池化，以减少参数量和计算量。随后，将 Split 层分为 3 部分：左右两部分进行平均池化和最大池化，再通过卷积提取特征；中间部分则直接进行卷积。通过这种方式，能在更高级别上提取全局特征，提高了模型的表达能力。最后，将 3 部分的结果拼接并输出，从而更好地融合多尺度信息，提升对复杂模式的捕捉能力。该模块通过减小特征图的空间尺寸，有效降低计算量和内存消耗，提高计算效率，使得在高分辨率图像上进行实时处理成为可能。同时，它能够提取高层次抽象特征，聚焦于重要信息，忽略次要细节，减少噪声影响并增强特征图的鲁棒性。此外，池化操作（如最大池化）可以保留显著特征，抑制噪声和无关信息，提高检测稳定性和特征表达能力。深度下采样模块结构如图 3 所示。

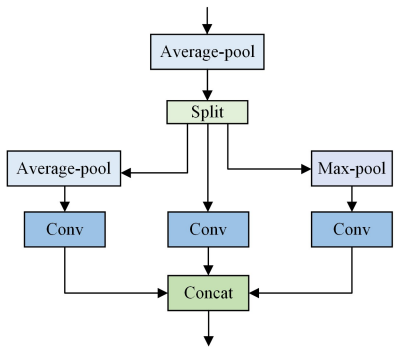

图 3 D-down 结构

Fig. 3 Structure of D-down

3.3 多分支注意力机制

医疗图像中常包含小目标，如肿瘤结节、血管结构等，这些小目标的检测对诊断和治疗至关重要。注意力机制能帮助目标检测模型聚焦于重要特征，从而提升小目标检测的精度。

然而，传统通道注意力方法存在缺陷：在计算权重时，输入张量经过全局平均池化，导致空间信息丢失，且通道与空间维度之间缺乏依赖性。尽管 CBAM 模型缓解了空间依赖问题，但通道和空间注意仍是独立计算的。本文提出了多分支注意力机制，它是一种轻量但有效的注意力机制，通过跨维度交互来捕捉空间和通道之间的关系。MA 由 3 个分支组成，每个分支负责捕获输入的空间维度和通道维度之间的交叉维度。给定形状为（$C\times H\times W$）的输入张量，每个分支负责聚合空间维度 H 或 W 与通道维度 C 之间的跨维度交互特征。在深度学习模型中，Transpose 操作通常用于改变张量的维度顺序，例如从（H，W，C）转置成（C，H，W）的形式，或者从（B，H，W，C）转置成（B，C，H，W）的形式等。然后，将张量传递到

Z 池，并将 3 维的维度缩减到 2 维，再将该维上的平均汇集特征和最大汇集特征连接起来，使得该层能够保留实际张量的丰富表示，同时缩小其深度，进一步减轻计算量。接着，将张量传递到内核大小为 $k \times k$ 的卷积层。注意力权重由 sigmoid 激活层生成并应用于置换的输入张量，之后将其置换回原始输入形状，最后需要经过排列变为 $C \times H \times W$ 维度特征。将前两个分支相加之后，利用 sigmoid 激活函数生成注意力权重。前两个分支和第三个分支分别提取不同维度的特征，并通过相加求平均的方式融合，减少了模型对单一注意力头的依赖，降低了过拟合风险。该机制通过多尺度特征融合显著增强了模型对特征的提取能力，有效减少了小目标的漏检和误检，特别是在复杂背景、多尺度目标和小目标检测中表现优越。图 4 为该注意力机制的结构图。

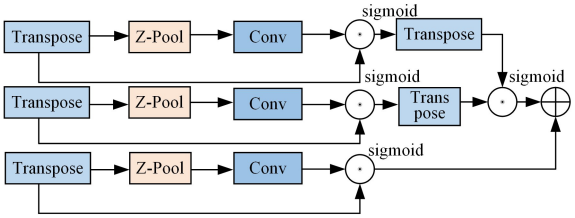

图 4 多分支注意力机制

Fig. 4 Multi-branch attention mechanism

3.4 联合优化策略

医疗图像通常目标过小，有时很难检测出病变部位，因此考虑改进损失函数。由于本文选用的数据集中小目标比较多，因此引入更适合小目标的 NWD 损失函数[26]来代替传统的交并比(IoU)损失函数。NWD 损失函数是用一个新的度量方法来计算框和框之间的相似度，其过程分为两个阶段。

1)边界框的高斯分布建模。常见的二维高斯分布的概率密度函数为：

$$f(x|\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{\exp\left(-\frac{1}{2}(x-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})\right)}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \tag{1}$$

