文章编号:1671-4598(2024)12-0035-08 DOI:10.16526/j.cnki.11-4762/tp.2024.12.006 中图分类号:TP391.41 文献标识码:A

基于改进 YOLOv7 的室内摔倒行为检测

陈华艳，张晓滨
（西安工程大学 计算机科学学院，西安 710048）

摘要：针对室内监控视频中老年人摔倒行为的检测问题，提出一种基于改进 YOLOv7 网络模型的实时摔倒行为检测算法；基于 YOLOv7 的目标检测模型传统使用跨步卷积来实现下采样特征，但这可能会使目标信息的特征模糊；为了解决这个问题，引入了新的下采样模块——鲁棒特征下采样，以改善下采样过程中目标信息特征的清晰度；此外，通过在网络的 concat 部分引入 CoordAttention 注意力机制，可更好地融合拼接后的特征图；实验结果表明，改进后的 YOLOv7 模型在摔倒行为检测方面具有较高的准确率和检测性能，准确率达到 98.88%，*mAP*50 值达到 98.83%，*mAP*50∶95 值达到 74.12%；这意味着该算法可以准确地检测老年人的摔倒行为，家人能够及时地发现，以便及时采取必要的救助措施。

关键词：摔倒检测；YOLOv7 网络模型；下采样；鲁棒特征下采样；CoordAttention 注意力机制

Indoor Falling Behavior Detection Algorithm Based on Improved YOLOv7

CHEN Huayan, ZHANG Xiaobin
(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: To address the problem of detecting falls for the elderly people in indoor surveillance video, a real-time fall behavior detection algorithm based on improved YOLOv7 network model was proposed. the strided convolution is traditionally used in the target detection model based on YOLOv7 to realize the downsampling feature, but this perhaps make the feature of the target information fuzzy. To solve this problem, a novel downsampling module, robust feature downsampling, is introduced to improve the clarity of target information features during downsampling. In addition, by introducing the CoordAttention attention mechanism in the concat section of the network, the spliced feature graphs can be better merged. Experimental results show that the improved YOLOv7 model has a high accuracy and detection performance in falling behavior detection, with an accuracy of 98.88%, *mAP*50 value of 98.83%, and *mAP*50∶95 value of 74.12%. This means that the algorithm can accurately detect the fall behavior of the elderly, so the family should promptly discover and make necessary rescue measures in a timely manner.

Keywords: falling detection; YOLOv7 network model; downsampling; robust feature downsampling; CoordAttention attention mechanism

0 引言

随着人口老龄化程度进一步加深，老年人独自居家养老的各种意外情况更是频繁发生，世界卫生组织研究表明，摔倒问题在中国老年人疾病负担中排名第八。中国每年摔倒的老年人数大约占总老年人数的 30%，摔倒成为老年人意外死亡的第二大原因。因此，如何保障老年人独居养老的安全性，成为家庭和社会的重要难题。老年人意外摔倒后的快速响应成为当下的研究热题，研究一种自动检测视频中摔倒行为的方法，及时地发现老年人摔倒并施加援助具有很重要的现实意义[1-2]。

基于计算机视觉的摔倒检测是对采集到的视频进行处理，检测是否存在摔倒行为。该方法由于具有固定摄像头获得连续供电保证实时监控、无需佩戴任何设备、不易受外界因素干扰、检测精度高等特点，受到广泛关注，成为摔倒检测研究的热点[3]。深度学习算法在行为检测领域得到广泛应用，双阶段目标检测和单阶段目标检测是两种常见的目标检测方法。双阶段目标检测方法通常包括两个主要阶段：提取候选框和对候选框进行分类与定位。典型的双阶段目标检测方法是 Faster R-CNN[4]。其主要步骤是首先使用区域提取网络（RPN）生成一组候选框，并对这些候选框进行分类和位置调整，因为它们在两个阶段都进行了专门的学习和优化，通常能够提供较高的目标检测准确性，但在速度上可能会有所牺牲。单阶段目标检测方法通过一个单一的网络直接预测目标的类别和位置。YOLO（You Only Look Once）系列[5]和 SSD（Single Shot MultiBox Detector）[6]是常见的单阶段目标检测方法。其一般只需要一次前向传播即可完成目标检测，因此速度相对较快，简单高效，适用于实时应用和对速度要求较高的场景。

随着 YOLO 系列的发展，文献［7］提出一种基于改进 YOLOv3[8]模型的实时摔倒检测算法，利用残差模块构建图

收稿日期:2023-11-01； **修回日期**:2023-12-13。
作者简介:陈华艳(1998-)，女，硕士研究生。
通讯作者:张晓滨(1970-)，男，硕士，副教授。
引用格式:陈华艳，张晓滨. 基于改进 YOLOv7 的室内摔倒行为检测[J]. 计算机测量与控制，2024，32(12)：35-42，87.

