

文章编号 1004-924X(2023)21-3178-14

## 强化特征提取能力的下颌骨骨折检测 3M-YOLOv5 网络

周 涛<sup>1,2</sup>, 杜玉虎<sup>1\*</sup>, 石道宗<sup>1</sup>, 彭彩月<sup>1</sup>, 陆惠玲<sup>3</sup>

(1. 北方民族大学 计算机科学与工程学院, 宁夏 银川 750021;

2. 北方民族大学 图像图形智能处理国家民委重点实验室, 宁夏 银川 750021;

3. 宁夏医科大学 医学信息与工程学院, 宁夏 银川 750004)

**摘要:**针对人工智能辅助骨折部位治疗时由于骨折部位通常伴随着出血等症状,不同体位所拍摄的 CT 影像存在较大差异,骨折部位大小不一,以及受到出血部位以及周围组织的干扰,骨折部位的特征提取不充分、骨折部位检测精度不高的问题,设计了一种 3M-YOLOv5 网络来检测下颌骨骨折部位。在特征提取网络中采用密集模块,利用密集连接特性提高网络的特征提取能力;采用局部全局注意力模块来提取 CT 影像的全局信息;构造一个轻量化的多尺度密集块,以较少的参数量提取骨折部位的多尺度特征;在特征增强网络中设计跨维度双向特征融合模块,使得特征图的高度、宽度以及通道之间有所交互,同时引入可训练的权重来平衡不同尺度特征图的融合重要性。为了验证 3M-YOLOv5 网络的有效性,在自建数据集上进行消融实验和对比实验。实验结果表明,在置信度阈值取 0.5 时,3M-YOLOv5 网络的 mAP 值、F1 值、召回率、精确率分别为 99.17%, 99.06%, 98.81% 和 99.32%。所提出的下颌骨骨折 CT 影像检测网络能够较好地检测出影像中的骨折部位,辅助医生制定治疗方案。

**关键词:**目标检测;下颌骨骨折;YOLOv5;跨维度注意力;密集连接神经网络

**中图分类号:**TP394.1;TH691.9 **文献标识码:**A **doi:**10.37188/OPE.20233121.3178

## Mandibular fracture detection with 3M-YOLOv5 network based on enhanced feature extraction capability

ZHOU Tao<sup>1,2</sup>, DU Yuhu<sup>1\*</sup>, SHI Daozong<sup>1</sup>, PENG Caiyue<sup>1</sup>, LU Huiling<sup>3</sup>

(1. College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China;

2. Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission,  
North Minzu University, Yinchuan 750021, China;

3. School of Medical Information & Engineering, Ningxia Medical University,  
Yinchuan 750004, China)

\* Corresponding author, E-mail:cy\_dyh@163.com

**Abstract:** For artificial intelligence assistance in the detection of fracture sites, the fracture sites are usually accompanied by bleeding and other symptoms. Further, CT images taken in different positions have large differences, the size of fracture sites varies, and the bleeding sites and surrounding tissues may inter-

收稿日期:2023-03-22;修订日期:2023-05-08.

基金项目:国家自然科学基金资助项目(No. 62062003);宁夏自然科学基金资助项目(No. 2022AAC03149)

fere with the detection of fracture sites, leading to insufficient feature extraction and the problem of low detection accuracy. Therefore, the 3M-YOLOv5 network is designed to detect mandibular fracture sites. First, the dense module is used in the feature extraction network to improve the feature extraction capability of the network by using the dense connection property. The local and global attention module (lgaM) is used to extract the global information of CT images. Second, a lightweight multiscale dense block (lmdM) is designed to extract the multiscale features of the fracture sites with fewer parameters. Third, a cross-dimension bidirectional feature fusion module (cdbfM) is designed in the feature enhancement network to make the height, width, and channel of the feature maps interact with each other, and trainable weights are introduced to balance the fusion importance of the feature maps with different scales. Finally, to verify the effectiveness of the proposed network, ablation and comparison experiments are conducted on a self-built dataset. The results show that when the confidence threshold is 0.5, the mAP value, F1 value, recall rate, and precision rate of the proposed network are 99.17%, 99.06%, 98.81%, and 99.32%, respectively. The proposed CT image detection network for mandibular fracture can better detect the fracture sites in the image than existing methods, which is a good reference for doctors to make a corresponding treatment plan based on the detection results.

**Key words:** target detection; mandibular fracture; YOLOv5; cross dimension attention; densely connected neural network

## 1 引 言

下颌骨由于特殊的生理结构,位于颌面部较为突出的位置,存在明显的解剖结构薄弱区域,在受到外力撞击时极易骨折。准确及时地确定下颌骨的骨折部位能够给予医生充足的参考信息,根据不同的骨折部位采取合适的治疗手段,避免术后下颌功能障碍、咬合关系错乱等问题,影响患者的生活质量<sup>[1]</sup>。当骨折程度较为严重时,骨折区域会出现骨头嵌插、压缩等情况,造成局部的密度增高,下颌中容易发生骨折的部位是正中联合部、颏孔区、下颌角部位以及髁突颈部,不同的骨折部位表现出来的症状不相同,在CT影像中的表现也不尽相同。在下颌骨骨折CT影像中,骨折部位的影像容易受到其余部位例如牙齿等部位的干扰,因为在CT影像中牙齿部位与骨骼部位均为白影,表示它们对X射线的吸收程度较高,而黑影表示对于X射线吸收程度较低的肌肉部位<sup>[2]</sup>。通常情况下,下颌骨的骨折伴随着或多或少的出血,出血部位对于X射线的吸收程度也较高。因此,由于出血情况以及其余部位的干扰,下颌骨骨折CT影像的人工阅片难度较大,通过人工智能的方法来辅助医生进行下颌骨骨

折部位的诊断具有重要的意义。

目前,人们开展了大量针对人体骨折部位检测的研究。Meng等<sup>[3]</sup>提出了一种用于肋骨骨折检测和分类的异构神经网络,由级联特征金字塔网络和分类网络组成,用来辅助放射科医生在CT图像上诊断和分类肋骨骨折。Zhou等<sup>[4]</sup>基于跨模态数据(临床信息和CT图像)进行肋骨骨折的自动检测和分类。应用基于快速区域的卷积神经网络(Fast Region Convolutional Neural Network, Faster R-CNN)整合CT图像和临床信息并使用结果合并技术将2D推断转换为3D损伤结果。Xue等<sup>[5]</sup>采用GA\_Faster R-CNN进行手部骨折的检测与定位,提出新的引导锚定方法使锚框生成更加准确和高效,大大提高了网络性能,并节约了计算量,同时采用平衡L1损失来适应学习任务的不平衡。Kitamura等<sup>[6]</sup>通过小样本、从头训练和多视图合并集成的卷积神经网络来检测踝关节骨折。Gao等<sup>[7]</sup>提出一种使用对侧、上下文和边缘增强模块的肋骨骨折诊断深度学习方法CCE-Net,使用多径融合机制作为主要架构来集成对侧、上下文和边缘增强模块所获得的特征信息。Liu等<sup>[8]</sup>采用Faster-RCNN辅助诊断股骨转子间骨折,同时对比了骨科主治医师与

人工智能的相关检测准确率、召回率等,指出人工智能诊断算法是一种有效的临床诊断方法,可作为骨科医师满意的临床助手。武等<sup>[9]</sup>基于胸部 X 光影像采用 YOLOv3 网络进行肋骨骨折的检测,证明了 CNN 可以提高肋骨骨折的检出率,有助于减少漏诊,避免医疗事故,减轻放射科医生的工作量。Warin 等<sup>[10]</sup>采用 Faster R-CNN 和 YOLOv5 来检测下颌骨骨折的 X 光影像,同时对比了这两种网络与口腔颌面外科医生在下颌骨骨折判断方面的准确率,检测网络实现了较好的下颌骨骨折分类性能。Wang 等采用 U-Net 和 ResNet 进行骨折 CT 影像的分类,首先利用 U-Net 对 CT 影像中的具体下颌骨部位进行分割,然后采用 ResNet 对分割后的部位进行具体分类,利用分类后的结果辅助医生对下颌骨骨折的诊断<sup>[11]</sup>。Vinayahalingam 等利用 Faster R-CNN 和 Swin-Transformer 网络对下颌骨部位口腔全景 X 光进行骨折部位的检测,利用 Swin-Transformer 作为主干特征提取网络,检测头部分仍采用 Faster R-CNN 的检测头,利用 Transformer 的自注意力机制进行较好的特征提取<sup>[12]</sup>。Son 等在口腔全景 X 光上利用 YOLOv4 网络进行下颌骨骨折的检测,对输入到网络中的 CT 影像进行单尺度亮度自适应变换和多尺度亮度自适应变换以增强图像的对比度<sup>[13]</sup>。

