



(12) 发明专利申请

(10) 申请公布号 CN 114973122 A

(43) 申请公布日 2022. 08. 30

(21) 申请号 202210467457.8

G06T 3/40 (2006.01)

(22) 申请日 2022.04.29

G06N 3/04 (2006.01)

(71) 申请人 韶关学院

地址 512005 广东省韶关市浈江区大学路
288号

(72) 发明人 郑楚伟 林辉 韩竺秦

(74) 专利代理机构 广州骏思知识产权代理有限公司 44425

专利代理师 程毅

(51) Int. Cl.

G06V 20/52 (2022.01)

G06V 10/82 (2022.01)

G06V 10/80 (2022.01)

G06V 10/774 (2022.01)

G06V 10/762 (2022.01)

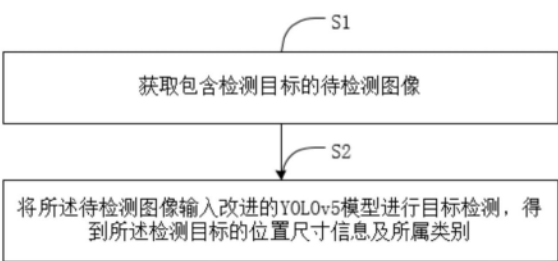
权利要求书5页 说明书14页 附图7页

(54) 发明名称

基于改进YOLOv5的安全帽佩戴检测方法

(57) 摘要

本发明涉及一种基于改进YOLOv5的安全帽佩戴检测方法,包括步骤:获取包含检测目标的待检测图像;将所述待检测图像输入改进的YOLOv5模型进行目标检测,得到所述检测目标的位置尺寸信息及所属类别。相对于现有技术,本发明提供一种基于改进YOLOv5的安全帽佩戴检测方法,通过多层的Swin transformer Block网络对特征进行提取,增强了模型对待检测图像的特征提取能力,通过自注意力机制提取特征,能够获得对用于识别检测目标贡献大的图像特征,并且通过多层次的特征提取,能够获得更丰富的图像特征,从而能够识别受到遮挡的检测目标以及亮度较低的检测目标,此外还能够辨别与检测目标形状相似的物体,误检漏检率低,模型的检测准确度高。



1. 一种基于改进YOLOv5的安全帽佩戴检测方法,其特征在于,包括步骤:

获取包含检测目标的待检测图像;

将所述待检测图像输入改进的YOLOv5模型进行目标检测,得到所述检测目标的位置尺寸信息及所属类别,其中,所述改进的YOLOv5模型包括特征提取模块、特征融合模块和结果预测模块,所述特征提取模块包括图块分割子模块、线性嵌入子模块、第一Swin-T子模块、第一图块拼接子模块、第二Swin-T子模块、第二图块拼接子模块、第三Swin-T子模块、第三图块拼接子模块、第四Swin-T子模块、第四图块拼接子模块和第五Swin-T子模块,所述特征提取模块对待检测图像的特征提取时,包括步骤:

将所述待检测图像输入所述图像分割子模块进行图像分割;

将分割后的所述待检测图像输入所述线性嵌入子模块进行线性变换;

将线性变换后的所述待检测图像输入所述第一Swin-T子模块进行特征提取,得到第一Swin-T特征图;

将所述第一Swin-T特征图输入所述第一图块拼接子模块进行下采样,得到第一层次特征图;

将所述第一层次特征图输入所述第二Swin-T子模块进行特征提取,得到第二Swin-T特征图;

将所述第二Swin-T特征图输入所述第二图块拼接子模块进行下采样,得到第二层次特征图;

将所述第二层次特征图输入所述第三Swin-T子模块进行特征提取,得到第三Swin-T特征图;

将所述第三Swin-T特征图输入所述第三图块拼接子模块进行下采样,得到第三层次特征图;

将所述第三层次特征图输入所述第四Swin-T子模块进行特征提取,得到第四Swin-T特征图;

将所述第四Swin-T特征图输入所述第四图块拼接子模块进行下采样,得到第四层次特征图;

将所述第四层次特征图输入所述第五Swin-T子模块进行特征提取,得到第五Swin-T特征图;

其中,所述第一Swin-T子模块、第二Swin-T子模块、第四Swin-T子模块和第五Swin-T子模块均包括两个Swin transformer Block网络,第三Swin-T子模块包括六个所述Swin transformer Block网络,所述Swin transformer Block网络用于对输入的特征图进行图像特征提取;

所述特征融合模块用于根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图;

所述结果预测模块用于根据多个不同网格尺寸的输出特征图预测得到检测目标的位置尺寸信息及所属类别。

2. 根据权利要求1所述的方法,其特征在于:所述Swin transformer Block网络包括四个LayerNorm层、一个多头自注意力层、两个MLP层、一个移位窗口多头自注意力层、四个DropPath层和四个残差连接层,所述Swin transformer Block网络对输入的特征图进行图

像特征提取时,包括步骤:

将所述输入的特征图输入所述LayerNorm层进行归一化处理;将归一化的所述输入的特征图输入所述多头自注意力层进行多头自注意力特征提取,得到多头自注意力特征图;将所述多头自注意力特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述多头自注意力特征图与所述输入的特征图输入所述残差连接层进行残差连接,得到第一中间特征图;

将所述第一中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第一中间特征图输入所述MLP层进行线性变换,得到第一变换特征图;将所述第一变换特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述第一变换特征图与所述第一中间特征图输入所述残差连接层进行残差连接,得到第二中间特征图;

将所述第二中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第二中间特征图输入所述移位窗口多头自注意力层进行像素移位的多头自注意力特征提取,得到移位多头自注意力特征图;将所述移位多头自注意力特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述移位多头自注意力特征图与所述第二中间特征图输入所述残差连接层进行残差连接,得到第三中间特征图;

将所述第三中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第三中间特征图输入所述MLP层进行线性变换,得到第二变换特征图;将所述第二变换特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述第二变换特征图与所述第三中间特征图输入所述残差连接层进行残差连接,得到作为所述Swin transformer Block网络输出的特征图。

3. 根据权利要求1所述的方法,其特征在于:所述第一图块拼接子模块、第二图块拼接子模块、第三图块拼接子模块和第四图块拼接子模块均包括图块分割层、concat层、LayerNorm层和全连接层,其中,所述图块分割层用于将输入的维度为[H,W,C]特征图中间隔为2的相邻像素划分为多个图块;所述concat层用于将分割的所述图块进行concat拼接,得到维度变为[H/2,W/2,4C]的特征图;所述LayerNorm层用于将所述concat层输出的特征图进行归一化;所述全连接层用于将所述LayerNorm层输出的特征图的通道数进行线性变换,得到维度为[H/2,W/2,2C]的特征图。

4. 根据权利要求1所述的方法,其特征在于:所述特征融合模块包括第一CONV层、第一UP层、第一Concat层、第一C3-Ghost层、第二CONV层、第二UP层、第二Concat层、第二C3-Ghost层、第三CONV层、第三UP层、第三Concat层、第三C3-Ghost层、第四CONV层、第四Concat层、第四C3-Ghost层、第五CONV层、第五Concat层、第五C3-Ghost层、第六CONV层、第六Concat层和第六C3-Ghost层,所述特征融合模块根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图时,包括步骤:

获取所述第五Swin-T特征图并输入所述第一CONV层进行卷积处理,得到第一卷积特征图;将所述第一卷积特征图输入所述第一UP层进行上采样操作;获取所述第四Swin-T特征图并与所述第一UP层输出的特征图共同输入所述第一Concat层进行Concat拼接;将所述第一Concat层输出的特征图输入所述第一C3-Ghost层进行卷积处理,得到第一输出特征图;

将所述第一输出特征图输入所述第二CONV层进行卷积处理,得到第二卷积特征图;将

所述第二卷积特征图输入所述第二UP层进行上采样操作;获取所述第三Swin-T特征图并与所述第二UP层输出的特征图共同输入所述第二Concat层进行Concat拼接;将所述第二Concat层输出的特征图输入所述第二C3-Ghost层进行卷积处理,得到第二输出特征图;

将所述第二输出特征图输入所述第三CONV层进行卷积处理,得到第三卷积特征图;将所述第三卷积特征图输入所述第三UP层进行上采样操作;获取所述第二Swin-T特征图并与所述第三UP层输出的特征图共同输入所述第三Concat层进行Concat拼接;将所述第三Concat层输出的特征图输入所述第三C3-Ghost层进行卷积处理,得到第三输出特征图;

将所述第三输出特征图输入所述第四CONV层进行卷积处理,得到第四卷积特征图;将所述第四卷积特征图与所述第三卷积特征图共同输入所述第四Concat层进行Concat拼接;将所述第四Concat层输出的特征图输入所述第四C3-Ghost层进行卷积处理,得到所述第四输出特征图;

将所述第四输出特征图输入所述第五CONV层进行卷积处理,得到第五卷积特征图;将所述第五卷积特征图与所述第二卷积特征图共同输入所述第五Concat层进行Concat拼接;将所述第五Concat层输出的特征图输入所述第五C3-Ghost层进行卷积处理,得到第五输出特征图;

将所述第五输出特征图输入所述第六CONV层进行卷积处理,得到第六卷积特征图;将所述第六卷积特征图与所述第一卷积特征图共同输入所述第六Concat层进行Concat拼接;将所述第六Concat层输出的特征图输入所述第六C3-Ghost层进行卷积处理,得到第六输出特征图。

5. 根据权利要求1所述的方法,其特征在于:所述特征融合模块包括第一CONV层、第一UP层、第一Concat层、第一C3-Ghost层、第二CONV层、第二UP层、第二Concat层、第二C3-Ghost层、第三CONV层、第三UP层、第三Concat层、第三C3-Ghost层、第四CONV层、第四Concat层、第四C3-Ghost层、第五CONV层、第五Concat层、第五C3-Ghost层、第六CONV层、第六Concat层和第六C3-Ghost层,所述特征融合模块根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图时,包括步骤:

获取所述第五Swin-T特征图并输入所述第一CONV层进行卷积处理,得到第一卷积特征图;将所述第一卷积特征图输入所述第一UP层进行上采样操作;获取所述第四Swin-T特征图并与所述第一UP层输出的特征图共同输入所述第一Concat层进行Concat拼接;将所述第一Concat层输出的特征图输入所述第一C3-Ghost层进行卷积处理,得到第一输出特征图;

将所述第一输出特征图输入所述第二CONV层进行卷积处理,得到第二卷积特征图;将所述第二卷积特征图输入所述第二UP层进行上采样操作;获取所述第三Swin-T特征图并与所述第二UP层输出的特征图共同输入所述第二Concat层进行Concat拼接;将所述第二Concat层输出的特征图输入所述第二C3-Ghost层进行卷积处理,得到第二输出特征图;

将所述第二输出特征图输入所述第三CONV层进行卷积处理,得到第三卷积特征图;将所述第三卷积特征图输入所述第三UP层进行上采样操作;获取所述第二Swin-T特征图并与所述第三UP层输出的特征图共同输入所述第三Concat层进行Concat拼接;将所述第三Concat层输出的特征图输入所述第三C3-Ghost层进行卷积处理,得到第三输出特征图;

将所述第三输出特征图输入所述第四CONV层进行卷积处理,得到第四卷积特征图;将

所述第三Swin-T特征图、第四卷积特征图与第三卷积特征图共同输入所述第四Concat层进行Concat拼接;将所述第四Concat层输出的特征图输入所述第四C3-Ghost层进行卷积处理,得到所述第四输出特征图;

将所述第四输出特征图输入所述第五CONV层进行卷积处理,得到第五卷积特征图;将所述第四Swin-T特征图、第五卷积特征图与第二卷积特征图共同输入所述第五Concat层进行Concat拼接;将所述第五Concat层输出的特征图输入所述第五C3-Ghost层进行卷积处理,得到第五输出特征图;

将所述第五输出特征图输入所述第六CONV层进行卷积处理,得到第六卷积特征图;将所述第六卷积特征图与所述第一卷积特征图共同输入所述第六Concat层进行Concat拼接;将所述第六Concat层输出的特征图输入所述第六C3-Ghost层进行卷积处理,得到第六输出特征图。

6. 根据权利要求4-5中任一项所述的方法,其特征在于:所述第一C3-Ghost层、第二C3-Ghost层、第三C3-Ghost层、第四C3-Ghost层、第五C3-Ghost层、第六C3-Ghost层对输入的特征图进行卷积处理时,包括步骤:

对输入的特征图进行标准卷积操作,使通道数量压缩,并通过N个串联的Ghost Bottleneck模块进行特征提取,得到第一C3-Ghost特征图;

对所述输入的特征图进行另一标准卷积操作,得到第二C3-Ghost特征图;

将所述第一C3-Ghost特征图和所述第二C3-Ghost特征图按通道维度进行Concat叠加,并通过卷积进行特征融合,得到输出的特征图;

其中,所述Ghost Bottleneck模块进行特征提取时的步骤包括:

将所述输入的特征图输入第一层Ghost module模块进行卷积操作,并通过BN层和具有稀疏性的Relu激活函数进行处理;

将经过所述BN层和Relu激活函数处理的特征图输入第二层Ghost module模块进行卷积操作,并通过另一BN层进行处理;

所述Ghost module模块进行卷积操作时的步骤包括:

对输入的特征图通过 1×1 的卷积核进行逐点卷积,并通过缩放因子将所述输入的特征图的通道数压缩,同时通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到浓缩特征图;

对所述浓缩特征图进行逐层卷积操作,再通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到冗余特征图;

将所述浓缩特征图和所述冗余特征图按通道维度进行Concat叠加,输出叠加结果。

7. 根据权利要求6所述的方法,其特征在于:所述缩放因子为2。

8. 根据权利要求4-5中任一项所述的方法,其特征在于,所述结果预测模块根据多个不同网格尺寸的输出特征图预测得到检测目标的位置尺寸信息及所属类别的步骤包括:

在所述第三输出特征图、第四输出特征图、第五输出特征图和第六输出特征图的每个空间点上,分别通过四个对应尺寸的先验锚框进行预测,得到预测的检测目标边框的坐标偏移量 (t_x, t_y) 、宽 t_w 、高 t_h 、概率值和预测类别的置信度;根据所述坐标偏移量 (t_x, t_y) 、宽 t_w 、高 t_h 得到所述检测目标的位置坐标和宽高,所述检测目标的位置坐标 (b_x, b_y) 的表达式为:

$$b_x = \sigma(t_x) + C_x, b_y = \sigma(t_y) + C_y$$

式中, C_x 、 C_y 分别为所述检测目标所在网格的左上角的坐标;

所述检测目标的宽 b_w 、高 b_h 的表达式为:

$$b_w = p_w e^{t_w}, b_h = p_h e^{t_h}$$

式中, p_w 、 p_h 分别为所述先验锚框的宽、高;