像快速特征提取网络，同时引入通道域注意力机制（SENet），实现对特征图的各个通道赋予不同的权重，提升模型检测准确性；其次，采用 CIoU 边界框回归损失函数，降低目标的漏检率，该算法的 AP 值为 92.1%，检测速度为 35 f/s，具有良好的检测性能。文献［9］提出了一种基于改进 YOLOv5s[10] 的老年人摔倒行为实时检测方法，在 Backbone 网络中使用非对称卷积块（ACB）卷积模块来代替现有的基本卷积，然后，在 Backbone 网络的残差结构中加入空间注意机制模块，以提取更多的特征位置信息。最后对特征层结构进行改进，该算法可以准确地检测摔倒行为的老年人，平均准确率达到 97.2%。文献［11］在 Ghost 模块中插入高效通道注意（ECA）模块，形成了 GED（Ghost-ECA-Dense）模块，利用 GED 模块构建 F-GED，这是一个参数更少、性能更好的主干，取代了 YOLOv5 中的 CSPDarknet 53。其次，在 YOLOv5 的颈部和头部用 Ghost 模块和深度可分离卷积代替冗余操作，构造 YOLOv5-R，平均精度达到 72.7%，每秒浮点运算次数和参数分别减少了 25% 和 47.2%，模型重量仅为 7.52 MB，每秒帧数达到 37 帧，显著减小了网络规模，提供实时检测。文献［12］为了解决当前人体摔倒检测的问题，提出了基于 YoloX-s[13] 结合轻量级 OpenPose[14] 提取人体骨架模型。该模型可以通过人体颈部和膝部关键点之间的角度变化率差值来识别人体摔倒，该方法总体准确率为 96.94%，其中摔倒行为检测的准确率为 97.92%，正常行为检测的准确率为 96.46%。文献［15］针对小目标检测更容易出现漏检等问题，提出一种改进的 YOLOv7 目标检测模型，对 YOLOv7 网络模型中的 MPConv 模块进行改进，以减少网络特征处理过程造成的特征损失，利用 ACmix 注意力模块提高网络对小尺度目标的敏感度，在此基础上，使用 SIoU 替换原 YOLOv7 网络模型中的 CIoU 来优化损失函数，减少损失函数自由度，提高网络鲁棒性，改进后的 YOLOv7 网络模型相比原网络，漏检情况得到明显改善，且 mAP 达到 71.1%。

人体摔倒检测是一项具有挑战性的研究任务，为了准确地检测到室内摔倒行为，本文提出一种改进的 YOLOv7 摔倒检测模型。通过对 YOLOv7 网络模型的采样模块进行改进，采用鲁棒特征下采样的方法，以提取和保留更多的特征信息，并在网络的 concat 处加入注意力机制，以更好地融合拼接后的特征图，提高模型的泛化能力。

1 YOLOv7 网络模型

YOLOv7[16] 是一种基于深度学习的目标检测算法。其核心思想是将目标检测任务转化为一个回归问题。它将输入图像分成一个固定大小的网格，并在每个网格单元中预测目标的类别和位置。与传统的目标检测算法相比，YOLOv7 的优势在于它能够实现实时的目标检测和定位。网络结构如图 1 所示。

YOLOv7 的网络结构主要由 Backbone 和 Head 两个组成部分构成：

1）Backbone：YOLOv7 基于 Darknet 架构，它是一种轻量级卷积神经网络。Darknet 由一系列卷积层和池化层组成，用于从原始图像中提取特征。在 YOLOv7 中，采用了 CSPDarknet53 作为主干网络。CSPDarknet53 采用了 Cross Stage Partial Network 结构，它在每个卷积层之前插入了残差连接，以提高信息流动和梯度传播。CSPDarknet53 具有 53 个卷积层，能够较好地提取图像的语义特征。

2）Head：YOLOv7 引入了特征金字塔网络 FPN（Feature Pyramid Network）模块。它通过构建多尺度的特征金字塔，使得网络能够处理不同尺度的目标。FPN 由不同尺度的特征图融合而成，它引入了横向连接和上采样操作，以获取具有高层次语义信息和丰富空间细节的特征图。YOLOv7 的输出层由几个预测层组成，每个预测层负责在不同尺度上预测目标的类别和位置。每个预测层生成一堆 Anchor boxes，然后根据预测结果和阈值来筛选出可能的目标框。最后，通过非极大值抑制（NMS）算法来消除重叠的目标框，以输出最终的检测结果。

YOLOv7 相比于前代版本的创新主要体现在以下几个方面：

1）模型结构：YOLOv7 采用了更深的主干网络 Darknet-53，提取更丰富的特征信息，同时引入 FPN 结构，增加了多尺度的特征融合，使得模型在检测小目标等场景下性能更好；