上述研究表明,采用人工智能的方法辅助骨折部位的检测具有较好的应用前景,能够给予医生充分的治疗参考信息。现有研究采用的检测网络多为通用网络结构,未针对于具体数据集进行特征提取网络的设计,对于骨折部位的关注程度不够充分。由于骨折部位大小不一、形状各异,同时受到出血以及其他未骨折部位的影响,当前相关骨折检测方法在进行影像特征提取时未考虑影像的全局特征表示,未提取相应的多尺度信息,不同尺度的特征图在进行特征融合时无法权衡其重要程度,特征图的通道维度、高度和宽度这三者之间未进行交互,缺乏必要信息的融合,存在检测精度不高的问题。

本文提出了一种下颌骨骨折检测网络 3M-YOLOv5,特征提取网络中采用密集模块进行改

进,引入密集连接思想,利用密集连接神经网络的特性增强网络的特征提取能力;在特征提取网络的开始阶段采用局部全局注意力模块(local and global attention Module, lgaM)来提取下颌骨 CT 影像的局部全局特征;在密集块结构中引入轻量化多尺度的思想,更好地提取 CT 影像的多尺度特征,关注到不同大小的骨折区域;设计了跨维度双向融合模块(cross dimension bidirectional fusion Module, cdbfM),利用跨维度注意力使得网络能够关注特征图的高、宽以及通道之间的相互关系,使得不同体位的 CT 影像特征互补,提升网络性能。

## 2 原 理

在提取 CT 影像特征时,通常利用卷积操作捕获局部信息,但是随着网络层数的加深,浅层网络的特征无法传递到深层网络中,导致深层网络无法获得边缘、形状和纹理等信息。密集连接神经网络通过构建当前层和后续其他层的密集连接,将浅层网络所提取的特征传递到深层网络中,能够实现信息的跨层传递<sup>[14]</sup>。因此,本文采用密集连接神经网络作为主要的特征提取网络,基于 YOLOv5\_s 网络提出了一种下颌骨骨折检测网络 3M-YOLOv5,其详细结构如图 1 所示。3M-YOLOv5 网络主要包括特征提取网络、cdbfM 模块以及 YOLO 检测头部分。特征提取网络主要包括 Stage 0 到 Stage 3 等 4 个阶段,其中 Stage 0 为预处理阶段,提取 CT 影像的局部全局特征,Stage 1 到 Stage 3 为重复堆叠的密集模块,分别重复堆叠 6, 12, 24 次。每个 Stage 的密集模块之后通过池化操作对特征图进行下采样,3 种不同尺度的特征图由此引出到 cdbfM 中进行特征增强。cdbfM 包括跨维度注意力模块(Cross Dimension Attention Module, CDA)以及双向特征融合模块(Bidirectional Feature Fusion Module, BFF)。特征增强后的特征图输入到 YOLO 检测头部分进行骨折部位的检测,3 个检测头从上至下分别负责检测较大的骨折区域、中等大小的骨折区域以及较小的骨折区域。

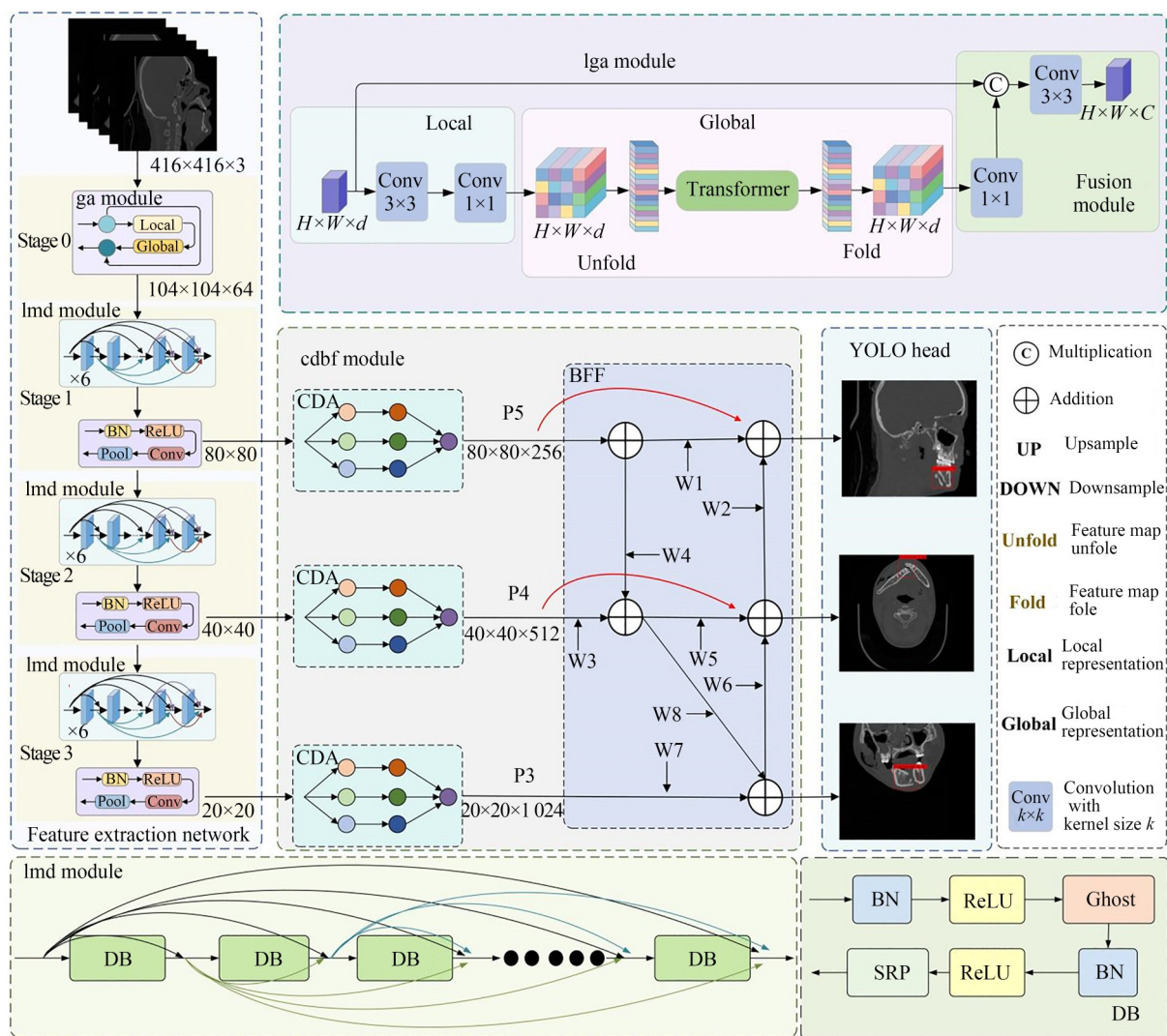


图1 3M-YOLOv5网络结构

Fig. 1 Structure of 3M-YOLOv5 network

## 2.1 特征提取网络

特征提取网络在Densenet121网络的基础上进行改进。Densenet121网络中有4个重复堆叠的密集块结构以及3个过渡层结构,原始网络中密集块的堆叠次数分别为6,12,24和16次,考虑到YOLOv5网络在3个不同尺度的特征图上进行目标检测,在网络结构设计时保留了Densenet121网络的前3个密集块以及过渡层结构作为特征提取网络。过渡层通过池化操作对特征图进行下采样,分别在3个过渡层结构中引出不同尺度的特征图进行特征增强。输入的CT影像首先经过lgaM获得全局信息,然后输入到重复堆叠的密集块结构中进行特征提取,最后在3个过渡层结构中分别输出尺寸为 $80 \times 80$ ,  $40 \times 40$ 和 $20 \times 20$ 的特征图,并输入到特征增强网络中进行特征增强。