经过非极大值抑制处理, 得到最终的所述检测目标的位置坐标、宽、高和预测类别的置信度, 确定所述置信度高的预测类别为对应检测目标的所属类别。

9. 根据权利要求8所述的方法, 其特征在于: 所述先验锚框的尺寸根据用于训练改进的YOLOv5模型的数据集更新, 包括步骤:

一, 通过轮盘赌算法依据概率大小抽选所述数据集中的实际边框作为初始聚类中心;

二, 计算每个所述实际边框与当前聚类中心的距离Loss, Loss距离的表达式为:

$$Loss = \min \sum_{i=1}^n \sum_{j=1}^k \left[1 - \frac{Box_i \cap Center_j}{Box_i \cup Center_j} \right]$$

式中, Box_i 为 n 个实际边框中的第 i 个实际边框的面积, $Center_j$ 为 k 个聚类中心中的第 j 个聚类中心的面积;

三, 将所述实际边框划分至与其具有最短距离的所述聚类中心所属的聚类类别中;

四, 计算每一所述聚类类别中所述实际边框坐标的中位数, 以该中位数更新对应类别的聚类中心, 重复执行步骤二~四, 直至得到 k 个位置稳定的聚类中心, 确定得到的所述位置稳定的聚类中心为先验锚框;

五, 计算每一所述实际边框分别与每一所述先验锚框的尺寸误差度, 获取每一所述实际边框对应的最小尺寸误差度值计算平均值, 确定该平均值为所述先验锚框的适应度, 确定适应度最高的所述先验锚框作为更新的先验锚框。

10. 根据权利要求9所述的方法, 其特征在于, 所述先验锚框的尺寸根据用于训练改进的YOLOv5模型的数据集更新, 还包括步骤六: 对所述更新的先验锚框进行线性变换, 将所述更新的先验锚框的最小宽度变换为0.8倍, 最大宽度变换为1.5倍, 并且保持宽高比例不变。

基于改进YOLOv5的安全帽佩戴检测方法

技术领域

[0001] 本发明涉及安全帽佩戴检测技术领域,尤其是涉及一种基于改进YOLOv5的安全帽佩戴检测方法。

背景技术

[0002] 中国作为世界第一制造业大国,施工现场安全问题是工业界极为关注的一个问题,正确佩戴安全帽能有效地防止或减轻工人在作业过程中发生安全事故对头部的伤害。然而在实际生产活动中,尽管明文要求施工人员一定要佩戴安全帽才能进入施工现场,仍难以杜绝个别人员出于侥幸心理或其他原因,在工作时不规范佩戴安全帽的现象。目前对安全帽佩戴情况的监管主要依赖人工,该方式效率较低且耗费人力,而且人工很难长时间专注监控,会存在一定的疏忽。现有技术中的一种基于YOLOv5的安全帽佩戴检测方法,通过将施工现场的图像输入YOLOv5模型,实现工人是否佩戴安全帽的自动检测,然而,施工现场人员密集,检测目标容易受到遮挡,以及有的施工现场光照环境较暗,这种检测方法难以识别被遮挡及处于暗处的检测目标,导致误检漏检率高。

发明内容

[0003] 本发明的目的在于克服现有技术的缺点与不足,提供一种基于改进YOLOv5的安全帽佩戴检测方法,能够识别被遮挡及尺寸小的检测目标,误检漏检率低,识别准确率高。

[0004] 本发明是通过以下技术方案实现的:一种基于改进YOLOv5的安全帽佩戴检测方法,包括步骤:获取包含检测目标的待检测图像;将所述待检测图像输入改进的YOLOv5模型进行目标检测,得到所述检测目标的位置尺寸信息及所属类别,其中,所述改进的YOLOv5模型包括特征提取模块、特征融合模块和结果预测模块,所述特征提取模块包括图块分割子模块、线性嵌入子模块、第一Swin-T子模块、第一图块拼接子模块、第二Swin-T子模块、第二图块拼接子模块、第三Swin-T子模块、第三图块拼接子模块、第四Swin-T子模块、第四图块拼接子模块和第五Swin-T子模块,所述特征提取模块对待检测图像的特征提取时,包括步骤:将所述待检测图像输入所述图像分割子模块进行图像分割;将分割后的所述待检测图像输入所述线性嵌入子模块进行线性变换;将线性变换后的所述待检测图像输入所述第一Swin-T子模块进行特征提取,得到第一Swin-T特征图;将所述第一Swin-T特征图输入所述第一图块拼接子模块进行下采样,得到第一层次特征图;将所述第一层次特征图输入所述第二Swin-T子模块进行特征提取,得到第二Swin-T特征图;将所述第二Swin-T特征图输入所述第二图块拼接子模块进行下采样,得到第二层次特征图;将所述第二层次特征图输入所述第三Swin-T子模块进行特征提取,得到第三Swin-T特征图;将所述第三Swin-T特征图输入所述第三图块拼接子模块进行下采样,得到第三层次特征图;将所述第三层次特征图输入所述第四Swin-T子模块进行特征提取,得到第四Swin-T特征图;将所述第四Swin-T特征图输入所述第四图块拼接子模块进行下采样,得到第四层次特征图;将所述第四层次特征图输入所述第五Swin-T子模块进行特征提取,得到第五Swin-T特征图;其中,所述第一

Swin-T子模块、第二Swin-T子模块、第四Swin-T子模块和第五Swin-T子模块均包括两个Swin transformer Block网络,第三Swin-T子模块包括六个所述Swin transformer Block网络,所述Swin transformer Block网络用于对输入的特征图进行图像特征提取;所述特征融合模块用于根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图;所述结果预测模块用于根据多个不同网格尺寸的输出特征图预测得到检测目标的位置尺寸信息及所属类别。

[0005] 相对于现有技术,本发明提供一种基于改进YOL0v5的安全帽佩戴检测方法,通过多层的Swin transformer Block网络对特征进行提取,增强了模型对待检测图像的特征提取能力,通过自注意力机制提取特征,能够获得对用于识别检测目标贡献大的图像特征,并且通过多层次的特征提取,能够获得更丰富的图像特征,从而能够识别受到遮挡的检测目标以及亮度较低的检测目标,此外还能够辨别与检测目标形状相似的物体,误检漏检率低,模型的检测准确度高。

[0006] 进一步地,所述Swin transformer Block网络包括四个LayerNorm层、一个多头自注意力层、两个MLP层、一个移位窗口多头自注意力层、四个DropPath层和四个残差连接层,所述Swin transformer Block网络对输入的特征图进行图像特征提取时,包括步骤:将所述输入的特征图输入所述LayerNorm层进行归一化处理;将归一化的所述输入的特征图输入所述多头自注意力层进行多头自注意力特征提取,得到多头自注意力特征图;将所述多头自注意力特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述多头自注意力特征图与所述输入的特征图输入所述残差连接层进行残差连接,得到第一中间特征图;将所述第一中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第一中间特征图输入所述MLP层进行线性变换,得到第一变换特征图;将所述第一变换特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述第一变换特征图与所述第一中间特征图输入所述残差连接层进行残差连接,得到第二中间特征图;将所述第二中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第二中间特征图输入所述移位窗口多头自注意力层进行像素移位的多头自注意力特征提取,得到移位多头自注意力特征图;将所述移位多头自注意力特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述移位多头自注意力特征图与所述第二中间特征图输入所述残差连接层进行残差连接,得到第三中间特征图;将所述第三中间特征图输入所述LayerNorm层进行归一化处理;将归一化的所述第三中间特征图输入所述MLP层进行线性变换,得到第二变换特征图;将所述第二变换特征图输入所述DropPath层进行随机失活;将所述DropPath层输出的所述第二变换特征图与所述第三中间特征图输入所述残差连接层进行残差连接,得到作为所述Swin transformer Block网络输出的特征图。