其中，$\boldsymbol{\mu}=\begin{bmatrix}c_x\\c_y\end{bmatrix}$，$\boldsymbol{\Sigma}=\begin{bmatrix}\frac{w^2}{4} & 0\\0 & \frac{h^2}{4}\end{bmatrix}$，$(c_x,c_y)$为中心坐标，$w$ 为边界框宽度，h 为高度。x 为高斯分布的坐标(x,y)，$\boldsymbol{\mu}$ 为高斯分布的均值向量，T 为矩阵的转置，$\boldsymbol{\Sigma}$ 为高斯分布的协方差矩阵。

2)归一化高斯 Wasserstein 距离。对于两个边界框，$A=(cx_a,cy_a,w_a,h_a)$，$B=(cx_b,cy_b,w_b,h_b)$，建模的高斯分布 N_a 和 N_b 如下：

$$w_2^2(N_a,N_b)=\left\|\left(\left[cx_a,cy_a,\frac{w^a}{2},\frac{h^a}{2}\right]^{\mathrm{T}},\left[cx_b,cy_b,\frac{w^b}{2},\frac{h^b}{2}\right]^{\mathrm{T}}\right)\right\|_2^2 \tag{2}$$

由于 $w_2^2(N_a,N_b)$ 是距离度量，不能直接用作相似度度量，因此使用它的指数形式归一化作为相似度度量，如下：

$$NWD(N_a,N_b)=\exp\left(-\frac{\sqrt{w_2^2(N_a,N_b)}}{c}\right) \tag{3}$$

其中，c 是与数据集相关的常数，$NWD(N_a,N_b)$为变量间的正态分布水平距离。

针对小目标检测问题，本文引入了一种基于归一化 Wasserstein 距离(NWD)的回归损失函数，该损失函数可以有效地衡量两个边界框之间的相似性，而不受目标尺度的影响。与传统的 IoU 损失函数相比，NWD 损失函数在小目标检测中表现出更好的鲁棒性和优化效果，因为它可以处理边界框之间的不重叠或微重叠情况以及边界框位置的微小偏差。引入 NWD 损失函数会导致模型收敛速度变慢，在综合考虑模型的精确性和实时性之后，将改进 $Wise\text{-}IoU$ 损失函数和 NWD 损失函数进行加权组合，形成一个综合的回归损失函数，并通过实验确定其比例。YOLOv7 中采用 $CIoU$[27-28]计算定位损失，计算式如下：

$$L_{\mathrm{CIoU}}=1-IoU+\frac{\rho^2(b^A,b^B)}{c^2}+\alpha v \tag{4}$$

其中，IoU 为交并比，是 anchorbox 和 box 的重叠度；b^B 和 b^A 分别为真实框和预测框的中心点；c 为真实框与预测框的最小外接矩形的对角线长度；ρ 为两点间的欧氏距离；v 是预测框与目标框宽高比的一致性；α 为平衡参数。由式(4)可知，v 不能反映真实的宽高和置信度的差异，降低了模型优化相似性问题的效率，且 $CIoU$ 为单调聚焦机制，要求数据集为高质量示例，以提高拟合边界框损失的能力。但当数据集中有低质量示例存在时，若将重点放在边界框对低质量示例的回归上，会降低模型的检测性能。因此，本文采用 $Wise\text{-}IoU$ 作为边界框损失函数。同时，使用 L_{IoU} 作为 BBR 损失，它以比例的形式有效地屏蔽了包围框大小的干扰，增强了对小目标的学习能力。L_{IoU} 的定义如下：

$$L_{\mathrm{IoU}}=1-IoU=1-\frac{w^iH^i}{s_u} \tag{5}$$

由式(5)可知，当边框之间没有重叠时，$\alpha L_{\mathrm{IoU}}/\alpha_{w_i}=0$，使得 L_{IoU} 由反向传播的梯度消失，故将边框回归损失改为以下范式：

$$L_i=L_{\mathrm{IoU}}+R_i \tag{6}$$

定义动态聚焦机制为 $f(\beta)$，其中 $\beta=L_{\mathrm{IoU}}/\overline{L_{\mathrm{IoU}}}$，通过将小梯度增益分配到具有小 β 的高质量锚框，并且把小梯度分配给低质量锚，降低了低质量样本对边界框回归的危害。$Wise\text{-}IoU$ 和 NWD 的比例情况如表 2 所列。

表 2 比例结果

Table 2 Proportional results

比例($Wise\text{-}IoU:NWD$)	$mAP_{0.5}$
1:9	0.799
3:7	0.797
5:5	0.797
7:3	**0.801**