40和 $20 \times 20$ 的特征图,并输入到特征增强网络中进行特征增强。

### 2.1.1 局部全局注意力模块

在进行骨折部位检测时,全局信息很重要。骨折部位约占整个CT影像的三分之一,下颌骨的骨折部位在CT影像中的位置信息可以通过提取全局特征来获得,通过局部全局信息的相结合提高网络的表征能力。利用Transformer结构可以获得影像的全局特征,在先前的Transformer应用中,采用Self-Attention来计算当前像素值与其他所有像素值的相关性,这种操作的计算量以及参数量较大,而在图像特征提取时,相邻像素间的差距并不是很大,像素所代表的信息相似<sup>[15]</sup>。因此,本文在特征提取网络的第一个阶段

引入 lgaM, 其结构如图 1 所示, 通过在每一个 patch 内计算 Self-Attention 来减少参数量, 通过卷积操作和 Transformer 结构增强网络的特征提取能力, 获得骨折部位的全局信息。lgaM 主要包括局部表示模块、全局表示模块以及融合模块 3 部分。首先, 使用一个卷积核尺寸为  $3 \times 3$  的卷积层来提取 CT 影像的局部特征, 然后通过一个卷积核尺寸为  $1 \times 1$  的卷积层调整通道数。接着, 通过全局表示模块来提取全局特征, 全局表示模块包括 Unfold, Transformer 和 Fold 3 部分, 然后再通过一个卷积核尺寸为  $1 \times 1$  的卷积层调整通道数。最后, 通过残差连接将全局建模后的特征图与最初输入的特征图在通道方向上进行拼接操作, 拼接后的特征图通过一个卷积核尺寸为  $3 \times 3$  的卷积层做特征融合得到输出。Unfold, Transformer 和 Fold 这 3 部分是 lgaM 的关键部分。假设输入全局表示模块的特征图尺寸为  $H \times W \times d$ , 首先将特征图进行 patch 划分, 每个 patch 的  $H$  和  $W$  都为 2。

如图 2 所示, Unfold 操作按照每个 patch 中相对位置相同的像素进行展开, 展开后的特征图通过 Transformer 部分进行自注意力计算, 在每一个片状特征图内分别计算自注意力, 这种操作能够极大地减少参数量。Fold 操作在自注意力计算完成之后将特征图按照相对位置“折叠”回原来的形状。自注意力的计算公式为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (1)$$

假设有输入特征图矩阵  $\mathbf{a}$ , 将节点分别经过可训练的共享矩阵  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$  变换后得到对应的  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 。  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  矩阵为:

$$\mathbf{Q} = \mathbf{a}\mathbf{W}_q, \mathbf{K} = \mathbf{a}\mathbf{W}_k, \mathbf{V} = \mathbf{a}\mathbf{W}_v. \quad (2)$$

将矩阵  $\mathbf{Q}, \mathbf{K}$  中的值进行点乘操作获得两者

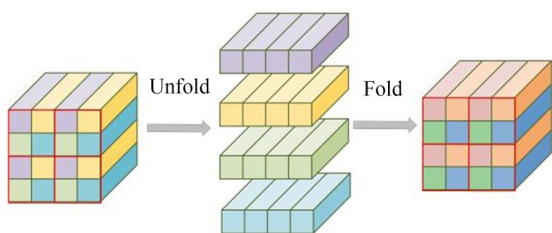


图 2 Unfold 以及 Fold 操作

Fig. 2 Unfold and Fold operations

的相关性, 点乘结果越大, 矩阵  $\mathbf{V}$  的值越大。点乘后的结果除以  $\sqrt{d_k}$  后经过 Softmax 函数处理, 其中  $d_k$  为向量  $k$  的维度, 处理后的结果与  $\mathbf{V}$  进行矩阵相乘即得到自注意力机制的计算值。

### 2.1.2 轻量化多尺度密集模块

在基本的密集块结构中, 网络的浅层特征通过跳跃连接向深层网络传递特征, 这种特征传递的机制能够将浅层的特征图信息传递到深层网络中, 有效地防止梯度消失, 使得深层网络也能够获得较强的纹理、边缘等信息。密集块结构中通过堆叠多个卷积层来进行特征提取, 在多次重复堆叠密集块结构后, 网络的参数量大增, 导致整体网络的参数量以及训练时间增大。在进行特征提取时, 普通卷积操作获得的特征图信息存在较多的冗余, 所包含的信息差别较小, 对于骨折部位的检测贡献较小, 这些差异较小的特征图并不需要通过卷积操作来获取, 可以通过简单的线性操作来代替, 从而减小网络的计算量<sup>[16]</sup>。因此, 考虑采用 Ghost 卷积模块来替换原始密集块结构中的  $1 \times 1$  普通卷积, 该模块在减少网络参数量的同时能够获得相同数量的特征图, 降低了网络的复杂度。

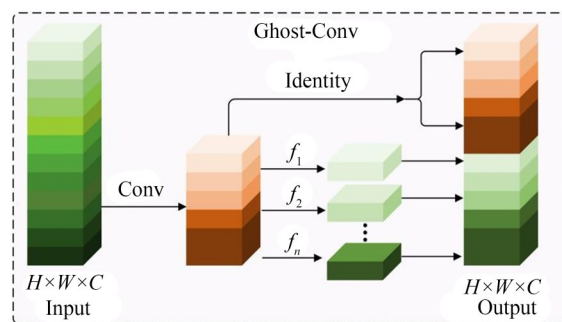


图 3 Ghost 卷积模块

Fig. 3 Ghost convolution module

Ghost 卷积模块是一种轻量化的卷积操作, 是 GhostNet 的主要部分, 它可以以较少的参数量来生成同样通道数量的特征图。如图 3 所示, Ghost 卷积模块的操作主要分为两步: 第一步使用普通的卷积操作压缩特征图的通道数, 通过普通卷积调整输入特征图的通道数, 这里设置该操作的卷积核个数为输入通道数的一半, 此时由于通道数的减少, 模块的计算量也随之减少, 将输入特征图进行信息整合; 第二步, 对通道数压缩

后的特征图进行逐通道方向的卷积操作,该操作与深度可分离卷积中的第一步操作相同,通过  $f_1 \sim f_n$  操作分别对第一步所得到的特征图做逐通道卷积;最后,将第一步与第二步操作得到的特征图在通道方向进行拼接,得到最终处理的特征图。

同样生成  $N$  个特征图,假设输入特征图通道数为  $C$ ,普通卷积的卷积核尺寸为  $K$ ,Ghost 卷积中第一次卷积生成的特征图个数为  $M$ ,第二次逐通道卷积生成  $S$  个特征图, $D$  为逐通道卷积的卷积核尺寸,最终输出  $N = M \times S$  个特征图。则普通卷积的参数量为:

$$P_1 = N \times C \times K \times K. \quad (3)$$

Ghost 卷积第一次卷积的参数量为:

$$P_{21} = \frac{N}{S} \times C \times K \times K. \quad (4)$$

第二次逐通道卷积的参数量为:

$$P_{22} = \frac{N}{S} \times (S-1) \times D \times D, \quad (5)$$

$$P_2 = P_{21} + P_{22}. \quad (6)$$

两者的参数比为:

$$\frac{P_1}{P_2} = \frac{N \times C \times K \times K}{\frac{N}{S} \times C \times K \times K + \frac{N}{S} \times (S-1) \times D \times D} \approx \frac{S \times C}{S + C - 1} \approx S. \quad (7)$$