[0007] 进一步地,所述第一图块拼接子模块、第二图块拼接子模块、第三图块拼接子模块和第四图块拼接子模块均包括图块分割层、concat层、LayerNorm层和全连接层,其中,所述图块分割层用于将输入的维度为 $[H, W, C]$ 特征图中间隔为2的相邻像素划分为多个图块;所述concat层用于将分割的所述图块进行concat拼接,得到维度变为 $[H/2, W/2, 4C]$ 的特征图;所述LayerNorm层用于将所述concat层输出的特征图进行归一化;所述全连接层用于将所述LayerNorm层输出的特征图的通道数进行线性变换,得到维度为 $[H/2, W/2, 2C]$ 的特征图。

[0008] 进一步地,所述特征融合模块包括第一CONV层、第一UP层、第一Concat层、第一C3-Ghost层、第二CONV层、第二UP层、第二Concat层、第二C3-Ghost层、第三CONV层、第三UP层、第三Concat层、第三C3-Ghost层、第四CONV层、第四Concat层、第四C3-Ghost层、第五CONV层、第五Concat层、第五C3-Ghost层、第六CONV层、第六Concat层和第六C3-Ghost层,所述特征融合模块根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图时,包括步骤:获取所述第五Swin-T特征图并输入所述第一CONV层进行卷积处理,得到第一卷积特征图;将所述第一卷积特征图输入所述第一UP层进行上采样操作;获取所述第四Swin-T特征图并与所述第一UP层输出的特征图共同输入所述第一Concat层进行Concat拼接;将所述第一Concat层输出的特征图输入所述第一C3-Ghost层进行卷积处理,得到第一输出特征图;将所述第一输出特征图输入所述第二CONV层进行卷积处理,得到第二卷积特征图;将所述第二卷积特征图输入所述第二UP层进行上采样操作;获取所述第三Swin-T特征图并与所述第二UP层输出的特征图共同输入所述第二Concat层进行Concat拼接;将所述第二Concat层输出的特征图输入所述第二C3-Ghost层进行卷积处理,得到第二输出特征图;将所述第二输出特征图输入所述第三CONV层进行卷积处理,得到第三卷积特征图;将所述第三卷积特征图输入所述第三UP层进行上采样操作;获取所述第二Swin-T特征图并与所述第三UP层输出的特征图共同输入所述第三Concat层进行Concat拼接;将所述第三Concat层输出的特征图输入所述第三C3-Ghost层进行卷积处理,得到第三输出特征图;将所述第三输出特征图输入所述第四CONV层进行卷积处理,得到第四卷积特征图;将所述第四卷积特征图与所述第三卷积特征图共同输入所述第四Concat层进行Concat拼接;将所述第四Concat层输出的特征图输入所述第四C3-Ghost层进行卷积处理,得到所述第四输出特征图;将所述第四输出特征图输入所述第五CONV层进行卷积处理,得到第五卷积特征图;将所述第五卷积特征图与所述第二卷积特征图共同输入所述第五Concat层进行Concat拼接;将所述第五Concat层输出的特征图输入所述第五C3-Ghost层进行卷积处理,得到第五输出特征图;将所述第五输出特征图输入所述第六CONV层进行卷积处理,得到第六卷积特征图;将所述第六卷积特征图与所述第一卷积特征图共同输入所述第六Concat层进行Concat拼接;将所述第六Concat层输出的特征图输入所述第六C3-Ghost层进行卷积处理,得到第六输出特征图。

[0009] 进一步地,所述特征融合模块包括第一CONV层、第一UP层、第一Concat层、第一C3-Ghost层、第二CONV层、第二UP层、第二Concat层、第二C3-Ghost层、第三CONV层、第三UP层、第三Concat层、第三C3-Ghost层、第四CONV层、第四Concat层、第四C3-Ghost层、第五CONV层、第五Concat层、第五C3-Ghost层、第六CONV层、第六Concat层和第六C3-Ghost层,所述特征融合模块根据所述第五Swin-T特征图、第四Swin-T特征图、第三Swin-T特征图、第二Swin-T特征图进行融合,得到多个不同网格尺寸的输出特征图时,包括步骤:获取所述第五Swin-T特征图并输入所述第一CONV层进行卷积处理,得到第一卷积特征图;将所述第一卷积特征图输入所述第一UP层进行上采样操作;获取所述第四Swin-T特征图并与所述第一UP层输出的特征图共同输入所述第一Concat层进行Concat拼接;将所述第一Concat层输出的特征图输入所述第一C3-Ghost层进行卷积处理,得到第一输出特征图;将所述第一输出特征图输入所述第二CONV层进行卷积处理,得到第二卷积特征图;将所述第二卷积特征图输入所述第二UP层进行上采样操作;获取所述第三Swin-T特征图并与所述第二UP层输出的特

征图共同输入所述第二Concat层进行Concat拼接;将所述第二Concat层输出的特征图输入所述第二C3-Ghost层进行卷积处理,得到第二输出特征图;将所述第二输出特征图输入所述第三CONV层进行卷积处理,得到第三卷积特征图;将所述第三卷积特征图输入所述第三UP层进行上采样操作;获取所述第二Swin-T特征图并与所述第三UP层输出的特征图共同输入所述第三Concat层进行Concat拼接;将所述第三Concat层输出的特征图输入所述第三C3-Ghost层进行卷积处理,得到第三输出特征图;将所述第三输出特征图输入所述第四CONV层进行卷积处理,得到第四卷积特征图;将所述第三Swin-T特征图、第四卷积特征图与第三卷积特征图共同输入所述第四Concat层进行Concat拼接;将所述第四Concat层输出的特征图输入所述第四C3-Ghost层进行卷积处理,得到所述第四输出特征图;将所述第四输出特征图输入所述第五CONV层进行卷积处理,得到第五卷积特征图;将所述第四Swin-T特征图、第五卷积特征图与第二卷积特征图共同输入所述第五Concat层进行Concat拼接;将所述第五Concat层输出的特征图输入所述第五C3-Ghost层进行卷积处理,得到第五输出特征图;将所述第五输出特征图输入所述第六CONV层进行卷积处理,得到第六卷积特征图;将所述第六卷积特征图与所述第一卷积特征图共同输入所述第六Concat层进行Concat拼接;将所述第六Concat层输出的特征图输入所述第六C3-Ghost层进行卷积处理,得到第六输出特征图。