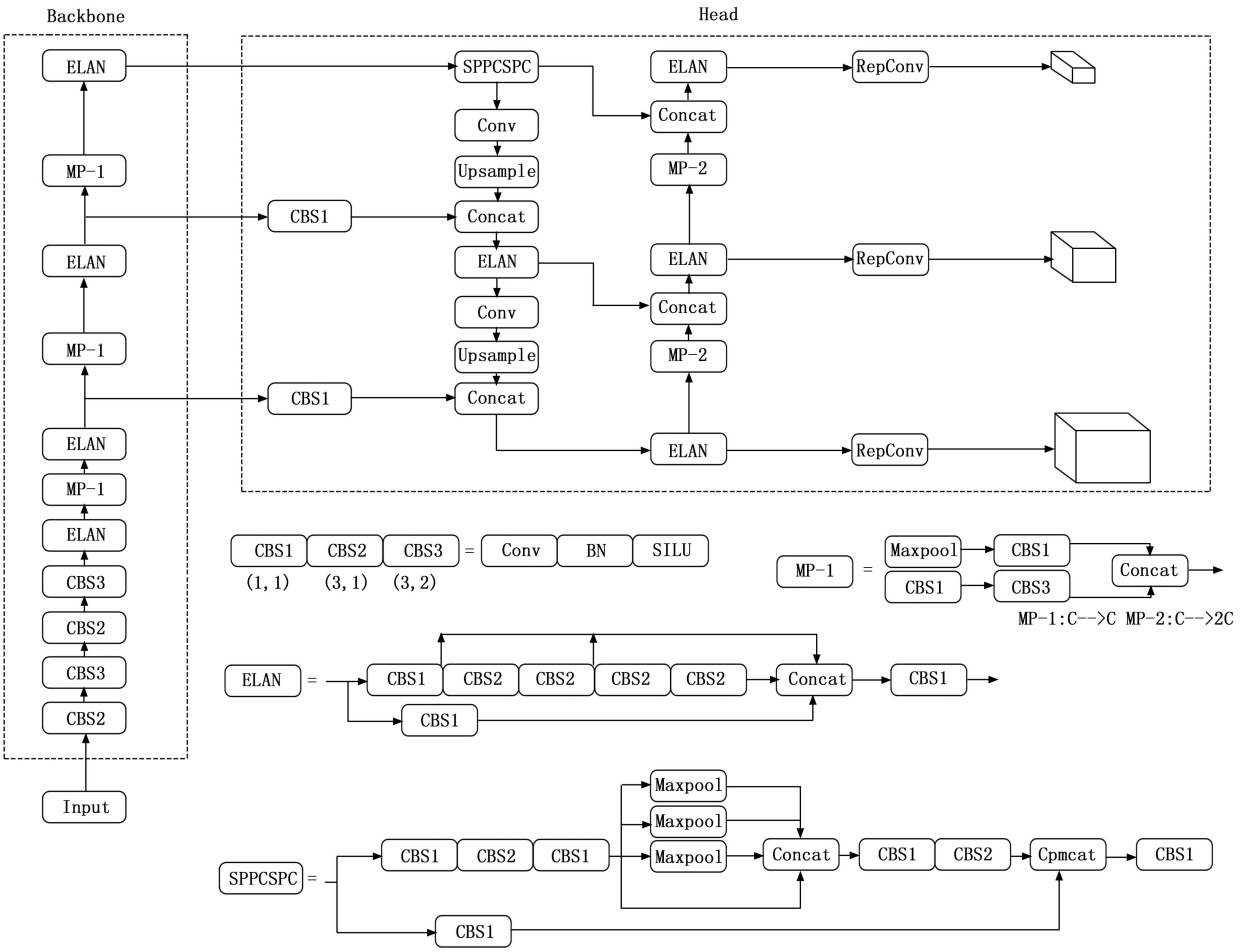
2）训练策略：YOLOv7 使用了一种新的训练策略，包括多尺度训练、分批次训练和动态权重调整等技术，这些策略有助于模型更好地适应不同尺度和形状的目标，并提高模型的鲁棒性；

3）性能优化：YOLOv7 针对目标检测的速度和精度进行了性能优化，通过改进网络结构和参数设置，提高了模型的检测速度，并在保持较高准确率的同时减少了计算成本。

人体的异常行为通常具有无规律、突发性和不可预见性等特点。因此，将在 YOLOv7 目标检测模型应用到摔倒检测模型上效果较好。虽然 YOLOv7 是一种强大的目标检测模型，但在小目标检测、定位精度、多类别预测等方面存在一些缺点。因此，将在 YOLOv7 模型的基础上进行改进。

2 下采样模块

下采样是计算机视觉中的一个基本操作，它可以降低图像或特征图的分辨率，同时保留重要的信息。在卷积神经网络等深度学习模型中，当输入数据的维度较高时，模型可能过于复杂，容易在训练集上过拟合，而在未见过的数据上表现较差。通过下采样，可以降低数据的维度，减少模型的复杂度，从而减少过拟合的可能性。另外，下采样还有助于提高模型的泛化能力。降低分辨率可以移除一些细节信息，而保留主要的结构和特征，这有助于模型更



注：Conv为卷积；BN为批归一化；SiLU为激活函数；Upsample为上采样模块；ELAN为高效聚合层网络； Maxpool为最大池化；SPPCSPC为空间金字塔池化结构；RepConv为重参数化卷积；Concat为特征拼接。

图 1　YOLOv7 网络架构

好地学习数据的一般规律，从而提高模型在未知数据上的表现。下采样模块可以通过不同的方法实现，常见的方法有：

池化操作（Pooling）[17]是一种常用的下采样技术，通常应用于卷积神经网络（CNN）中，在卷积层之后进行。它将输入图像或特征图划分为不重叠的区域（比如 2×2 的窗口），然后对每个区域进行汇聚操作。最常见的池化操作有最大池化（Maxpooling）和平均池化（Average Pooling）。在 Maxpooling 中，每个区域内的最大值会被选取作为该区域的代表值；而在 Average Pooling 中，每个区域内的数值会被求取平均值作为代表值。这样可以有效地减少特征图的尺寸，实现对特征的空间降维，同时保留重要的特征信息。跨步卷积（Strided Convolution）[18]也是一种常用的下采样模块，其工作原理是，在进行卷积操作时，卷积核每次滑动的步长大于 1，这就使得输出特征图的尺寸缩小。通过增大步幅可以降低输出特征图的尺寸，从而实现下采样的效果。相比于池化操作，跨步卷积能够更好地保留空间上的细节信息，因为它不会像池化操作那样丢弃一些特征值。跨步卷积已经在深度学习中的一些应用中被证明是有效的，并且在一些网络架构中得到广泛使用，例如在 MobileNet 和 EfficientNet 等轻量级网络中。抗齿卷积（Anti-aliased CNNs）[19]通过将低通滤波器集成到现有的卷积神经网络中，例如最大池化和跨步卷积等操作之前，有效地降低了高频噪声，并提高了图像分类性能。

YOLOv7 模型在下采样的过程中，通常是使用跨步卷积来实现下采样特征，但考虑到摔倒检测类别单一，需要细粒度的识别，虽然跨步卷积可以实现下采样的过程中信息融合，但是跨步卷积同时也会导致目标信息的特征变得模糊。为了解决这个问题，本文将原来的下采样部分替换成鲁棒特征下采样（RFD，robust feature downsampling）[20]，以弥补图像中卷积下采样的不足。