由此可见,使用 Ghost 卷积的参数量约为普通卷积的  $1/S$ 。在密集块结构中引入 Ghost 卷积

来替换原始的  $1 \times 1$  普通卷积,能够在减少参数量的同时获得数量不变的特征图。为了便于将模块引入到密集块结构中,第一步普通卷积操作中卷积核的个数以及第二步逐通道卷积操作中卷积核的个数均设为输入特征图通道数的一半。通过控制卷积核的个数确保输入输出特征图的尺寸不变,可以将 Ghost 模块直接引入到密集块结构中。

本文采用的数据集有 3 种不同体位的下颌骨 CT 影像,分别是冠状位、矢状位和轴状位。不同体位甚至是同一体位的骨折部位影像尺寸不同,采用同样尺寸的卷积核进行特征提取时获得的局部特征感受野有限,而在进行特征提取时关注多尺度特征有助于网络性能的提升。故考虑在密集块中利用结构重参数化模块 (Structure Reparameterization Module, SRP) 进行改进,如图 4 所示。该模块在训练时使用 3 条不同的分支对图像进行特征提取,在推理时通过结构重参数化操作将 3 条分支合并为一条分支。训练时,第一条分支为卷积核尺寸为  $1 \times 1$  的卷积操作以及批量归一化 (Batch Normalization, BN), 第二条分支为卷积核尺寸为  $3 \times 3$  的卷积操作以及 BN, 第三条分支的特征图不经过卷积只有 BN 操作,将三条分支所获得的特征图进行相加,相加后的特征图再经过挤压激励模块 (Squeeze and Excitation Module, SE Module) 进行进一步精炼。SE 模块首先通过自适应池化操作将特征图压缩成

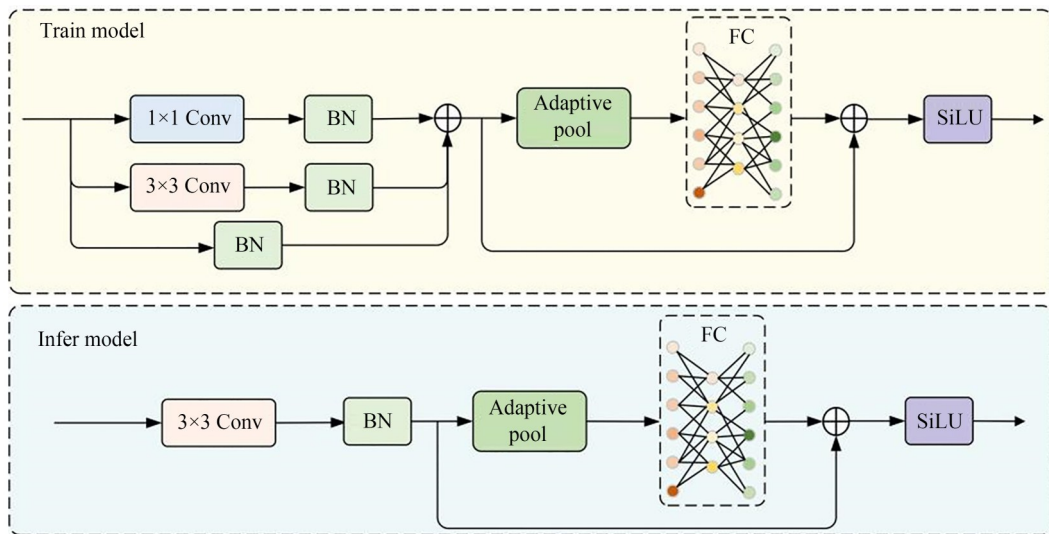


图 4 结构重参数化模块

Fig. 4 Structure reparameterization module

高宽为 1 的长条状,再通过两个全连接层获得注意力信息,将注意力信息与原始特征图相乘使网络关注更有意义的特征,最后将特征图经过 SiLU 激活函数的处理,得到最终的特征图。SiLU 激活函数在数值为负值时并不是直接将数值置零,而是采用极小值来代替,这种操作避免了网络中的神经元为负值时无法收敛的情况。 $1 \times 1$  和  $3 \times 3$  的卷积操作分别提取不同尺度的特征,关注不同尺寸局部区域的下颌骨影像特征。两条分支通过采用不同尺寸的卷积核来提取不同尺度的特征,将不同尺度的特征图进行相加操作,能够融合具有不同尺寸感受野的卷积层所提取的下颌骨 CT 影像特征。

## 2.2 跨维度双向融合模块

### 2.2.1 跨维度注意力模块

在特征提取完成后,会输出 3 种不同尺度的特征图,尺寸分别为  $20 \times 20 \times 1\,024$ ,  $40 \times 40 \times 512$ ,  $80 \times 80 \times 256$ ,将这 3 种不同尺度的特征图输入到特征增强网络中进行处理。此时获得的特征图高度、宽度以及通道之间没有信息的交互,而不同通道提取的信息有着位置、形状等的相关性,特征图的通道和高度以及通道和宽度之间也有着相应的联系,对于下颌骨骨折部位的判断至关重要。本文数据集中包含冠状位、轴状位和矢状位 3 个体位的 CT 影像,不同体位所蕴含的信息也不相同,通过跨维度的特征融合使得 3 个体位的信息有所交互。因此,在输入到特征增强网络中之前,采用跨维度注意力模块,使得特征图的高度和通道之间、宽度和通道之间的信息有所交互,如图 5 所示。同时,本文数据集中采用的 3 个体位的影像也能够补充信息,利用三支结构捕获跨维度交互来计算注意力权重,通过旋转操作和残差变换建立维度、

影像间的依存关系<sup>[17]</sup>。

跨维度注意力模块共有 3 个相互平行的分支,第一条分支负责关注特征图的通道维度  $C$  和空间维度  $W$  的相互关系,第二条分支负责关注特征图的通道维度  $C$  和空间维度  $H$  的相互关系,第三条分支负责捕获特征图的高度  $H$  和宽度  $W$  之间的依赖关系。具体实现过程如下:

第一条分支首先将输入的特征图沿着高度方向逆时针旋转  $90^\circ$  得到形状为  $W \times H \times C$  的特征图,然后通过 Z-POOL, Conv 以及 Sigmoid 激活函数得到注意力权重值, Z-POOL 操作对输入的特征图进行平均池化和最大池化,减小特征图的尺寸,同时提取有意义的特征。最后,将注意力权重值与原始特征图相乘,此时得到的特征图再沿着高度方向顺时针旋转  $90^\circ$  恢复到和输入特征图一致的形状;第二条分支与第一条分支的操作类似,负责计算通道维度  $C$  和空间维度  $W$  的注意力权重,特征图的旋转方式是沿着宽度方向  $W$  逆时针旋转  $90^\circ$ ,此时特征图形状变为  $H \times C \times W$ ,在高度方向进行 Z-POOL 操作,将特征图的高度缩减为 2,再经过 Conv 以及 Sigmoid 激活函数得到注意力权重值,将权重值与原始特征图相乘,最后再顺时针旋转  $90^\circ$  恢复到原来的形状;第三条分支不经过旋转操作,只通过 Z-POOL, Conv 以及 Sigmoid 激活函数得到注意力权重值,再将注意力权重值与原始特征图相乘得到处理后的特征图。最后,对 3 个分支获得的特征图进行相加操作以及取平均值操作。

### 2.2.2 双向特征融合模块

在网络训练过程中,输入到其中的骨折 CT 影像中骨折部位的尺寸不一,在进行卷积以及下采样操作时,随着网络的加深,较大的骨折区域得以保留,而较小的骨折区域可能丢失,因此需要进行特征增强,将不同分辨率的特征图进行融合。深层网络的特征包含抽象的语义信息,但是缺乏空间信息;浅层网络的特征包括边缘、轮廓及形状等影像的原始信息,但是缺乏语义信息,通过将深层特征与浅层特征进行融合,将纹理信息与语义信息融合在一块,能够获得更具表征性的特征图信息<sup>[18]</sup>。