[0010] 进一步地,所述第一C3-Ghost层、第二C3-Ghost层、第三C3-Ghost层、第四C3-Ghost层、第五C3-Ghost层、第六C3-Ghost层对输入的特征图进行卷积处理时,包括步骤:对输入的特征图进行标准卷积操作,使通道数量压缩,并通过N个串联的Ghost Bottleneck模块进行特征提取,得到第一C3-Ghost特征图;对所述输入的特征图进行另一标准卷积操作,得到第二C3-Ghost特征图;将所述第一C3-Ghost特征图和所述第二C3-Ghost特征图按通道维度进行Concat叠加,并通过卷积进行特征融合,得到输出的特征图;其中,所述GhostBottleneck模块进行特征提取时的步骤包括:将所述输入的特征图输入第一层Ghost module模块进行卷积操作,并通过BN层和具有稀疏性的Relu激活函数进行处理;将经过所述BN层和Relu激活函数处理的特征图输入第二层Ghost module模块进行卷积操作,并通过另一BN层进行处理;所述Ghost module模块进行卷积操作时的步骤包括:对输入的特征图通过 1×1 的卷积核进行逐点卷积,并通过缩放因子将所述输入的特征图的通道数压缩,同时通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到浓缩特征图;对所述浓缩特征图进行逐层卷积操作,再通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到冗余特征图;将所述浓缩特征图和所述冗余特征图按通道维度进行Concat叠加,输出叠加结果。

[0011] 进一步地,所述缩放因子为2。

[0012] 进一步地,所述结果预测模块根据多个不同网格尺寸的输出特征图预测得到检测目标的位置尺寸信息及所属类别的步骤包括:在所述第三输出特征图、第四输出特征图、第五输出特征图和第六输出特征图的每个空间点上,分别通过四个对应尺寸的先验锚框进行预测,得到预测的检测目标边框的坐标偏移量 (t_x, t_y) 、宽 t_w 、高 t_h 、概率值和预测类别的置信度;根据所述坐标偏移量 (t_x, t_y) 、宽 t_w 、高 t_h 得到检测目标的位置坐标和宽高,检测目标的位置坐标 (b_x, b_y) 的表达式为:

[0013] $b_x = \sigma(t_x) + C_x, b_y = \sigma(t_y) + C_y$

[0014] 式中, C_x 、 C_y 分别为所述检测目标所在网格的左上角的坐标;

[0015] 检测目标的宽 b_w 、高 b_h 的表达式为:

$$[0016] \quad b_w = p_w e^{t_w}, b_h = p_h e^{t_h}$$

[0017] 式中, p_w 、 p_h 分别为所述先验锚框的宽、高;

[0018] 经过非极大值抑制处理, 得到最终的所述检测目标的位置坐标、宽、高和预测类别的置信度, 确定所述置信度高的预测类别为对应检测目标的所属类别。

[0019] 进一步地, 所述先验锚框的尺寸根据用于训练改进的YOLOv5模型的数据集更新, 包括步骤: 一, 通过轮盘赌算法依据概率大小抽选所述数据集中的实际边框作为初始聚类中心; 二, 计算每个所述实际边框与当前聚类中心的距离Loss, Loss距离的表达式为:

$$[0020] \quad Loss = \min \sum_{i=1}^n \sum_{j=1}^k \left[1 - \frac{Box_i \cap Center_j}{Box_i \cup Center_j} \right]$$

[0021] 式中, Box_i 为 n 个实际边框中的第 i 个实际边框的面积, $Center_j$ 为 k 个聚类中心中的第 j 个聚类中心的面积;

[0022] 三, 将所述实际边框划分至与其具有最短距离的所述聚类中心所属的聚类类别中; 四, 计算每一所述聚类类别中所述实际边框坐标的中位数, 以该中位数更新对应类别的聚类中心, 重复执行步骤二~四, 直至得到 k 个位置稳定的聚类中心, 确定得到的所述位置稳定的聚类中心为先验锚框; 五, 计算每一所述实际边框分别与每一所述先验锚框的尺寸误差度, 获取每一所述实际边框对应的最小尺寸误差度值计算平均值, 确定该平均值为所述先验锚框的适应度, 确定适应度最高的所述先验锚框作为更新的先验锚框。

[0023] 进一步地, 所述先验锚框的尺寸根据用于训练改进的YOLOv5模型的数据集更新, 还包括步骤六: 对所述更新的先验锚框进行线性变换, 将所述更新的先验锚框的最小宽度变换为0.8倍, 最大宽度变换为1.5倍, 并且保持宽高比例不变。

[0024] 为了更好地理解和实施, 下面结合附图详细说明本发明。

附图说明

[0025] 图1为实施例基于改进YOLOv5的安全帽佩戴检测方法的流程示意图;

[0026] 图2为图1所示步骤S2中所述改进的YOLOv5模型的结构示意图;

[0027] 图3为实施例Swin transformer Block网络的结构示意图;

[0028] 图4为实施例多头自注意力层的结构示意图;

[0029] 图5为实施例第一多头自注意力模块的流程示意图;

[0030] 图6为实施例MLP层的结构示意图;

[0031] 图7为实施例移位窗口多头自注意力层的结构示意图;

[0032] 图8为实施例第二多头自注意力模块的流程示意图;

[0033] 图9为实施例第一C3-Ghost层、第二C3-Ghost层、第三C3-Ghost层、第四C3-Ghost层、第五C3-Ghost层、第六C3-Ghost层的算法流程示意图;

[0034] 图10为实验(1)中通过传统YOLOv5模型输出的检测结果图像;

[0035] 图11为实验(1)中改进的YOLOv5模型输出的检测结果图像;

[0036] 图12为实验(2)中通过传统YOLOv5模型输出的检测结果图像;

- [0037] 图13为实验(2)中改进的YOLOv5模型输出的检测结果图像；
[0038] 图14为实验(3)中通过传统YOLOv5模型输出的检测结果图像；
[0039] 图15为实验(3)中改进的YOLOv5模型输出的检测结果图像。

具体实施方式

[0040] 为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

[0041] 应当明确，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。

[0042] 下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反，它们仅是如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0043] 在本申请的描述中，需要理解的是，术语“第一”、“第二”、“第三”等仅用于用于区别类似的对象，而不必用于描述特定的顺序或先后次序，也不能理解为指示或暗示相对重要性。对于本领域的普通技术人员而言，可以根据具体情况理解上述术语在本申请中的具体含义。此外，在本申请的描述中，除非另有说明，“多个”是指两个或两个以上。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0044] 请参阅图1，其为本实施例的基于改进YOLOv5的安全帽佩戴检测方法的流程示意图。该方法包括步骤：

[0045] S1：获取包含检测目标的待检测图像；

[0046] S2：将所述待检测图像输入改进的YOLOv5模型进行目标检测，得到所述检测目标的位置尺寸信息及所属类别。

[0047] 在步骤S1中，所述待检测图像可以通过摄像头在施工现场采集，摄像头采集图像后通过数据线或网络传输到能够执行计算机程序的处理器中，摄像头所采集的图像可以是图片数据或视频流数据，当所采集的图像是视频流数据时，抽取其中的帧图像作为待检测图像。

[0048] 待检测图像还可以是用于训练和测试改进的YOLOv5模型的数据集，所述数据集中的图像为包含了施工现场人戴安全帽的图像，所述数据集可以是已有公开的安全帽开源数据集，此外为了提高模型的识别效果，还可以通过网络爬虫或实地采集施工现场不同光照环境下的人戴安全帽图像来扩充已有公开的安全帽开源数据集，其中，使用LabelIMG标注工具制作通过网络爬虫或实地采集的人戴安全帽图像数据集的标签，包括安全帽所在区域的实际边框。本实施例将数据集按照9:1的比例随机划分为训练集和测试集。

[0049] 待检测图像中的检测目标可以是人戴安全帽的图像、未戴安全帽人头的图像等需要被识别的图像信息。

[0050] 请参阅图2，其为步骤S2中所述改进的YOLOv5模型的结构示意图，该模型包括特征提取模块、特征融合模块和结果预测模块，其中特征提取模块用于提取待检测图像中的图