3　基于改进 YOLOv7 的室内摔倒行为检测模型

3.1　下采样模块改进

YOLOv7 模型在下采样的过程中，通常是使用 $k=3$、$s=2$ 的跨步卷积来实现下采样特征。为了避免这类由于步长为 2 的卷积对目标网络所造成的特征丢失，鲁棒特征下

采样方法在下采样期间将初始特征表示复制到两个副本中，分别表示为 X 和 M。随后，使用两种不同的技术方法对这些复制的特征进行降采样，并将结果特征图融合以产生一致且稳健的表示。X 执行卷积下采样，融合特征图的局部信息，提高泛化能力。M 执行最大池化以避免关键信息和细粒度信息丢失。所得到的下采样特征图被拼接并减少到通道的二分之一维度，以增加鲁棒性。通过合并两个特征图，鲁棒特征下采样方法克服了卷积下采样的局限性，并在检测中表现出色，并且可以捕获复杂细节，从而产生更健壮的特征表示。具体模块细节如图 2 所示。

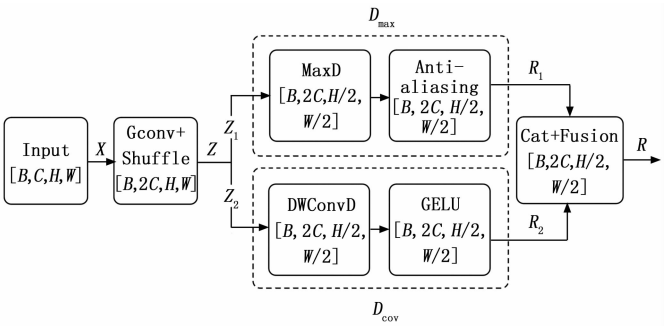


图 2 鲁棒特征下采样（RFD）

输入 X 的特征大小为 $[B, C, H, W]$，其中 B 表示 Batch _ size，C 表示通道数，H 表示特征图的高，W 表示特征图的宽。首先，通过带参数的组卷积（$I = C, O = 2C, K = 3, S = 1, P = 1, G = C$），得到输出特征映射 Y。括号中的参数表示输入通道（I），输出通道（O），卷积核大小（K），步长（S），填充（P），分组数（G）。由于组卷积会导致通道之间的信息交换阻塞。为了解决这个问题，我们执行部分洗牌操作。该操作填补了通道信息交换的空白，仅微量增加计算量。这一措施已被证明可以提高准确性。部分洗牌运算将特征映射 Y 划分为 8 个块，记为 $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$。每一个块包含 D 个通道。然后我们将特征映射 y_1 和 y_5 连接起来，得到一个输出特征，通过参数（$I = 2D, O = 2D, K = 3, S = 1, P = 1$）卷积，得到 Y_1。我们不再进一步处理其他块，并将剩余块与 Y_1 连接以获得输出特征映射 Z，即 $Z = \{Y_1, y_2, y_3, y_4, y_6, y_7, y_8\}$，其特征大小为 $[B, 2C, H, W]$。

$$Z = Shuffle[GConv(X)] \tag{1}$$

接着创建 Z 的副本记为 Z_1 和 Z_2，并将它们分别输入最大池化下采样（D_{max}）和卷积下采样（D_{cov}）。D_{max} 内核大小为 2×2 和步长为 2，这保留了重要的特征。由于 D_{max} 下采样忽略位置信息导致特征位移。因此，我们添加了 anti-aliasing 操作到 D_{max} 中，有效地缓解了最大池化下采样导致特征位移的问题。在 D_{max} 之后，它特征大小为 $[B, 2C, H/2, W/2]$，记为 R_1。可以表示为：

$$D_{max} = BN\{\text{Anti-aliasing}[maxpooling(Z)]\} \tag{2}$$

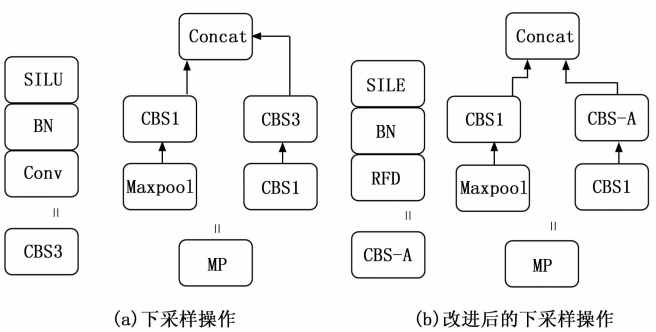
其中：GConv、Shuffle、BN 和 Anti-aliasing 分别表示组卷积、部分洗牌操作、批归一化和抗锯齿操作。Dcov 使用参数为（$I = 2C, O = 2C, K = 3, S = 2, P = 1, G = 2C$）的可分离卷积，即步长为 2 的深度可分离卷积，然后是 GELU 激活。经过 Dcov 后，特征尺寸为 $[B, 2C, H/2, W/2]$，记为 R_2。可以表示为：

$$Dconv = BN(GELU(DWConvD(Z))) \tag{3}$$

然后，我们将 R_1 和 R_2 连接起来，通过（$I = 4C, O = 2C, K = 1, S = 1$）的卷积进行融合，得到特征图 R，特征大小 $[B, 2C, H/2, W/2]$：