在原始 YOLOv5 网络的特征增强网络中,采

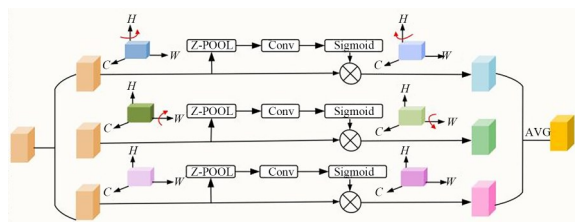


图 5 跨维度注意力模块

Fig. 5 Cross dimension attention module

用特征金字塔网络 (Feature Pyramid Network, FPN) 上下采样拼接的形式来进行特征融合, 将小尺寸的特征图进行上采样与大尺寸的特征图相加, 将大尺寸的特征图进行下采样与小尺寸的特征图相加, 通过上下采样以及相加操作使得不同分辨率、不同尺寸的特征图有所交互, 将网络深层的语义信息传递到浅层特征图中, 同时将浅层网络所提取的纹理、边缘和位置等信息传递到深层网络, 从而融合不同分辨率的特征图, 但是融合时每个特征图的重要程度无法确定。对不同分辨率的特征图进行简单的相加融合操作会导致不同尺寸、不同分辨率的特征图对于融合后特征增强的特征图贡献相同, 尺寸较大的特征图所占比重较大, 而网络深层语义信息丰富的小尺寸特征图所占的比重较小。

如图 1 所示, 跨维度双向融合模块在每个上采样以及下采样的分支中引入可训练的权重值来平衡不同尺度特征图融合时的重要程度, 同时增加了跨尺度的特征拼接操作, 将原始的特征图与特征增强处理后的特征图进行相加操作, 类似于残差网络的结构, 这种结构能够增强网络的表征能力, 对不同尺度的特征图进行区分。

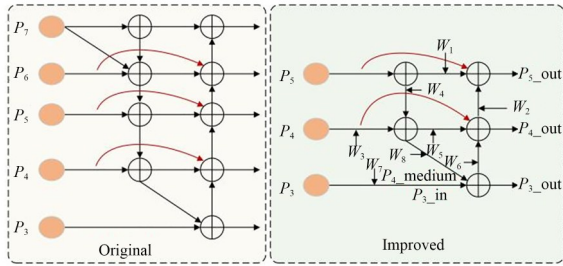


图 6 双向特征融合模块

Fig. 6 Bidirectional feature fusion module

如图 6 所示, 原始融合模块中有 5 个输出, 其中尺寸较大的  $P_6, P_7$  特征图是由特征图  $P_5$  进行上采样得到的,  $P_6, P_7$  特征图结合了尺寸较小的特征图作为输出。YOLOv5 网络中的特征提取网络输出 3 种类型的特征图, 3 种不同尺寸的特征图已经能够较好地满足下颌骨骨折检测的需求, 因此, 本文只采用尺寸较小的 3 个特征图作为检测特征图。特征图  $P_3$  未进行下采样操作, 仅通

过横向连接与特征图  $P_4$  下采样后的特征图进行拼接, 只有单个输出边, 故省略横向的残差连接以减少参数量。输出特征图的计算公式为:

$$Out = \sum_i \frac{W_i \times I_i}{\epsilon + \sum_j W_j}, \quad (8)$$

其中:  $W_i$  为可训练的权重值,  $I_i$  为输入特征图,  $\epsilon$  为一个极小值, 防止分母为 0。图 6 中,  $P_{3\_out}$  为:

$$P_{3\_out} = Conv \left( \frac{W_7 \times P_{3\_in}}{W_7 + W_8 + \epsilon} \right) + Conv \left( \frac{W_8 \times Re\ size(P_{4\_medium})}{W_7 + W_8 + \epsilon} \right). \quad (9)$$

由于  $P_{4\_medium}$  特征图与  $P_3$  特征图的尺寸不同, 因此需要对其进行下采样操作,  $Re\ size$  代表对特征图进行上采样或者下采样操作。式 (7) 中分子为将所有要进行特征融合操作的输入特征图与可训练权重逐个相乘, 分母为所有的权重值与极小值  $\epsilon$  相加。通过引入可训练权重来控制不同尺度特征图融合时的比重, 从而控制不同分辨率的特征图对最终骨折部位检测的贡献值。

### 3 实验与结果分析

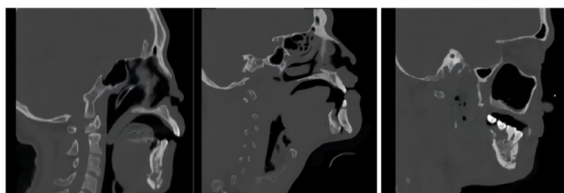
#### 3.1 下颌骨骨折数据集

实验采用的数据集为下颌骨骨折 CT 影像数据集。从宁夏某三甲医院颌面外科处获得下颌骨部位影像的原始图像 5 861 张, 由专业医师对下颌骨骨折部位进行标注, 使用开源 Labelimg 软件标出下颌骨的骨折部位。标签文件为 .xml 格式, 在本数据集中只有骨折这一种类型, 其标注名称为 fractured。标签文件中还包括真实框的高度、宽度、图像的原始尺寸, 以及真实框的坐标等信息。训练集以及验证集按照 9:1 的比例进行划分, 得到训练集影像 4 746 张, 验证集影像 528 张。如图 7 所示, 数据集包含 3 个体位的 CT 影像, 分别为冠状位、矢状位和轴状位影像。每个体位的 CT 影像选取 3 张, 3 个体位共计 9 张 CT 影像。冠状位 CT 是指从患者的正前方向患者进行 CT 影像的拍摄, 此时获得的 CT 影像成为冠状位 CT 影像; 矢状位 CT 是指从患者身体的右侧对患者进行 CT 影像的拍摄, 此时获得的 CT 影像成为矢状位 CT 影像; 轴状位 CT 是指从患者头顶

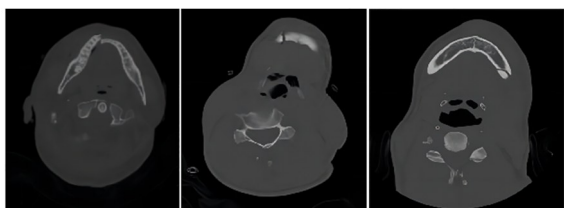
正上方对患者进行 CT 影像的拍摄,此时获得的 CT 影像成为轴状位 CT 影像。通过在不同的体位拍摄 CT 影像,结合不同角度的影像来判断骨折部位,能够获得更充分的信息。



(a) 冠状位CT影像  
(a) The coronal CT images



(b) 矢状位CT影像  
(b) The sagittal CT images



(c) 轴状位CT影像  
(c) The axial CT images

图7 下颌骨CT影像数据集

Fig. 7 Mandibular CT image dataset

### 3.2 实验环境与评价指标

#### 3.2.1 实验环境

实验中,服务器操作系统为 Windows 10 专业版 64 位系统,计算机内存为 64 GB,搭载 Intel Gold 5218 2.3GHz 处理器,显卡为 Titan RTX 24GB, cuda 版本为 11.4。优化器采用 Adam 优化器,动量设置为 0.9,学习率设置为 0.001,权重衰退值为 0.000 1。对每一个网络进行 epoch=100 的迭代,最终选择准确率最高的网络权重进行比较。

#### 3.2.2 评价指标

为了验证网络的有效性,在实验中对比了置信度阈值为 0.5 时网络的 F1 值、召回率、精确率、mAP 值等指标。精确率( $P$ )、召回率( $R$ )和 F1 值