像特征;特征融合模块用于将特征提取模块所提取的图像特征进行融合,并输出多个不同网格尺寸的输出特征图;结果预测模块用于根据多个不同网格尺寸的输出特征图预测得到检测目标的位置尺寸信息及所属类别。

[0051] 具体的,特征提取模块包括图块分割(Patch Partition)子模块、线性嵌入(Linear Embedding)子模块、第一Swin-T(Swin transformer)子模块、第一图块拼接(Patch Merging)子模块、第二Swin-T子模块、第二图块拼接子模块、第三Swin-T子模块、第三图块拼接子模块、第四Swin-T子模块、第四图块拼接子模块和第五Swin-T子模块,特征提取模块对待检测图像的特征提取时,包括步骤:

[0052] 将待检测图像输入图像分割子模块进行图像分割;

[0053] 将分割后的待检测图像输入线性嵌入子模块进行线性变换;

[0054] 将线性变换后的待检测图像输入第一Swin-T子模块进行特征提取,得到第一Swin-T特征图;

[0055] 将第一Swin-T特征图输入第一图块拼接子模块进行下采样,得到第一层次特征图;

[0056] 将第一层次特征图输入第二Swin-T子模块进行特征提取,得到第二Swin-T特征图;

[0057] 将第二Swin-T特征图输入第二图块拼接子模块进行下采样,得到第二层次特征图;

[0058] 将第二层次特征图输入第三Swin-T子模块进行特征提取,得到第三Swin-T特征图;

[0059] 将第三Swin-T特征图输入第三图块拼接子模块进行下采样,得到第三层次特征图;

[0060] 将第三层次特征图输入第四Swin-T子模块进行特征提取,得到第四Swin-T特征图;

[0061] 将第四Swin-T特征图输入第四图块拼接子模块进行下采样,得到第四层次特征图;

[0062] 将第四层次特征图输入第五Swin-T子模块进行特征提取,得到第五Swin-T特征图。

[0063] 其中,图像分割子模块对待检测图像进行图像分割时,将所输入维度为 $[H, W, CH]$ 的待检测图像中每 $a \times a$ 个相邻像素分割为一个图块,并沿着通道方向展开,得到维度为 $[H/a, W/a, a^2CH]$ 的待检测图像,其中 H 为待检测图像的高, W 为待检测图像的宽, CH 为待检测图像的通道数。在一具体实施中, a 取值为4。

[0064] 线性嵌入子模块对分割后的待检测图像进行线性变换时,对分割后的待检测图像的每个像素的通道数据做线性变换,得到维度为 $[H/a, W/a, C]$ 的待检测图像,其中 C 为调整图像通道数以适应特征融合模块的超参数。在一具体实施中,超参数 C 设置为64。

[0065] 第一Swin-T子模块、第二Swin-T子模块、第四Swin-T子模块和第五Swin-T子模块中均包括两个用于对输入的特征图进行图像特征提取的Swin transformer Block网络,第三Swin-T子模块包括六个所述Swin transformer Block网络。请参阅图3,其为Swin transformer Block网络的结构示意图,Swin transformer Block网络包括四个LayerNorm

层、一个多头自注意力 (W-MSA) 层、两个MLP层、一个移位窗口多头自注意力 (SW-MSA) 层、四个DropPath层和四个残差连接层。Swin transformer Block网络进行图像特征提取时, 包括步骤:

[0066] 将输入特征图 Z^{l-1} 输入LayerNorm层进行归一化处理; 将归一化的输入特征图 Z^{l-1} 输入多头自注意力层进行多头自注意力特征提取, 得到多头自注意力特征图; 对多头自注意力特征图输入DropPath层, 使Swin Transformer Block中的多分支路径随机失活, 其作为一种正则化策略, 以此提高模型的泛化能力, 防止过拟合; 将DropPath层输出的多头自注意力特征图与特征图 Z^{l-1} 输入残差连接层进行残差连接, 得到第一中间特征图 \hat{Z}^l ;

[0067] 将残差连接得到第一中间特征图 \hat{Z}^l 输入LayerNorm层进行归一化处理; 将归一化的第一中间特征图 \hat{Z}^l 输入MLP层进行线性变换, 得到第一变换特征图; 将第一变换特征图输入DropPath层进行随机失活; 将DropPath层输出的第一变换特征图与第一中间特征图 \hat{Z}^l 输入残差连接层进行残差连接, 得到第二中间特征图 Z^l ;

[0068] 将第二中间特征图 Z^l 输入LayerNorm层进行归一化处理; 将归一化的第二中间特征图 Z^l 输入移位窗口多头自注意力层进行像素移位的多头自注意力特征提取, 得到移位多头自注意力特征图; 将移位多头自注意力特征图输入DropPath层进行随机失活; 将DropPath层输出的移位多头自注意力特征图与第二中间特征图 Z^l 输入残差连接层进行残差连接, 得到第三中间特征图 \hat{Z}^{l+1} ;

[0069] 将第三中间特征图 \hat{Z}^{l+1} 输入LayerNorm层进行归一化处理; 将归一化的第三中间特征图 \hat{Z}^{l+1} 输入MLP层进行线性变换, 得到第二变换特征图; 将第二变换特征图输入DropPath层进行随机失活; 将DropPath层输出的第二变换特征图与第三中间特征图 \hat{Z}^{l+1} 输入残差连接层进行残差连接, 得到输出特征图 Z^{l+1} 。

[0070] 更具体的, 请参阅图4, 其为多头自注意力层的结构示意图, 多头自注意力层包括第一窗口分割 (Window Partition) 模块、第一多头自注意力 (Multi-Head Self-Attention) 模块和第一窗口重组 (Window Reverse) 模块, 其中第一窗口分割模块用于将输入的特征图分割为多个 $M \times M$ 相邻像素的互不重叠的独立窗口, 即将特征图分割为多个图块向量, 以将第一多头自注意力模块的计算限制在每个独立窗口内部, 从而减少计算量。

[0071] 第一多头自注意力模块用于对每个独立窗口分别进行多头的缩放点积注意力计算, 得到每个独立窗口对应的多头自注意力特征。具体地, 请参阅图5, 其为第一多头自注意力模块的流程示意图, 第一多头自注意力模块中的步骤包括: 对每个独立窗口的图块向量在通道维度进行线性变换, 使通道数增加两倍, 同时在特征维度上分割为 h 个子空间, h 为注意力头 (head) 的个数; 通过 h 个不同的参数矩阵 W^Q 、 W^K 、 W^V 分别在 h 个子空间中对每个像素的查询 Q (Query)、键 K (Key) 和权重 V (Value) 进行线性变换, 并进行缩放点积注意力计算; 将 h 个计算结果输入Concat模块和Linear模块, 通过可学习的权重矩阵 W^0 进行拼接融合, 以联合来自不同子空间中学习到的特征信息, 得到多头自注意力特征。其中, 第 i 个注意力头的缩放点积注意力计算结果 $head_i$ 的表达式为:

[0072] $head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$

[0073] 式中, W_i^Q 为第 i 个参数矩阵 W^Q ; W_i^K 为第 i 个参数矩阵 W^K ; W_i^V 为第 i 个参数矩阵 W^V ; $\text{Attention}()$ 为归一化的缩放点积模型, 其表达式为:

[0074] $Attention(Q, K, V) = SoftMax(QK^T / \sqrt{d} + B)V$

[0075] 式中, $Q, K, V \in R^{M^2 \times d}$, QK^T 为不同像素点进行信息交互的过程, 通过点积来计算不同像素点之间的相似度; d 为一个 query 和 key 向量的维度, 通过除以 \sqrt{d} 进行缩放操作能保证梯度的稳定性; B 为可学习的相对位置编码 (Relative Position Bias), $B \in R^{M^2 \times M^2}$ 。

[0076] 多头自注意力特征的拼接融合表达式为:

[0077] $MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h) W^0$

[0078] 第一窗口重组模块用于对每个独立窗口的多头自注意力特征进行还原拼接, 得到完整的多头自注意力特征图。

[0079] 请参阅图6, 其为MLP层的结构示意图, MLP层包括两个Linear层、一个GELU层、两个Dropout层, MLP层对输入的特征图进行线性变换时, 包括步骤: 将输入的特征图通过第一个Linear层进行线性变换, 得到通道数为原来的四倍的特征图; 将四倍通道数的特征图通过GELU层进行非线性激活函数的激活, 以增加网络模型的非线性; 将GELU层输出的特征图通过第一个Dropout层进行随机失活操作, 以避免网络过度依赖某些局部的特征, 使模型泛化性增强; 将第一个Dropout层输出的特征图通过第二个Linear层进行线性变换, 得到通道数与MLP层输入的特征图相同的特征图; 将第二个Linear层输出的特征图通过第二个Dropout层进行随机失活操作后, 确定为输出的特征图。

[0080] 请参阅图7, 其为移位窗口多头自注意力层的结构示意图, 移位窗口多头自注意力层包括循环移位 (Cyclic Shift) 模块、第二窗口分割模块、第二多头自注意力模块、第二窗口重组模块和移位还原 (Reverse Cyclic Shift) 模块, 其中, 循环移位模块用于将输入的特征图顶部的 $M/2$ 行的像素移动至底部, 将最左边的 $M/2$ 列像素移动至最右边。

[0081] 第二窗口分割模块用于将移位后的特征图分割为多个 $M \times M$ 相邻像素的互不重叠的独立窗口。

[0082] 第二多头自注意力模块用于对每个独立窗口分别进行多头的缩放点积注意力计算, 得到每个独立窗口对应的移位多头自注意力特征。请参阅图8, 其为第二多头自注意力模块的流程示意图, 第二多头自注意力模块中的步骤包括: 对每个独立窗口的图块向量在通道维度进行线性变换, 使通道数增加两倍, 同时在特征维度上分割为 h 个子空间, h 为注意力头 (head) 的个数; 通过 h 个不同的参数矩阵 W^Q 、 W^K 、 W^V 分别在 h 个子空间中对像素点的查询 Q 、键 K 和权重 V 进行线性变换, 并进行缩放点积注意力计算, 计算中加入掩码机制, 将独立窗口中移位前位于不相邻区域的像素点之间的注意力权重系数设置为0, 以隔离输入的特征图中不相邻区域的像素点之间无效的信息交流, 在一具体实施中, 将独立窗口中移位前位于不相邻区域的像素点的相似度结果减去100, 则经过softmax归一化后的结果为0; 将 h 个计算结果输入Concat模块和Linear模块, 通过可学习的权重矩阵 W^0 进行拼接融合, 得到移位多头自注意力特征。

[0083] 第二窗口重组模块用于对每个独立窗口的移位多头自注意力特征进行还原拼接。

[0084] 移位还原模块用于将还原拼接后的特征图的最右边的 $M/2$ 列像素移动至最左边, 将底部的 $M/2$ 行的像素移动至顶部, 以将循环移位后的特征图的像素位置进行还原, 得到移位多头自注意力特征图。

[0085] 第一图块拼接子模块、第二图块拼接子模块、第三图块拼接子模块和第四图块拼

接子模块均包括图块分割层、concat层、LayerNorm层和全连接层。第一图块拼接子模块、第二图块拼接子模块、第三图块拼接子模块和第四图块拼接子模块对输入的特征图进行下采样时,通过图块分割层将输入的维度为 $[H, W, C]$ 特征图中间隔为2的相邻像素划分为多个图块;通过concat层将分割的图块进行concat拼接,以使输入的特征图维度变为 $[H/2, W/2, 4C]$;通过LayerNorm层将特征图进行归一化;通过全连接层将特征图的通道数线性变换,使输入的特征图维度变为 $[H/2, W/2, 2C]$ 。

[0086] 特征融合模块包括第一CONV层、第一UP层、第一Concat层、第一C3-Ghost层、第二CONV层、第二UP层、第二Concat层、第二C3-Ghost层、第三CONV层、第三UP层、第三Concat层、第三C3-Ghost层、第四CONV层、第四Concat层、第四C3-Ghost层、第五CONV层、第五Concat层、第五C3-Ghost层、第六CONV层、第六Concat层和第六C3-Ghost层,特征融合模块对特征提取模块所提取的图像特征进行融合时,包括步骤:

[0087] 获取第五Swin-T特征图并输入第一CONV层进行卷积处理,得到第一卷积特征图;将第一卷积特征图输入第一UP层进行上采样操作;获取第四Swin-T特征图并与第一UP层输出的特征图共同输入第一Concat层进行Concat拼接;将第一Concat层输出的特征图输入第一C3-Ghost层进行卷积处理,得到第一输出特征图;

[0088] 将第一输出特征图输入第二CONV层进行卷积处理,得到第二卷积特征图;将第二卷积特征图输入第二UP层进行上采样操作;获取第三Swin-T特征图并与第二UP层输出的特征图共同输入第二Concat层进行Concat拼接;将第二Concat层输出的特征图输入第二C3-Ghost层进行卷积处理,得到第二输出特征图;

[0089] 将第二输出特征图输入第三CONV层进行卷积处理,得到第三卷积特征图;将第三卷积特征图输入第三UP层进行上采样操作;获取第二Swin-T特征图并与第三UP层输出的特征图共同输入第三Concat层进行Concat拼接;将第三Concat层输出的特征图输入第三C3-Ghost层进行卷积处理,得到第三输出特征图;

[0090] 将第三输出特征图输入第四CONV层进行卷积处理,得到第四卷积特征图;将第四卷积特征图与第三卷积特征图共同输入第四Concat层进行Concat拼接;将第四Concat层输出的特征图输入第四C3-Ghost层进行卷积处理,得到第四输出特征图;

[0091] 将第四输出特征图输入第五CONV层进行卷积处理,得到第五卷积特征图;将第五卷积特征图与第二卷积特征图共同输入第五Concat层进行Concat拼接;将第五Concat层输出的特征图输入第五C3-Ghost层进行卷积处理,得到第五输出特征图;

[0092] 将第五输出特征图输入第六CONV层进行卷积处理,得到第六卷积特征图;将第六卷积特征图与第一卷积特征图共同输入第六Concat层进行Concat拼接;将第六Concat层输出的特征图输入第六C3-Ghost层进行卷积处理,得到第六输出特征图。

[0093] 在具体实施中,第一UP层、第二UP层和第三UP层通过最邻近插值算法进行上采样操作。

[0094] 在一优选的特征融合模块对特征提取模块所提取的图像特征进行融合实施例中,上述步骤的将第四卷积特征图与第三卷积特征图共同输入第四Concat层进行Concat拼接可替换为:将第三Swin-T特征图、第四卷积特征图与第三卷积特征图共同输入第四Concat层进行Concat拼接;上述步骤的将第五卷积特征图与第二卷积特征图共同输入第五Concat层进行Concat拼接可替换为:将第四Swin-T特征图、第五卷积特征图与第二卷积特征图共