$$R = fusion(concatenate(R_1, R_2)) \tag{4}$$

RFD 通过融合来自多种下采样方法的特征来提取更鲁棒的下采样特征图，使后续的模块能够更好地捕获关键信息并提高网络的整体性能。在 YOLOv7 网络中，用于下采样的 CBS 模块，它是由一个 $k=3$、$s=2$ 的卷积层，一个归一化层，还有一个激活函数层组成。MP-1，MP-2 模块上的下分支在 $k=1$、$s=1$ 的卷积后连接一个 $k=3$、$s=2$ 的卷积。其中，$k=1$、$s=1$ 的卷积用来特征提取，$k=3$、$s=2$ 的卷积用来下采样，提取到图像中更多的细节信息。然而，当选择 $k=3$、$s=2$ 的卷积时，卷积过程会造成一些细粒度的丢失，从而使得网络特征表示学习的效率降低，本模型将网络中 CBS 和 MP 模块中 $k=3$、$s=2$ 卷积层替换成 RFD 模块，为了与原 CBS 和 MP 模块区分，将新设计的模块命名为 CBS－1，MP'，如图 3 所示。



(a)下采样操作 (b)改进后的下采样操作

图 3 改进后的下采样操作

3.2 CoordAttention 注意力模块

在深度学习中，通过引入注意力机制，神经网络能够自动地学习并选择性地关注当前任务目标更加关键的信息，提高模型的性能和泛化能力[21]。其中，最典型的注意力机制包括空间注意力机制（CBAM）[22] 和通道注意力机制（SENet）[23]。SENet 侧重于图像中每个通道的信息，但忽略了它们在图像中的位置关系。为弥补这一缺失，CBAM 引入了在通道上进行全局池化的方法来考虑位置信息，然而这种方式仅能获取局部信息，无法捕捉到更长范围依赖关系。继 SENet，CBAM 之后，Hou 等人提出的坐标注意力机制（CA，coordattention）[24]，它通过对输入序列的每个坐标进行注意力加权，以增强模型对于空间位置信息的关注。具体来说，CoordAttention 首先将输入序列的每个坐标表示为一个向量，然后通过一个可学习的线性变换得到该坐标的权重，最后将该权重乘以该坐标对应的特征图，得

到加权后的特征图。CA 模块如图 4 所示。

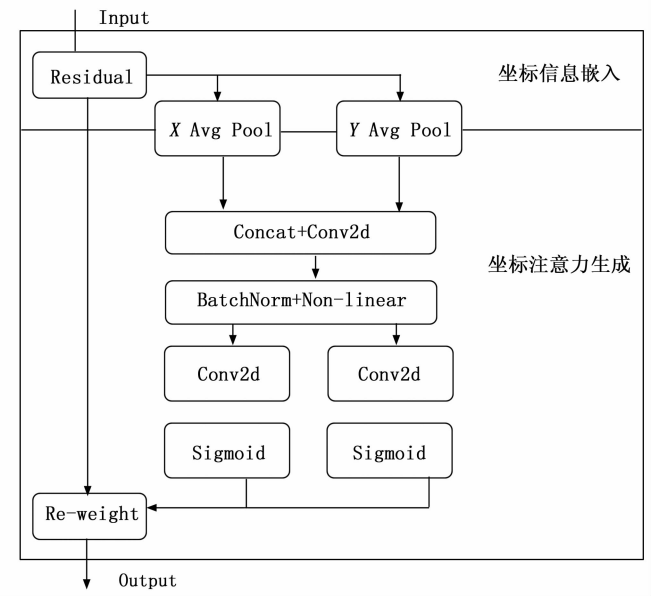


图 4　CA 模块

CA 模块通过准确的位置信息对通道关系和长期依赖性进行编码，具体来说，它包括两个步骤：坐标信息嵌入和坐标注意力生成。

3.2.1　坐标信息嵌入

全局池化方法通常用于对通道注意编码空间信息进行全局编码，但它的缺点是将全局空间信息压缩到通道描述符中，从而难以保留位置信息。为了使注意力模块能够捕捉到具有精确位置信息的远程空间交互，对全局池化进行分解，转化为水平方向和垂直方向的一维特征编码操作，并使用以下计算公式进行处理：

$$ZC = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} xc(i,j) \tag{5}$$

式中，i 为高度，j 为宽度；$xc(i,j)$ 为第 c 个通道的输入信息；Zc 为第 c 个通道的输出信息；特征图的高和宽分别定义为 H 和 W。

对于每个通道，首先使用尺寸为（H，1）和（1，W）的池化核，分别沿着水平坐标和垂直坐标进行编码。因此，高度为 h 的第 c 通道的输出用以下公式表示：

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant i < W} X_c(h,i) \tag{6}$$

同样地，对于宽度为 w 的第 c 通道，可以表示其输出为：

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leqslant j < H} X_c(j,w) \tag{7}$$

3.2.2　坐标注意力生成

经过上述坐标信息嵌入的步骤，可以更准确地获得全局感受野，并且更精确地编码位置信息。为了充分利用这种表征，模型引入坐标注意力生成方法，以便能够准确捕捉感兴趣区域，并有效地捕捉通道之间的关系。