注：加粗数据为最优值。

最终确定损失函数如下：

$$Loss_{\mathrm{loc}}=IoU_{\mathrm{loss}}\times 0.7+NWD_{\mathrm{loss}}\times 0.3 \tag{7}$$

其中，$Loss_{\mathrm{loc}}$ 为最终确定的损失函数，IoU_{loss} 为 $Wise\text{-}IoU$ 损失函数，NWD_{loss} 为 NWD 损失函数。

4 实验与结果分析

4.1 数据集与实验参数设置

实验数据集从 roboflow 上获取，roboflow 是一个大型的 AI 图像识别以及模型训练数据标注平台。Tumor 数据

集包括脑瘤图像 9 900 张，包含 3 类目标。对图像进行网络训练，按照 8∶1∶1 的比例随机划分训练集、验证集和测试集，其中训练集数据 7 920 张，验证集数据 990 张，测试集数据 990 张。

另外一个数据集 Liver 也从 roboflow 上获取，该数据集包括肝脏图像 9 271 张，包含 3 类目标。对图像进行网络训练，按照 8∶1∶1 的比例随机划分训练集、验证集和测试集，其中训练集数据 7 416 张，验证集数据 927 张，测试集数据 928 张。为提高数据集的多样性，对每张图片进行了多次标注，每种方式都包含了不同的角度、尺度、变换，以模拟真实世界中的视觉场景。实验采用操作系统 Windows10，CPU 为 Intel Xeon Silver 4214R，GPU 为 NVDIA GeForce RTX3080Ti，12GB 显存；深度学习框架为 PyTorch 1.13.1，CUDA 11.1。对每一种算法，统一设置 epochs 为 200。

4.2 评价指标

本文所使用的模型评价指标包括召回率（Recall，R）、精确率（Precision，P）、平均精度均值（Mean Average Precision，mAP）、每秒传输帧数（Frames Per Second，FPS），以及模型参数量（Params）等。其中，精确率是指正确检测目标的比例，召回率表示数据集中实际检测出的目标的比例，mAP 的值是由精确率和召回率计算得到的，即：

$$P=TP/(TP+FP)*100\% \tag{8}$$

$$R=TP/(TP+FN)*100\% \tag{9}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{10}$$

$$mAP=\sum_{i=1}^{n}AP(i)*100\% \tag{11}$$

$$F_1\text{-}Score=(2*P*R)/(P+R) \tag{12}$$

其中，TP（True Positive）为正例且代表预测正确，FP（False Positive）为正例但表示预测错误，FN（False Negative）为负例且表示预测错误。

4.3 对比实验

采用了 11 种不同的模型进行目标检测的对比实验，其中包括 Faster R-CNN，DETR，Tood，Atss，Ddq，Dino，YOLOv5n，YOLOv6n，YOLOv8n，YOLOv9c，以及本文提出的 MD-Det 方法。通过 $mAP_{0.5}$，$mAP_{0.5:0.95}$，P，R，FPS，Params 这 6 个指标来评估模型性能。为了验证 MD-Det 的优越性和泛化性，在两个数据集都进行了对比实验。

Tumor 数据集上的结果如表 3 所列。结果表明，本文算法相较于 YOLOv5n，P 和 R 分别提升了 4.1 个百分点和 4.9 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 4.5 个百分点和 5.7 个百分点，参数量 $Params$ 降低了 2%，表明本文模型更加轻量；与 YOLOv6n 相比，P 和 R 分别提升了 5.3 个百分点和 4.4 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 4.6 个百分点和 4.8 个百分点，参数量 $Params$ 降低了 40.9%，说明本文模型在轻量化方面更加具有优势；与 YOLOv9c 相比，P 和 R 分别提升了 2.3 个百分点和 1.6 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 1.4 个百分点和 0.3 个百分点，参数量 $Params$ 降低了 80.6%。尽管模型的其他性能指标均有所提升，但 FPS 略有下降。这表明优化策略有效增强了模型性能，但也引入了额外的计算开销，导致检测速度有所减缓。然而，这种权衡在医疗图像这种对高精度要求极高的应用中依然是可接受的。因此，本文模型在显著提升性能的同时，成功保持了轻量化，兼顾了检测精度和计算效率。