的计算公式分别为:

$$P = \frac{TP}{TP + FP}, \quad (10)$$

$$R = \frac{TP}{TP + FN}, \quad (11)$$

$$F1 = \frac{2 \times (P \times R)}{P + R}, \quad (12)$$

其中:TP 表示在真实正样本中被预测为正样本的样本数量,FP 表示在真实负样本中被预测为正样本的样本数量,FN 表示在真实正样本中被预测为负样本的样本数量。在置信阈值为 0.5 时,对评价指标进行了比较。当预测框与真实框的交并比(Intersection over Union, IoU)大于 0.5 时,该部分为骨折部分,计算该样本为 TP 真阳性样本。当  $0 < \text{IoU} \leq 0.5$  时,将样本计算为 FP。当 IoU 阈值为 0 时,即当真实框与预测框不相交时,未检测到真实框,并将样本计数为 FN。精确率表示在预测为阳性样本的所有样本中,真阳性样本所占的比例,即真实骨折样本在预测为骨折的样本中所占的比例。召回率是指预测的阳性样本在真实阳性样本中的比例,即网络正确预测的样本在所有真实骨折样本中的比例。以召回率为横轴,以精确率为纵轴,绘制出  $P$ - $R$  曲线。该曲线与横轴所围成的面积即为平均精度(Average Precision, AP),其计算公式为:

$$AP = \int_0^1 P(R) dR. \quad (13)$$

mAP 表示所有类别 AP 值的平均值,一般情况下,评价时置信度阈值取 0.5 时所有类别的平均 AP 值来计算 mAP。本文数据集中只有下颌骨骨折一种类型,因此 mAP 等于 AP,其计算公式为:

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}. \quad (14)$$

为了直观地考察不同网络的性能,本文还根据相关实验数据绘制了雷达图,对不同网络的 F1 值、召回率、精确率和 mAP 值进行了可视化。雷达图中,中心区域的数值为零,从中心向四周辐射,数值逐渐增大,数值越大代表网络的各项评价指标越高,网络性能越好,不同颜色的折线分别代表了不同网络。在对比实验部分,给出了不同网络的 FPS(Frame Per Second),即每秒所处理的帧数。FPS 是目标检测算法中另一个重要

评估指标,数值越大,表明网络的检测速度越快。在评估网络的整体性能时,要综合考虑网络的 F1 值、召回率、精确率、mAP 值以及 FPS 做出全面的评估。

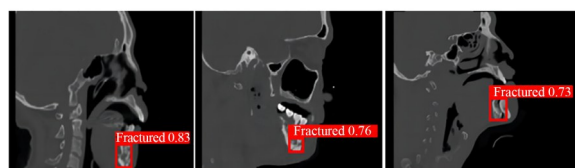
### 3.3 实验结果

图 8 是下颌骨骨折 CT 影像,分别为冠状位影像、矢状位影像和轴状位影像的检测结果。在影像中使用矩形框将骨折部位框出,并且在矩形框的上方给出了“fractured”字样,同时给出了该矩形框部位为骨折的置信度值。



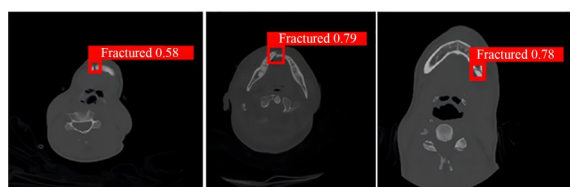
(a) 冠状位影像检测结果

(a) The detection results of coronal images



(b) 矢状位影像检测结果

(b) The detection results of sagittal images



(c) 轴状位影像检测结果

(c) The detection results of axial images

图 8 下颌骨骨折 CT 影像检测结果

Fig. 8 Detection results of mandibular fracture CT images

#### 3.3.1 消融实验

3M-YOLOv5 结构基于 YOLOv5 网络进行改进,设计了 3 个模块改进网络性能,分别是 lga 模块、lmd 模块以及 cdbf 模块。lmd 模块包括 Ghost 卷积模块以及 SRP 模块, cdbf 模块包括 CDA 模块以及 BFF 模块。在消融实验中,通过分别引入不同的模块来验证其有效性。实验一采用 lga 模块、SRP 模块、CDA 模块和 BFF 模块对网络进行改进,去掉 lmdM 模块中的 Ghost 卷积模块,使用基本的  $1 \times 1$  卷积操作进行特征提

取;实验二采用 lga 模块、Ghost 卷积模块、CDA 模块和 BFF 模块对网络进行改进,去掉 lmdM 模块中的 SRP 模块,使用基本的  $3 \times 3$  卷积操作进行特征提取;实验三采用 lga 模块、Ghost 卷积模块、SRP 模块和 CDA 模块对网络进行改进,去掉 BFF 模块,使用原始网络中的 FPN 网络进行特征增强;实验四使用 lga 模块、Ghost 卷积模块、SRP 模块和 BFF 模块进行网络改进,去掉 CDA 模块;实验五使用 Ghost 卷积模块、SRP 模块、CDA 模块和 BFF 模块对网络进行改进,去掉网络开始阶段的 lga 模块;实验六去掉网络开始阶段的 lga 模块和 BFF 模块,只保留 Ghost 卷积模块、SRP 模块和 CDA 模块;实验七只保留 lga 模块、CDA 模块和 BFF 模块,去掉密集块结构中的 Ghost 卷积模块和 SRP 模块,采用基本的  $1 \times 1$  和  $3 \times 3$  卷积进行特征提取;实验八利用特征提取网络引入的 lga 模块、Ghost 卷积模块和 SRP 模块对网络进行改进,去掉 CDA 模块和 BFF 模块;实验九只保留 CDA 模块和 BFF 模块,在骨干网络中使用原始的密集块结构进行特征提取;实验十引入所有模块来改进网络。

消融实验结果如表 1 所示。实验一网络的 mAP 值较引入所有模块时下降了 0.08%,参数量较最终网络结构上升了 1.969 MB,证明 Ghost 卷积模块相较于  $1 \times 1$  卷积能够减少参数量。各项评价指标也有所下降,证明在密集块结构中引入 Ghost 卷积模块所带来的性能提升有限。Ghost 卷积模块虽然能够以较少的参数量来获得同样数量的特征图,但是由于其最终特征图中有一半的特征图没有经过处理,而是类似于残差结构直接与另一半特征图进行通道拼接,因此它所带来的性能提升有限。实验二网络的 mAP 值较引入所有模块时下降了 0.95%,置信度阈值取 0.5 时,网络的 F1 值、召回率、精确率分别下降 1.97, 2.05 以及 1.89,证明了 SRP 模块中的  $1 \times 1$  以及  $3 \times 3$  卷积能够关注到不同尺寸的局部区域,同时在模块中引入的 SiLU 激活函数能够较好地防止神经元在负值时的死亡,有利于网络的收敛。实验三网络的 mAP 值较引入所有模块时下降了 0.76%,证明了采用 BFF 模块能够提升网络的性能,在进行不同分辨率特征图的融合拼接

表 1 消融实验结果

Tab. 1 Result of ablation experiment

网 络	参数量/MB	GFLOPS	精确率	召回率	F1 值	mAP 值
YOLOv5	7.064	16.477	0.953 8	0.948 9	0.953 1	0.956 1
实验一	12.220	70.187	0.986 2	0.971 0	0.978 5	0.990 9
实验二	9.997	58.818	0.974 3	0.967 6	0.970 9	0.982 2
实验三	8.924	54.124	0.982 8	0.972 7	0.977 7	0.984 1
实验四	10.251	61.116	0.982 9	0.977 9	0.980 4	0.989 0
实验五	9.389	52.430	0.982 8	0.971 0	0.976 9	0.987 8
实验六	8.780	46.764	0.973 3	0.930 2	0.951 2	0.977 3
实验七	11.966	67.862	0.986 0	0.957 4	0.971 5	0.979 2
实验八	8.924	54.098	0.974 4	0.971 0	0.972 7	0.979 1
实验九	11.104	59.675	0.881 6	0.913 1	0.897 1	0.920 8
实验十	10.251	61.143	0.993 2	0.988 1	0.990 6	0.991 7