在信息嵌入的变换之后，将上面的变换进行沿空间维数进行 concatenate 操作，然后使用一个 1×1 卷积变换函数 F_1 对其进行压缩操作，δ 为非线性激活函数，对水平方向和垂直方向进行编码，得到中间特征映射，公式表示为：

$$f = \delta[F_1(Z^h, Z^w)] \tag{8}$$

然后沿着空间维数，将 f 进行分解操作，分为 f^h，f^w。利用 2 个卷积变换函数分别将它们变换为具有与输入通道数相同的张量，公式如下：

$$g^h = \delta[F_h(f^h)] \tag{9}$$

$$g^w = \delta[F_w(f^w)] \tag{10}$$

这里 δ 为 sigmoid 激活函数，目的是降低模型的复杂性和计算开销，然后对输出进行扩展，分别作为水平和垂直方向的 attention weights，最终 Coordinate Attention Block 的输出 y 可以写成：

$$y(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{11}$$

通过以上计算得到的特征信息利用残差连接与输入特征进行加和操作，更新权重后得到输出结果。YOLOv7 的 concat 操作将来自不同层级的特征图进行拼接，融合了不同尺度和语义的特征信息。本文通过在 concat 处引入 CA 模块，如图 5 所示，可以使得模型对于不同特征的关注程度有所区别，更加强调对于摔倒检测任务而言更为重要和有区分度的特征，使得模型对于关键特征有所强化，提升目标定位精度，并且抑制冗余信息，改善模型的性能和准确度。

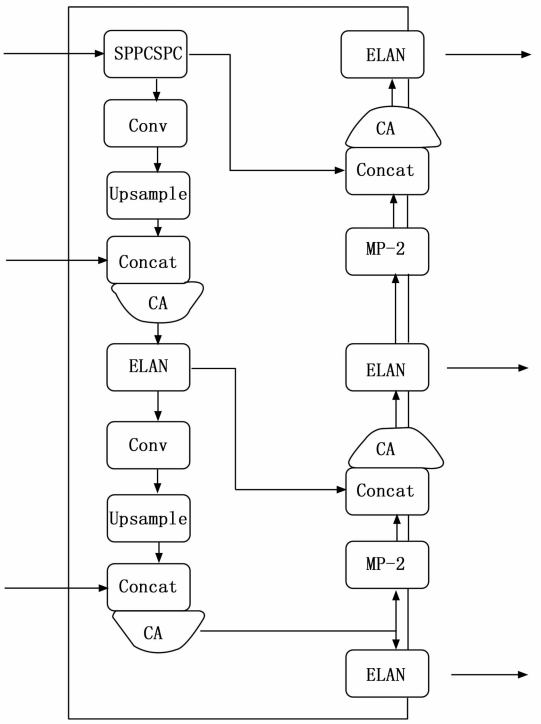


图 5　改进后的 concat 模块

4　实验结果与分析

4.1　实验环境与配置

实验的软件环境和硬件环境如表 1 所示。

表 1 软、硬件实验环境

环境	配置	版本
软件环境	操作系统	Windows 10
	框架	PyTorch1.11.0
	CUDA	CUDA11.7
	PyCharm	Python 3.8
硬件环境	处理器	Intel(R)Core(TM)i9-13900HX CPU
	内存	32G RAM
	显卡	NVIDIA GEFORCE RTX 3090ti
	显存	8 G

实验参数如表 2 所示。

表 2 实验参数配置

训练批次	训练次数	初始学习率	输入图像尺寸	优化器
8	20	0.01	640×640	Adam

4.2 实验数据集

实验采用由 Charfi 等人创建的 Le2i FDD（Le2i Fall Detection Dataset）数据集[25]。该数据集包括了 9 名受试者进行 3 种不同类型的摔倒活动（前向摔倒、平衡丧失、坐着时摔倒）和 6 种常见的日常生活活动（坐着、行走、移动椅子、家务等），共捕获 191 个视频。这些视频是通过单个 RGB 相机拍摄的，并在 4 个不同的环境中进行拍摄，包括家庭、咖啡室、办公室和演讲室。在拍摄过程中，考虑到了各种因素的变化，如光线条件、衣物、衣物颜色、质地、阴影、反射以及相机视角等。数据集示例如图 6 所示。

图 6 数据集示例

为满足实验需求，从中各选用 130 个训练视频和 13 个测试视频。Le2i 数据集标注格式是以列表的形式表示类别，由于摔倒帧标注不准确，需要手动找到每一个视频的摔倒帧进行重新标注，对 Le2i FDD 数据转换成 voc 格式的 xml 文件；在转换文件后，将 voc 格式数据集转换为 yolo 格式。

4.3 评价指标

本实验针对人体摔倒行为检测任务，选择了 3 个常用的指标来评估模型的性能，具体指标如下所示。

精确度（Precision）表示所识别摔倒中真正的摔倒行为所占比例，公式如下：

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{12}$$

召回率（Recall）表示正确识别出来的摔倒行为数量占其总数的比例，公式如下：

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{13}$$

其中：TP 为目标被正确识别的数量，FN 为本类预测为其他类的数量，FP 为错将其他类预测为本类的数量。

平均精度（AP）表示单个类别目标识别精度，平均精度均值（mAP）表示 c 个类别目标 AP 的平均值，公式如下：

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{14}$$

$$mAP = \frac{1}{c}\sum_{i=1}^{c} APi \tag{15}$$

其中：P 为精确度，R 为召回率，$mAP50$ 表示将 IoU 阈值取 0.5 时的 mAP 值，$mAP50:95$ 表示在不同阈值（从 0.5～0.95，步长为 0.05）上的平均 mAP。

4.4 不同下采样方法的试验结果

改进模型以 RFD 鲁棒下采样方法替代原始 YOLOv7 网络模型中的下采样模块。为了验证改进的有效性，本文与 3 种常见的下采样方法（Maxpooling、SoftPool 和 LIP）进行对比实验，即分别将这些下采样方法替换原始 YOLOv7 模型中的下采样模块，并进行训练，随后在同一个数据集上进行对比测试。试验结果如表 3 所示，明显表明采用 RFD 的效果最为显著，精度提升了 1.12 个百分点。