表 3 Tumor 数据集上的实验结果

Table 3 Experimental results on Tumor dataset

方法	P	R	$mAP_{0.5}$	mAP_S	mAP_M	mAP_L	$mAP_{0.5:0.95}$	FPS	$Params$
FasterRcnn[9]	0.719	0.583	0.712	0.219	0.509	0.721	0.397	72.0	4.13×10^7
DETR[19]	0.730	0.575	0.721	0.227	0.519	0.726	0.436	75.9	4.15×10^7
Tood[29]	0.743	0.549	0.723	0.235	0.527	0.733	0.424	19.1	3.20×10^7
Ddq[30]	0.791	0.667	0.774	0.293	0.586	0.780	0.473	17.0	4.84×10^7
Atss[31]	0.701	0.542	0.696	0.204	0.495	0.697	0.389	20.6	3.21×10^7
Dino[32]	0.789	0.623	0.782	0.286	0.576	0.785	0.472	20.2	4.75×10^7
YOLOv5n	0.871	0.692	0.772	0.298	0.568	0.774	0.482	178.4	5.00×10^7
YOLOv6n	0.859	0.697	0.771	0.296	0.582	0.788	0.491	**215.2**	8.30×10^7
YOLOv8n	0.891	0.703	0.792	0.301	0.591	0.802	0.516	194.6	6.00×10^7
YOLOv9c	0.889	0.725	0.803	0.318	0.614	0.813	0.536	179.6	2.53×10^7
MD-Det	**0.912**	**0.741**	**0.817**	**0.327**	**0.623**	**0.824**	**0.539**	139.8	$\mathbf{4.90\times10^7}$

注：加粗数据为最优值。

Liver 数据集上的结果如表 4 所列。Dino 相比 MD-Det，尽管召回率 R 有所提升，但其精确度 P 却显著下降，导致误检增加。此外，Dino 的参数量显著增大，造成计算效率和推理速度明显下降。相比 Dino，本文模型在精确度 P 方面实现了 6.8 个百分点的显著提升，表明本文模型正确识别目标的能力增强；同时，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提高了 3.0 个百分点和 1.5 个百分点，表明本文方法在精确度 P 和召回率 R 之间达成了良好的平衡；参数量也下降了 89.6%，表明本文模型在轻量化方面优势明显。因此，本文方法在综合性能上超过了 Dino 模型。YOLOv5n 相比 MD-Det，尽管精确度 P 有所提升，但其召回率 R 却下降明显，导致漏检增多。相比 YOLOv5n，本文方法在召回率 R 方面实现了 5.9 个百分点的显著提升，表明本文模型能检测到更多的真实目标实例；同时，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升 1.8 个百分点和 1.3 个百分点，说明本文方法在 P 和 R 之间取得了较好的平衡；参数量下降了 2%，表明本文模型更加轻量。所以，本文方法的综合性能要优于 YOLOv5n。MD-Det 与 YOLOv6n 相比，P 和 R 分别提升了 1.2 个百分点和 1.6 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 0.8 个百分点和 1.0 个百分点，参数量 $Params$ 降低了 40.9%，说明本文模型在轻量化方面依然优势明显；与 YOLOv9c 相比，P 和 R 分别提升了 3.0 个百分点和 0.3 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 0.4 个百分点和 0.2 个百分点，参数量 $Params$ 降低了80.6%，表明本文模型依然保持了很好的轻量化设计。本文模型

的 *FPS* 相较于 YOLOv5n，YOLOv6n 和 YOLOv9c 略有下降，说明优化策略有效提升了模型的综合性能，但也增加了计算开销，导致检测速度减慢。但经过对模型综合性能的验证，这是可接受的。本文模型在提升性能的同时，成功保持了轻量化设计。

由表 3 和表 4 的结果可以看出，本文方法 MD-Det 不仅在包含大小病灶时的检测效果优于其他对比方法，而且在小病灶上的检测效果也优于其他对比方法。

表 4 Liver 数据集上的实验结果

Table 4 Experimental results on Liver dataset

方法	*P*	*R*	$mAP_{0.5}$	mAP_S	mAP_M	mAP_L	$mAP_{0.5:0.95}$	*Fps*	Params
FasterRcnn[9]	0.621	0.463	0.608	0.152	0.426	0.610	0.357	76.1	4.13×10^7
DETR[19]	0.636	0.478	0.629	0.172	0.434	0.631	0.384	77.1	4.15×10^7
Tood[29]	0.653	0.538	0.642	0.181	0.453	0.652	0.370	37.4	3.20×10^7
Ddq[30]	0.721	0.683	0.714	0.224	0.521	0.719	0.413	17.6	4.84×10^7
Atss[31]	0.655	0.535	0.653	0.185	0.467	0.668	0.372	48.9	3.21×10^7
Dino[32]	0.695	**0.684**	0.689	0.229	0.528	0.737	0.406	23.6	4.75×10^7
YOLOv5n	**0.798**	0.595	0.701	0.207	0.505	0.704	0.408	171.8	5.00×10^7
YOLOv6n	0.751	0.638	0.711	0.223	0.513	0.717	0.411	**220.7**	8.30×10^7
YOLOv8n	0.752	0.631	0.708	0.215	0.509	0.715	0.414	196.2	6.00×10^7
YOLOv9c	0.733	0.651	0.715	0.226	0.520	0.734	0.419	181.3	2.53×10^7
MD-Det	0.763	0.654	**0.719**	**0.234**	**0.531**	**0.740**	**0.421**	144.9	$\mathbf{4.90\times10^7}$