时能够较好地平衡不同特征图的融合权重,同时在特征提取网络中同时引入(1)、(2)、(3)、(4)这 4 个模块能够增强网络的特征提取能力,较好地获取到不同大小骨折区域的特征以及骨折部位在整个下颌骨 CT 影像中的全局信息。实验四网络的 mAP 值下降了 0.27%,证明引入 CDA 模块能够关注特征图的高度、宽度和通道之间的关系,进一步提升网络性能,但是性能提升有限;实验五中去掉 lgaM 模块后,置信度阈值取 0.5 时网络的 mAP 值、F1 值、召回率、精确率分别下降 0.39%,1.37%,1.71%,1.04%,同时参数量与计算量也有所下降,lgaM 模块以较小的代价来提取全局特征信息,证明 lgaM 模块能够在特征提取网络的开始阶段获得全局表征信息,同时 Un-fold 以及 Fold 操作能够减少计算注意力时的参数量。实验六网络的 mAP 值下降较多,与原始 YOLOv5 网络相比仅提升了 2.12%,证明了在网络的最初始阶段引入 lgaM 模块能够使网络更好地捕获全局特征信息,从而提升网络的特征提取能力,同时 BFF 模块在进行特征增强时能够通过可训练的权重来决定不同分辨率的特征图融合的权重。实验七网络 mAP 值仅提升了 2.31%,同时网络的参数量也有所提升,证明了原始密集块结构中的卷积操作参数量较大,特征提取能力较好,但是将原始卷积操作分别替换为 Ghost 卷积模块以及 SRP 模块后,网络的特征提取能力能

够进一步的提升。实验八网络的 mAP 值提升了 2.3%,证明采用 BFF 模块能够平衡不同分辨率特征图融合时的权重,同时 CDA 模块能够关注特征图的高度、宽度和通道之间的关系,进一步提升网络性能。实验九网络性能下降得最多,甚至远不如原始 YOLOv5 网络,F1 值等各项评价指标也大幅下降,证明了仅采用两个模块来改进网络无法很好地进行下颌骨骨折部位的检测。实验十同时引入 5 个模块,网络的各项指标最大,证明本文所提出的 3M-YOLOv5 网络能够较好地进行下颌骨骨折部位的检测。图 9 给出了消融实验的雷达图,通过该图可以看出,实验十的折线位于最外侧,各项性能最优。

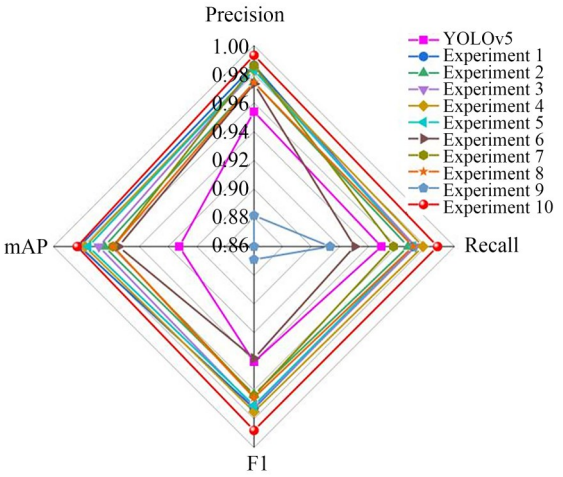


图 9 消融实验雷达图

Fig. 9 Radar map of ablation experiment

### 3.3.2 对比实验

本文通过对比不同目标检测网络在下颌骨骨折检测方面的各项评价指标,验证了所提出的 3M-YOLOv5 下颌骨骨折检测网络的有效性。在提出的下颌骨骨折 CT 影像数据集上,采用 3M-YOLOv5 网络与 YOLOv3 网络、YOLOv4 网络、FasterRCNN、CenterNet、YOLOX

网络以及 YOLOv7 网络进行了检测。其中,YOLOX 选用 YOLOX\_s 版本进行对比;YOLOv7 网络有针对于边缘 GPU、普通 GPU 和云 GPU 的 3 种基本网络,分别命名为 YOLOv7tiny、YOLOv7 和 YOLOv7-W6,本文选取适用于普通 GPU 版本的 YOLOv7 网络。实验结果如表 2 所示。

表 2 不同实验结果对比

Tab. 2 Comparison of experiment results with different networks

网 络	参数量/MB	GFLOPS	精确率	召回率	F1 值	mAP 值	FPS
YOLOv3 <sup>[19]</sup>	236.32	66.171	0.844 6	0.277 7	0.417 9	0.638 9	11.819 3
YOLOv4 <sup>[20]</sup>	245.53	60.527	0.908 3	0.691 7	0.785 3	0.851 5	9.465 7
YOLOv5	7.064	16.477	0.953 8	0.948 9	0.953 1	0.956 1	14.051 6
Faster RCNN <sup>[21]</sup>	137.099	370.210	0.919 7	0.858 6	0.888 1	0.927 4	7.059 4
CenterNet <sup>[22]</sup>	32.665	70.217	0.877 5	0.524 7	0.656 7	0.794 0	<b>24.972 8</b>
YOLOX <sup>[23]</sup>	8.938	26.635	0.918 4	0.901 2	0.909 7	0.911 8	22.099 2
YOLOv7 <sup>[24]</sup>	37.195	104.761	0.910 3	0.812 6	0.858 7	0.901 6	12.669 1
3M-YOLOv5	10.251	61.143	<b>0.993 2</b>	<b>0.988 1</b>	<b>0.990 6</b>	<b>0.991 7</b>	2.682 6

YOLOv3 网络采用 Darknet-53 作为主干特征提取网络,网络结构中采用了卷积层进行特征的提取,通过采用步长为 2 的卷积层来进行下采样操作,没有使用池化层,避免了池化操作所带来的特征信息丢失等问题<sup>[25]</sup>。但是由于其特征提取不够充分,仅通过两次上采样进行特征图拼接,未进行特征增强,高分辨率特征图与低分辨率特征图之间没有交互,缺少信息互补,因此整体效果并不理想。YOLOv4 网络是 YOLO 系列网络中第一次采用小尺寸特征图上采样与大尺寸特征图下采样进行特征网络的增强,在特征提取网络中还引入了空间金字塔池化(Spatial Pyramid Pool, SPP)结构来提取不同尺度的特征,但是网络性能的提升效果有限。Faster RCNN 网络是常用的两阶段检测网络,它首先生成区域建议候选框,再对候选框进行分类。网络的检测效果较好,但是网络的参数量以及计算量数倍于其他 YOLO 系列网络,极大增加了训练以及部署时的成本。CenterNet 是一种无锚框的检测网络,不依赖大量的锚框,其参数量以及计算量较少,但是由于它只在高语义信息的特征图上进行特征检测,未结合大尺寸、大分辨率的特征图,丢失

了较多的细节信息,因此其检测效果较差。YOLOX 网络的参数量和 YOLOv5 相差无几,计算量提升较多,模型结构更加复杂,同时图片每秒的检测速度也有所提升,但是对于骨折部位的特征提取不够充分,因此整体性能提升较少。YOLOv7 网络的性能与 YOLOX 网络接近,但是由于二者都是通用检测网络,对于下颌骨骨折部位的特征提取不够充分,其性能甚至弱于 YOLOv5 网络。本文提出的 3M-YOLOv5 在置信度阈值取 0.5 时的 mAP 值、F1 值、召回率和精确率分别为 0.991 7, 0.990 6, 0.988 1, 0.993 2, 与其他网络相比,整体性能提升最多。3M-YOLOv5 网络的 FPS 值在所有网络中最低,其检测速度最慢。但是考虑到在医学图像辅助诊断中,漏诊、误诊等现象是不被允许的,精度要求是第一位的,其次才会考虑检测速度。因此,综合考虑网络的整体性能,以牺牲速度来换取准确率,网络的整体表现还是可以接受的。图 10 给出了对比实验的雷达图,通过该图可以看出,3M-YOLOv5 网络的 mAP 值、F1 值、召回率和精确率折线均位于最外侧,各项性能最优。