表 3 使用不同下采样方法的试验对比 %

模型	$mAP50$	$mAP50:95$	精确度
YOLOv7	98.16	69.05	97.12
YOLOv7+Maxpooling	97.42	67.23	96.62
YOLOv7+SoftPool	98.27	70.32	97.76
YOLOv7+LIP	98.33	71.45	97.84
YOLOv7+RFD	98.57	72.83	98.24

4.5 添加不同注意力机制的试验结果

为了使模型更加精准地识别和定位图中的目标，更加强调对于摔倒检测任务而言更为重要和有区分度的特征，本研究在改进原始 YOLOv7 网络模型中下采样模块的基础上，尝试在网络的 concat 处添加不同的注意力机制（SE、CBAM），并进行对比测试。测试结果如表 4 所示。结果表明，CA 模块添加到网络的 concat 处检测效果最好。

表 4 添加不同注意力机制的试验对比 %

模型	$mAP50$	$mAP50:95$	精确度
YOLOv7+RFD	98.57	72.83	98.24
YOLOv7+RFD+SE	98.68	73.62	98.31
YOLOv7+RFD+CBAM	98.77	73.91	98.74
YOLOv7+RFD+CA	98.83	74.12	98.88

4.6 消融实验

为了验证各改进模块在 YOLOv7 网络模型中的有效性，采用原始 YOLOv7 的实验结果作为基准进行消融实验，以

评估各个改进模块对本文检测算法的影响，具体的实验数据如表 5 所示。

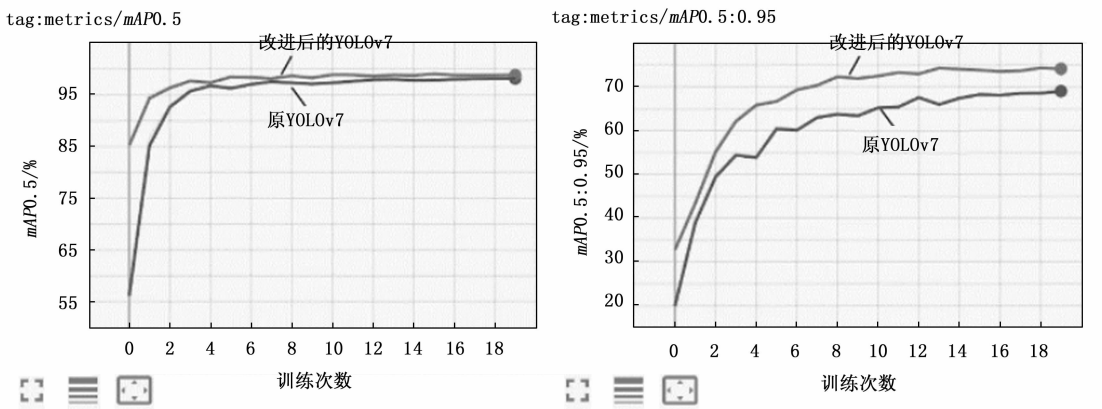
表 5　消融实验结果　　%

改进网络结构	*mAP*50	*mAP*50∶95	精确度
YOLOv7	98.16	69.05	97.12
YOLOv7＋RFD	98.57	72.83	98.24
YOLOv7＋RFD＋CA	98.83	74.12	98.88

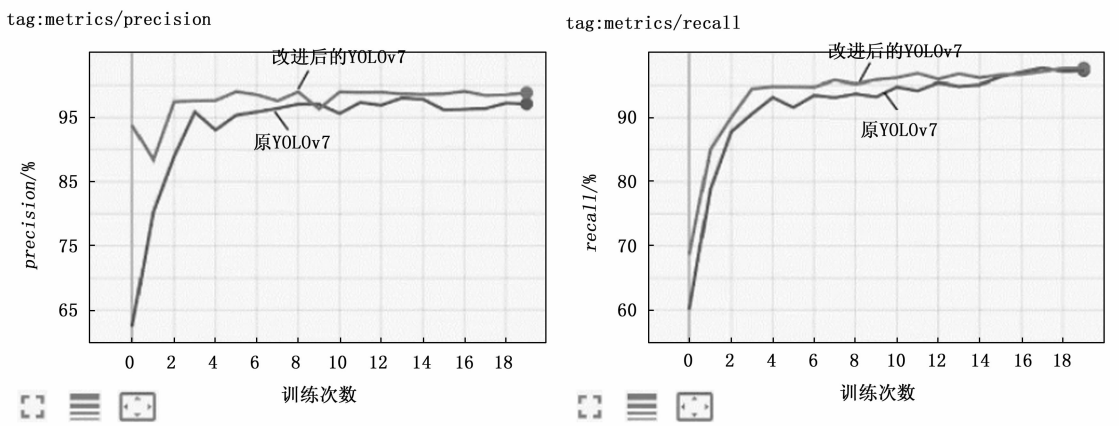
从表 5 中可以看出，原始 YOLOv7 网络模型在 Le2i FDD 数据集上的 *mAP*50 值达到 98.16%，*mAP*50∶95 值达到 69.05%以及精确度达到 97.12%，通过逐步增加改进的两个模块，观察到各评价指标在人体摔倒检测任务中基本上都有所提升。这表明各个模块对于改进算法的有效性都具有积极的影响，能够提升检测任务的准确性和性能。首先将 YOLOv7 网络模型中 CBS 和 MP 模块中 $k=3$、$s=2$ 卷积层替换成 RFD 模块，*mAP*50 值从 98.16%提升至 98.57%，*mAP*50∶95 值提升了 3.78 个百分点以及精确度提升了 1.12 个百分点，证明改进后模型会提取和保留更多的特征信息。其次在网络的 concat 处加入 CoordAttention 注意力机制后，*mAP*50 值提升了 0.26 个百分点，*mAP*50∶95 值提升了 1.29 个百分点以及精确度提升至 98.88%，证明改进后可以更好地融合拼接后的特征图，提升目标定位精度，并且抑制冗余信息，改善模型的性能和准确度。