注：加粗数据为最优值。

4.4 消融实验

为验证本文提出的多分支注意力机制（Multi-branch Attention，MA）、深度下采样模块（D-down）、特征提取模块（C2f-DWR）以及引入综合损失函数的有效性，以 YOLOv8n 为基准模型，在相同实验条件下进行消融实验，实验结果如表 5 所列。在 YOLOv8n 的基础上，逐步添加改进模块，并根据模型评价标准比较各模块的优化效果。加入 MA 后，$mAP_{0.5}$ 提高了 0.2 个百分点，$mAP_{0.5:0.95}$ 提升了 0.3 个百分点，召回率 *R* 增加了 1 个百分点，但是精确率 *P* 下降了 0.9 个百分点。尽管精确率有所下降，但模型的整体性能得到了提升。MA 增强了对特征图空间信息的学习，突出了与目标相关的区域，使检测模型更智能地选择和聚焦重要特征。D-down 模块则部分替换了骨干网络中的卷积层。D-down 模块通过减少参数量来降低模型的复杂度，进而提高模型的运行效率。在 backbone 中，D-down 可用于在特征图的不同层之间进行下采样，而在 head 部分，它可以帮助进一步细化特征图的分辨率，用于更精确的目标检测。该模块使得模型的参数量下降约 13%，虽然精确率下降 0.7 个百分点，但召回率提高 2.4 个百分点，模型的 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别上升 0.9 个百分点和 1 个百分点。D-down 使得模型性能有较大提升的情况下，降低了模型的复杂度。C2f-DWR 模块替换骨干网络中的部分 C2f 层之后，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 1 个百分点和 1.5 个百分点，精确率虽下降了 0.4 个百分点，但召回率 *R* 提升了 2.1 个百分点，模型参数量也有所下降。修改损失函数之后，模型的 *P* 和 *R* 分别提升了 0.8 个百分点和 1.6 个百分点，$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 1.1 个百分点和 0.6 个百分点，也实现了模型性能的综合提升。最后，将上述模块同时组合到一起，其模型精确度 *P* 提高 2.1 个百分点，召回率 *R* 提高 3.8 个百分点，$mAP_{0.5}$ 提高 2.5 个百分点，$mAP_{0.5:0.95}$ 提升 2.3 个百分点，模型的参数量也下降 18%。实验结果表明，与 YOLOv8n 相比，本文算法成功实现了模型轻量化与高性能的平衡。本文模型在最小化参数量的同时显著提升了效率，相比其他模型展现出较高的性能均衡性和实用价值。

表 5 消融实验对比

Table 5 Comparison of ablation experiments

方法	MA	D-down	C2f-DWR	Wiou+NWD	*P*	*R*	$mAP_{0.5}$	$mAP_{0.5:0.95}$	*Params*
基准模型+					0.891	0.703	0.792	0.516	6.0×10^6
	√				0.882	0.713	0.794	0.519	6.0×10^6
		√			0.884	0.727	0.801	0.526	5.2×10^6
			√		0.887	0.724	0.802	0.531	5.9×10^6
				√	0.899	0.719	0.803	0.522	6.0×10^6
	√	√	√	√	**0.912**	**0.741**	**0.817**	**0.539**	$\mathbf{4.9\times10^6}$

注：加粗数据为最优值。

4.5 数据分析

为了更好地证明本文方法的优越性，从 F1，PR 和热力图等几个方面再次进行论证。F1 分数作为模型精确率和召回率的调和平均，综合考虑了这两者的权衡，是用于衡量二分类模型精确率的重要指标。YOLOv8n 的 F1 分数曲线如图 5 所示。YOLOv9c 的 F1 分数曲线如图 6 所示。MD-Det 的 F1 分数曲线如图 7 所示。可以明显看出，YOLOv9c 和 MD-Det 相较于基准模型 YOLOv8n，在各个种类均取得较高的 F1 分数，MD-Det 表现更为明显。