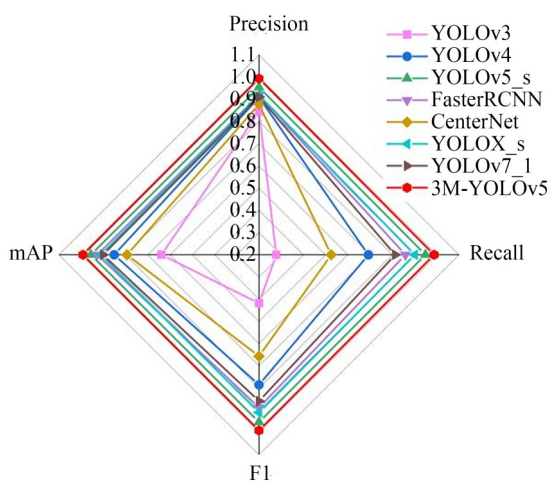


图 10 对比实验雷达图

Fig. 10 Radar map of comparison experiment

## 4 结 论

下颌骨骨折时需要根据不同的骨折部位采取不同的治疗手段,准确及时地定位骨折部位能够给予医生充足的治疗参考信息,当前相关检测网络存在特征提取不充分、检测精度不高等问题。本文提出了一种用于下颌骨骨折检测的 3M-YOLOv5 网络,它基于 YOLOv5 网络进行改进。首先,引入密集连接思想来改进特征提取网

络,在 3 个过渡层中分别引出 3 种不同尺度的特征图输入到特征增强网络中;其次,在重复堆叠的密集块中引入 Ghost 卷积模块以及结构重参数化模块,使得特征提取网络能够提取多尺度信息,从而增强网络的特征提取能力;然后,在跨维度双向融合模块中采用跨维度注意力来融合特征图的高度、宽度以及通道之间的信息;最后采用双向特征融合模块,在特征图进行上下采样的过程中加入可训练的权重,更好地平衡了不同尺度特征图融合时的重要程度。实验结果表明,改进后的 3M-YOLOv5 网络的 F1 值、召回率、精确率以及 mAP 值分别为 99.06%, 98.81%, 99.32% 和 99.17%。

针对单一类别即下颌骨骨折类型,改进后网络的 mAP 值能够达到 99% 以上,表明 YOLOv5 网络在单一目标检测领域也能取得很好的效果。但是,本文未对下颌骨的骨折部位给出明确的分类,只检测出骨折部位。未来工作要对具体的下颌骨骨折部位给出明确的分类,例如具体到髁突、下颌体、下颌角等具体部位的骨折信息,同时可以进行下颌骨骨折 CT 影像的实时分析,帮助医师在进行 CT 影像拍摄时能够更快更好地找到下颌骨的骨折部位,从而选取更加有代表性的 CT 影像供给主治医师进行治疗的参考。

## 参考文献:

- [1] ROCCIA F, SOBRERO F, RAVEGGI E, *et al.* European multicenter prospective analysis of the use of maxillomandibular fixation for mandibular fractures treated with open reduction and internal fixation [J]. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 2023, 124(1): 101376.
- [2] ZHOU T, HOU S B, LU H L, *et al.*. Exploring the improved mechanism of U-Net and its application in medical image segmentation [J]. *Journal of Biomedical Engineering*, 2022, 39(04): 806-825.
- [3] MENG X H, WU D J, WANG Z, *et al.* A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance[J]. *Skeletal Radiology*, 2021, 50(9): 1821-1828.
- [4] ZHOU Q Q, TANG W, WANG J S, *et al.* Automatic detection and classification of rib fractures based on patients' CT images and clinical information via convolutional neural network[J]. *European Radiology*, 2021, 31(6): 3815-3825.
- [5] XUE L Y, YAN W N, LUO P, *et al.* Detection and localization of hand fractures based on GA\_Faster R-CNN [J]. *Alexandria Engineering Journal*, 2021, 60(5): 4555-4562.
- [6] KITAMURA G, CHUNG C Y, MOORE B E. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, *de novo* training, and multiview incorporation [J]. *Journal of Digital Imaging*, 2019, 32 (4) : 672-677.
- [7] YUAN G, LIU H Z, JIANG L, *et al.* CCE-Net: a rib fracture diagnosis network based on contralateral, contextual, and edge enhanced modules[J]. *Biomedical Signal Processing and Control*, 2022, 75: 103620.
- [8] LIU P R, LU L, CHEN Y F, *et al.* Artificial intelligence to detect the femoral intertrochanteric fracture: the arrival of the intelligent-medicine era[J]. *Frontiers in Bioengineering and Biotechnology*,

- 2022, 10: 927926.
- [9] WU J F, LIU N J, LI X J, *et al.* Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study[J]. *BMC Medical Imaging*, 2023, 23(1): 1-12.
- [10] WARIN K, LIMPRASERT W, SUEBNUKARN S, *et al.* Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs[J]. *International Journal of Oral and Maxillofacial Surgery*, 2022, 51(11): 1488-1494.
- [11] WANG X B, XU Z N, TONG Y H, *et al.* Detection and classification of mandibular fracture on CT scan using deep convolutional neural network[J]. *Clinical Oral Investigations*, 2022, 26(6): 4593-4601.
- [12] VINAYAHALINGAM S, VAN NISTELROOIJ N, VAN GINNEKEN B, *et al.* Detection of mandibular fractures on panoramic radiographs using deep learning[J]. *Scientific Reports*, 2022, 12: 19596.
- [13] SON D M, YOON Y A, KWON H J, *et al.* Automatic detection of mandibular fractures in panoramic radiographs using deep learning[J]. *Diagnostics*, 2021, 11(6): 933.
- [14] ZHOU T, YE X Y, LU H L, *et al.* Dense convolutional network and its application in medical image analysis[J]. *BioMed Research International*, 2022, 2022: 23848301-22.
- [15] MEHTA S, RASTEGARI M. MobileViT: lightweight, general-purpose, and mobile-friendly vision transformer[C]. *Computer Vision and Pattern Recognition*, 2021, arXiv preprint arXiv: 2110.02178.
- [16] ZHANG S, CHE S B, LIU Z, *et al.* A real-time and lightweight traffic sign detection method based on ghost-YOLO[J]. *Multimedia Tools and Applications*, 2023, 82(17): 26063-26087.
- [17] WANG S, QU Z, LI C, *et al.* BANet: Small and multi-object detection with a bidirectional attention network for traffic scenes[J]. *Engineering Applications of Artificial Intelligence*, 2023, 117: 105504.
- [18] CHEN Y T, XIA R L, YANG K, *et al.* MFFN: image super-resolution via multi-level features fusion network [J]. *The Visual Computer*, 2023: 1-16.
- [19] JOSEPH R, ALI F. YOLOv3: An incremental improvement [C]. *Computer Vision and Pattern Recognition*, 2018, arXiv preprint arXiv: 1804.02767.
- [20] ALEXEY B, WANG C Y, LIAO H Y M, *et al.* . YOLOv4: Optimal speed and accuracy of object detection [J]. *Image and Video Processing*, 2020, arXiv preprint arXiv:2004.10934, 2020.
- [21] REN S Q, HE K M, 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*, 2017, 39(6): 1137-1149.
- [22] ZHOU X Y, WANG D Q, KRÄHENBÜHL P, *et al.* Objects as Points [C]. *Computer Vision and Pattern Recognition*, 2019: 1904.07850.
- [23] GE Z, LIU S, WANG F, *et al.* YOLOX: Exceeding YOLO Series in 2021 [EB/OL]. 2021: arXiv: 2107.08430. <https://arxiv.org/abs/2107.08430.pdf>.
- [24] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-art for Real-time Object Detectors [EB/OL]. 2022: arXiv: 2207.02696. <https://arxiv.org/abs/2207.02696.pdf>.
- [25] ZHOU T, CHANG X Y, LU H L, *et al.* Pooling operations in deep learning: from “invariable” to “variable” [J]. *BioMed Research International*, 2022, 2022: 1-17.

## 作者简介:



周 涛(1977—),男,博士,教授,博士生导师,CSIG 理事,CSS 理事,主要从事计算机辅助诊断,医学图像分析与处理,模式识别等方面的研究。E-mail:zhoutaonxmu@126.com

## 通讯作者:



杜玉虎(1999—),男,硕士研究生,主要从事图像图形智能处理方面的研究。E-mail:cy\_dyh@163.com