4.7　实验结果分析

为了直观地说明本文改进模型的有效性，选取了室内环境下不同姿态对模型进行试验，正常行为活动标记为 up，摔倒行为标记为 down。本文模型和 YOLOv7 模型在 Le2i FDD 数据集上进行训练，模型训练结束后，在测试集上对两种模型分别进行测试。图 7（a）为 YOLOv7 模型和本文模型针对 *mAP*50，*mAP*50∶95 的检测结果，图 7（b）为 YOLOv7 模型和本文模型针对 *Precision*，*Recall* 的检测结果。横坐标为训练轮次（epoch），纵坐标为不同的评价指标，得到精度均值 *AP*、召回率 *Recall* 和精确率 *Precision* 的值如表 6 所示。由表可知，两种模型均可对室内环境下的摔倒行为进行检测，但本文模型的各项评价指标值均高于 YOLOv7 模型，精确度达到 98.88%，对摔倒行为的检测更为准确。



(a) *mAP*50, *mAP*50∶95对比结果



(b) *Precision*, *Recall*对比结果

图 7　YOLOv7 模型和本文模型针对不同评价指标的检测结果

表 6 YOLOv7 模型和改进 YOLOv7 模型检测结果

图片尺寸	模型	*mAP*50 /%	*mAP*50-95 /%	召回率 /%	精确度 /%
640×640	YOLOv7	98.16	69.05	97.2	97.12
	本文算法	98.83	74.12	97.59	98.88

图 8 展示了部分可视化结果。上述实验的检测置信度，依次为 0.98，0.99，0.97。检测结果表明本文提出改进后 YOLOv7 网络模型在室内环境下的摔倒行为检测方面具有较好的性能，有望应用于实际场景中，为独居老人提供准确的摔倒检测服务。

(a) 原YOLOv7检测结果

(b) 改进的YOLOv7检测结果

图 8 改进的 YOLOv7 与原 YOLOv7 网络检测结果对比

5 结束语

本文优化了 YOLOv7 的下采样模块，通过采用鲁棒特征下采样，在提取丰富特征的同时，改善下采样过程中目标信息特征的清晰度，提高摔倒检测的准确性。此外，通过在网络的 Concat 部分引入 CoordAttention 注意力机制，更好地融合拼接后的特征图，实现了模型检测速度的提升。实验结果表明，本文改进的 YOLOv7 模型可以准确检测视频中的摔倒行为，实时性较好。下一步将考虑如何在满足实时性和准确性的前提下，降低模型的参数量和计算量，使得能部署到边缘终端设备上，以便植入到家用摄像机中。

参考文献：

[1] TANWAR R, NANDAL N, ZAMANI M, et al. Pathway of trends and technologies in fall detection: a systematic review [J]. Healthcare. MDPI, 2022, 10 (1): 172.

[2] MONTERO-ODASSO M, VAN DER VELDE N, MARTIN F C, et al. World guidelines for falls prevention and management for older adults: a global initiative [J]. Age and Ageing, 2022, 51 (9): 1-36.

[3] LIANJ Z, ZHENGYC, CHENLIN T. Review of fall detection method based on wearable devices [J]. Computer Engineering and Applications, 2019, 55 (18): 8-14.

[4] REN S, HE K, GIRSHICK R, et al. Faster r-CNN: towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.

[5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //European Conference on Computer Vision, 2016: 21-37.

[7] 高正中，孙 健，张天航，等. 基于改进 YOLOv3 的室内摔倒行为检测算法 [J]. 实验技术与管理，2022，39 (11)：132-137.

[8] YU Y, WANG M, WANG Z, et al. Surface defect detection of hight-speed railway hub based on improved YOLOv3 algorithm [C] //2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). Chongqing, China, 2021: 1386-1390.

[9] CHEN T, DING Z, LI B. Elderly fall detection based on improved YOLOv5s network [J]. IEEE Access, 2022, 10: 91273-91282.

[10] WU LZ, XIALW, QIAN Z, et al. An object detection method of falling person based on optimized YOLOv5s [J]. Journal of Graphics, 2022, 43 (5): 791.

[11] REN J, WANG Z, ZHANG Y, et al. YOLOv5-R: lightweight real-time detection based on improved YOLOv5 [J]. Journal of Electronic Imaging, 2022, 31 (3): 033033.

[12] SHI D, ZHU W, CHENG R, et al. Human fall detection algorithm based on YoloX-s and lightweight OpenPose [C] //2022 15th International Conference on Advanced Computer Theory and Engineering (ICACTE). IEEE, 2022: 23-28.

[13] LIU B, HUANG J, LIN S, et al. Improved YOLOX-S abnormal condition detection for power transmission line corridors [C] //2021 IEEE 3rd International Conference on Power Data Science (ICPDS). IEEE, 2021: 13-16.

[14] 尹志成，徐熙平，孙也尧. 基于 OpenPose 的摔倒行为检测技术研究 [J]. 长春理工大学学报（自然科学版），2021，44 (3)：15-21.

[15] 戚玲珑，高建瓴. 基于改进 YOLOv7 的小目标检测 [J]. 计算机工程，2023，49 (1)：41-48.

[16] 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 [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[17] SUN M, SONG Z, JIANG X, et al. Learning pooling for convolutional neural network [J]. Neurocomputing, 2017, 224: 96-104.

（下转第 87 页）