热力图是一种用于展示数据空间分布的图形方式，每个单元格通过颜色深浅来表示数据值的大小。在热力图中，每个数据点或区域都被分配一个颜色值，该值反映了该位置的数据密集程度或数值大小。通常，高数值或高密集区域使用暖色调表示（如红色），低数值或低密集区域则使用冷色调表示（如蓝色），而中间数值则使用黄色或绿色等颜色来表示。在目标检测过程中，热力图能够直观地展示模型对输入的关注区域，高置信度区域通常以较亮的颜色显示，而低置信度区域则较暗。图 8 展示了本文模型与基线模型的热力图对比。

可以看出，本文模型在检测中的表现优于基线模型：背景检测覆盖较少且亮度较暗；而在目标区域，热力图颜色较亮且面积较大，表明模型对目标的检测置信度较高。这验证了本文模型在小目标检测中表现优异，并显示出其在复杂场景下能有效减少背景干扰，聚焦关键特征。

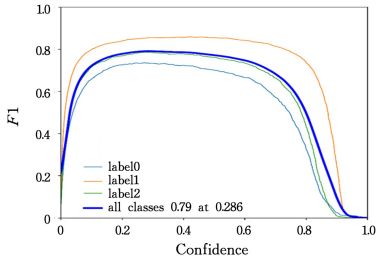

图 5　YOLOv8n 算法 F1 分数曲线

Fig. 5　F1 fraction curve of YOLOv8n algorithm

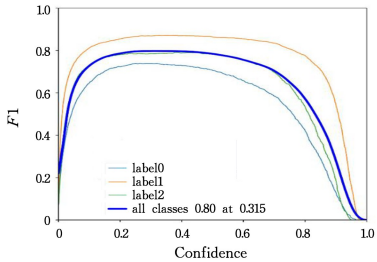

图 6　YOLOv9c 算法 F1 分数曲线

Fig. 6　F1 fraction curve of YOLOv9c algorithm

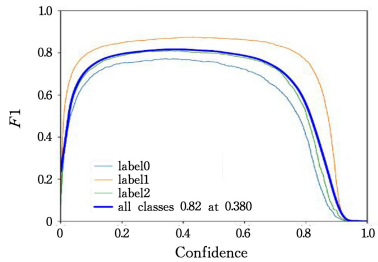

图 7　MD-Det 算法 F1 分数曲线

Fig. 7　F1 fraction curve of MD-Det algorithm

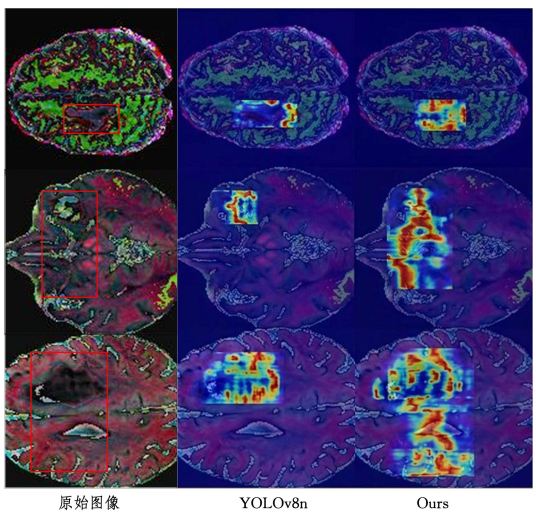

图 8　热力图对比(电子版为彩图)

Fig. 8　Thermal map comparison

精确率-召回率曲线(Precision-Recall Curve，P-R 曲线)作为模型重要的评估指标之一，用于评估分类模型的性能，展示了模型在不同阈值下的精确率(纵轴)与召回率(横轴)之间的权衡关系。曲线的形状可以揭示模型的不同特征：曲线越接近右上角，说明模型在高精度和高召回率之间取得了更好的平衡。YOLOv8n 的 P-R 曲线如图 9 所示。YOLOv9c 的 P-R 曲线如图 10 所示。MD-Det 的 P-R 曲线如图 11 所示。从整体可以看出，MD-Det 的 P-R 曲线优于 YOLOv8n 的 P-R 曲线。

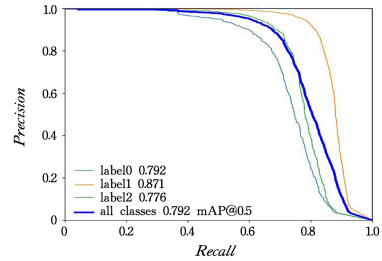

图 9　YOLOv8n 算法的 P-R 曲线

Fig. 9　P-R curve of YOLOv8n algorithm

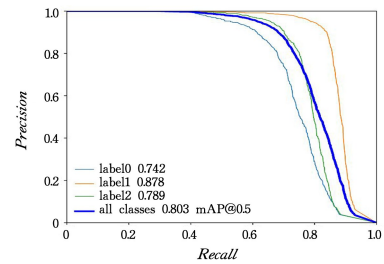

图 10　YOLOv9c 算法的 P-R 曲线

Fig. 10　P-R curve of YOLOv9c algorithm

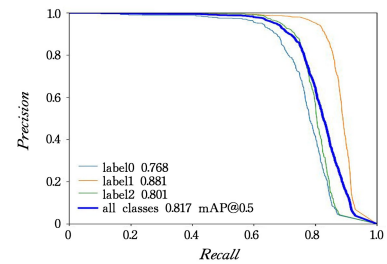

图 11　MD-Det 算法的 P-R 曲线

Fig. 11　P-R curve of MD-Det algorithm

4.6　结果分析

本文模型与基线模型在医疗场景下的目标检测结果如图 12 所示，其中涵盖了不同尺度的脑瘤目标。本文算法在各尺寸目标的检测中能够提取更精确的特征信息，展现出更高的检测精度，明显优于基线模型，对同一目标的置信度更高。在医疗图像分析中，高置信度的诊断结果可直接用于临床决策，而低置信度的结果则可进行复查或者由专家审阅，为模型改进提供了依据，提升了决策的可靠性和准确性。通过比较基线模型和本文方法的检测结果可以看出，本文方法在处理多尺度和复杂背景的医疗图像时，更能聚焦于关键病变区域，有效减少漏检和误检，为提升医疗决策的可靠性和精度提供了坚实的数据支持。

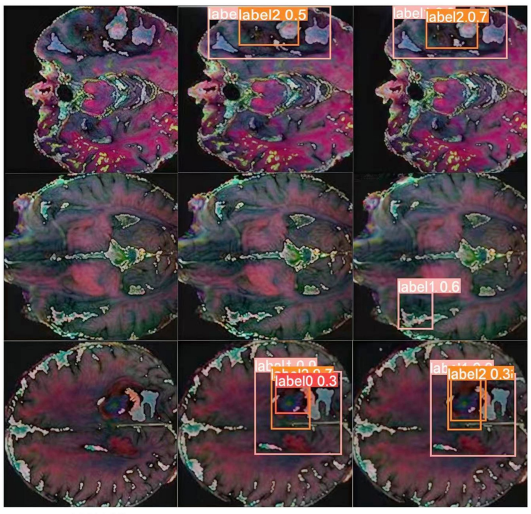

图 12 Tumor 上的检测结果

Fig. 12 Test results on Tumor

YOLOv8n 模型在小目标和图像质量较差的情况下可能会出现漏检和误检，而本文算法有效改善了这一问题。与其他检测算法相比，模型的综合性能也有了显著提升。

结束语 根据实际应用中医疗图像检测存在目标小造成的误检、漏检，以及图像质量不好所造成的目标难以检测等问题，本文提出了基于改进 YOLOv8n 的医疗图像目标检测算法 MD-Det。改进之后的 MD-Det 算法的总体精度不仅显著提高，对于误检、漏检等问题也起到了明显的缓解作用。MD-Det 相对于基准网络 YOLOv8n 进行了以下改进：1)在网络中增加 C2f-DWR 模块，简化了学习过程，并且确保了相关特征更有序和更有效地被提取；2)将多分支注意力模块加到基准网络 backbone 的末尾，该注意力模块加快了收敛速度并提高了精度；3)将网络中的部分卷积更换为 D-down 模块，该设计可以在保持计算效率的同时，增强特征的提取能力，通过引入更深的网络层次和更复杂的连接方式来有效地捕捉图像中的上下文信息，极大地提高了对目标的识别和定位精度；4)引入 *NWD* 回归损失函数，将 *Wise-IoU* 损失函数和 *NWD* 损失函数进行加权组合。*NWD* 损失函数通过 Wasserstein 距离提供了一种更鲁棒和有效的方式来衡量分布之间的差异，这样能够改善模型的训练稳定性，提高生成样本的质量，并更好地捕捉细节和结构信息。

本文算法的检测精度仍有提升空间，可以考虑在算法中引入自蒸馏策略，但直接增加该策略会导致算法参数量急剧增大，从而影响算法性能。目前尚未找到一种较好的平衡方法。另外，还可以考虑对模型进行压缩，但大部分模型压缩方法都会对精度产生较大影响，目前尚未找到理想的压缩方案。这些问题仍待解决，在后续研究中将会重点加以关注。

参 考 文 献

[1] LINT Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European Conference on Computer Vision. 2014: 740-755.

[2] NAGARAJANM B, HUBER M B, SCHLOSSBAUER T, et al. Classification of small lesions in dynamic breast mri: eliminating the need for precise lesion segmentation through spatiotemporal analysis of contrast enhancement[J]. Machine Vision and Applications, 2013, 24(7): 1371-1381.

[3] SETIO A A A, TRAVERSO A, DE BEL T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge[J]. Medical Image Analysis, 2017, 42: 1-13.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409. 1556, 2014.

[6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[7] SONG L M, WANG S P, LI Y P, et al. A weld feature points detection method based on improved YOLO for welding robots in strong noise environment[J]. Signal, Image and Video Processing, 2022, 17(5): 1801-1809.

[8] GIRSHICK R. Fast R-CNN[C]// Proceedings of 2015 IEEE International Conference on Computer Vision. IEEE, 2015: 1440-1448.

[9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 91-99.

[10] 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). IEEE, 2016: 779-788.

[11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 6517-6525.

[12] LEE S, BAE J S, KIM H, et al. Liver lesion detection from weakly-labeled multi-phase CT volumes with a grouped single shot multibox detector [C] // Medical Image Computing and Computer Assisted Intervention. Springer, 2018: 693-701.

[13] ALBAHLI S, NIDA N, IRTAZA A, et al. Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour[J]. IEEE Access, 2020, 8: 198403-198414.

[14] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004. 10934, 2020.

[15] ZHANG Z, ZHANG X, LIN X, et al. Ultrasonic diagnosis of breast nodules using modified faster R-CNN[J]. Ultrasonic Imaging, 2019, 41(6): 353-367.

[16] LI F, HUANG H, WU Y, et al. Lung nodule detection with a 3d convnet via iou self-normalization and maxout unit[C] // ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 1214-1218.

[17] ZHAO Y, WANG Z, LIU X, et al. Pulmonary Nodule Detection

Based on Multiscale Feature Fusion[J]. Computational and Mathematical Methods in Medicine，2022，22(41)：1-13.

[18] DING J，LI A，HU Z，et al. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks[C]// Medical Image Computing and Computer Assisted Intervention. Springer，2017：559-567.

[19] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end object detection with transformers[C]// European Conference on Computer Vision. Cham：Springer，2020：213-229.

[20] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc.，2017：6000-6010.

[21] HU J，SHEN L，SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE，2018：7132-7141.

[22] JADERBERG M，SIMONYAN K，ZISSERMAN A，et al. Spatial transformernetworks [J]. arXiv：1506. 02025，2015.

[23] WOO S，PARK J，LEE J Y，et al. CBAM：convolutional block attention module[C]// European Conference on Computer Vision. Cham：Springer，2018：3-19.

[24] ZHAO J，DING Z，ZHOU Y，et al. Oriented Former：An End-to-End Transformer-Based Oriented Object Detector in Remote Sensing Images[C]// IEEE Transactions on Geoscience and Remote Sensing. IEEE，2024：1-1.

[25] CHEN Y Z，WANG S Q，ZHOU W，et al. small target detection of fish swarm based on SPD-Conv structure and NAM attention mechanism [J]. Computer Science，2018，51(S1)：438-444.

[26] ZHOU X，JIANG L，GUAN X J，et al. Infrared small target detection algorithm with complex background based on YOLO-NWD[C]// Proceedings of the 4th International Conference on Image Processing and Machine Vision. ACM，2022：6-12.

[27] YE L M，CHEN W W. A method for detecting cascaded insulator defects that combines semantic segmentation and object detection [J]. Computers and Modernization，2023(6)：82-88.

[28] JIANG R Q，YE Z C，PENG Y P，et al. Lightweight target detection algorithm for weak UAV targets[J]. Advances in Lasers and Optoelectronics，2022，59(8)：109-120.

[29] FENG C，ZHONG Y，GAO Y，et al. Tood：Task-aligned one-stage object detection[C]// 2021 IEEE/CVF International Conference on Computer Vision(ICCV). IEEE Computer Society，2021：3490-3499.

[30] ZHANG S，WANG X，WANG J，et al. Dense distinct query for end-to-end object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023：7329-7338.

[31] ZHANG S，CHI C，YAO Y，et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：9759-9768.

[32] ZHANG H，LI F，LIU S，et al. Dino：Detr with improved denoising anchor boxes for end-to-end object detection[J]. arXiv：2203. 03605，2022.

GU Chengjie，born in 1985，Ph.D，professor. His main research interests include trusted network architecture and security in cyberspace.

ZHU Dongjun，born in 1991，Ph.D，lecturer. His main research interests include deep learning and machine vision.

（责任编辑：柯颖）