同输入第五Concat层进行Concat拼接。在该优选实施例中,通过增加同一层级的原始输入节点和输出节点之间的横向跳跃连接,使同层级上的特征图能够共享彼此的语义信息,能够加强特征融合以提高模型精度。

[0095] 请参阅图9,其为第一C3-Ghost层、第二C3-Ghost层、第三C3-Ghost层、第四C3-Ghost层、第五C3-Ghost层、第六C3-Ghost层的算法流程示意图。第一C3-Ghost层、第二C3-Ghost层、第三C3-Ghost层、第四C3-Ghost层、第五C3-Ghost层、第六C3-Ghost层对输入的特征图进行卷积处理时,包括步骤:

[0096] 对输入的特征图进行标准卷积操作,使通道数量压缩,并通过N个串联的GhostBottleneck模块进行特征提取,得到第一C3-Ghost特征图;

[0097] 同时,对输入的特征图进行另一标准卷积操作,得到第二C3-Ghost特征图;

[0098] 将第一C3-Ghost特征图和第二C3-Ghost特征图按通道维度进行Concat叠加,并通过卷积进行特征融合,得到输出的特征图。

[0099] 更具体的,Ghost Bottleneck模块对输入的特征图进行特征提取时,包括步骤:

[0100] 将输入的特征图输入第一层Ghost module模块进行卷积操作,并通过BN (Batch Normalization) 层和具有稀疏性的Relu激活函数进行处理,其中BN层用于保证每一层网络的输入具有相同的分布,Relu激活函数用于避免反向传播的梯度消失现象;

[0101] 将经过BN层和Relu激活函数处理的特征图输入第二层Ghost module模块进行卷积操作,并通过另一BN层进行处理。此时不使用Relu激活函数是因为ReLU激活函数负半轴存在的硬饱和置0会使其输出数据分布不为零均值而导致神经元失活,从而降低网络的性能。

[0102] 其中Ghost module模块对输入的特征图进行卷积操作时,包括步骤:

[0103] 对输入的特征图通过 1×1 的卷积核进行逐点卷积,再通过缩放因子ratio将输入的特征图的通道数压缩,同时通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到浓缩特征图,在本实施例中,缩放因子ratio为2,将输入的特征图的通道数压缩为原来的一半;

[0104] 对浓缩特征图进行逐层卷积操作,再通过BatchNorm2d层进行归一化操作,并经过SiLU激活函数处理,得到冗余特征图;

[0105] 将浓缩特征图和冗余特征图按通道维度进行Concat叠加,并输出叠加结果。

[0106] 结果预测模块根据不同网格尺寸的第三输出特征图、第四输出特征图、第五输出特征图和第六输出特征图获得检测目标的位置尺寸信息及所属类别,其中网格尺寸最大的输出特征图用于检测尺寸小的目标,网格尺寸最小的特征图用于检测尺寸大的目标,具体包括步骤:

[0107] 在不同网格尺寸的第三输出特征图、第四输出特征图、第五输出特征图和第六输出特征图的每个空间点上,分别通过四个对应尺寸的先验锚框进行预测,得到预测的目标边框的坐标偏移量 (t_x, t_y) 、宽 t_w 、高 t_h 、概率值和预测类别的置信度;根据预测的目标边框的坐标偏移量 (t_x, t_y) 、目标边框的宽 t_w 、高 t_h 得到检测目标的位置坐标和宽高,检测目标的位置坐标 (b_x, b_y) 的表达式为:

[0108] $b_x = \sigma(t_x) + C_x, b_y = \sigma(t_y) + C_y$

[0109] 式中, C_x 、 C_y 分别为检测目标所在网格的左上角于特征图中的坐标,

[0110] 检测目标的宽 b_w 、高 b_h 的表达式为:

$$[0111] \quad b_w = p_w e^{t_w}, b_h = p_h e^{t_h}$$

[0112] 式中, p_w 、 p_h 分别为先验锚框的宽、高;

[0113] 最后经过非极大值抑制处理,得到最终的检测目标的位置坐标、宽高和预测类别的置信度,确定置信度高的预测类别为对应检测目标的所属类别。在一具体实施中,最终的检测目标的位置坐标、宽高和检测目标的所属类别将在待检测图像上进行标记,并作为检测结果图像输出,检测目标的所属类别为正确佩戴安全帽和未佩戴安全帽。

[0114] 在对本实施例的改进的YOLOv5模型进行训练时,将模型输出的待检测图像中的检测目标的位置尺寸信息及所属类别输入损失函数,进行预测数据与实际数据的差距计算,通过损失函数对模型进行参数调整。

[0115] 在一优选实施例中,根据用于训练改进的YOLOv5模型的数据集更新改进的YOLOv5模型中的先验锚框尺寸,具体包括步骤:

[0116] 一,在数据集中通过轮盘赌算法依据概率大小抽选实际边框作为初始聚类中心;

[0117] 二,计算数据集中每个实际边框与当前聚类中心的距离Loss, Loss距离的表达式为:

$$[0118] \quad Loss = \min \sum_{i=1}^n \sum_{j=1}^k \left[1 - \frac{Box_i \cap Center_j}{Box_i \cup Center_j} \right]$$

[0119] 式中, Box_i 为n个实际边框中的第i个实际边框的面积, $Center_j$ 为k个聚类中心中的第j个聚类中心的面积;

[0120] 三,将实际边框划分至与其具有最短距离的聚类中心所属的聚类类别中;

[0121] 四,计算每一聚类类别中实际边框坐标的中位数,以该中位数更新对应类别的聚类中心,重复执行步骤二~四,直至选出k个位置稳定的聚类中心,确定得到的聚类中心为先验锚框;

[0122] 五,计算每一实际边框分别与每一先验锚框的尺寸误差度,获取每一所述实际边框对应的最小尺寸误差度值计算平均值,确定该平均值为先验锚框的适应度,确定适应度最高的先验锚框作为更新的先验锚框;

[0123] 为了更好的发挥改进的YOLOv5模型的多尺度目标检测能力,还包括步骤六:对先验锚框进行线性变换,将先验锚框的最小宽度变换为0.8倍,最大宽度变换为1.5倍,并且保持宽高比例不变。

[0124] 以下为改进的YOLOv5模型与传统YOLOv5模型的检测对比实验:

[0125] (1) 将检测目标被遮挡的待检测图像分别输入传统YOLOv5模型与改进的YOLOv5模型,得到检测结果图像。请参阅图10和图11,其中图10为通过传统YOLOv5模型输出的检测结果图像,图11为改进的YOLOv5模型输出的检测结果图像,可见,改进的YOLOv5模型仍然可以将被遮挡的检测目标检测出来,而传统YOLOv5模型无法检测该被遮挡的检测目标。

[0126] (2) 将包含圆形控制器图像的待检测图像分别输入传统YOLOv5模型与改进的YOLOv5模型,得到检测结果图像。请参阅图12和图13,其中图12为通过传统YOLOv5模型输出的检测结果图像,图13为改进的YOLOv5模型输出的检测结果图像,可见,传统YOLOv5模型将圆形控制器误认为是未佩戴安全帽的人头,而改进的YOLOv5模型可以分辨图像中的圆形控

制器不是人头,具有更高的检测准确度。

[0127] (3) 将弱光照下得到的待检测图像分别输入传统YOLOv5模型与改进的YOLOv5模型,得到检测结果图像。请参阅图14和15,其中图14为通过传统YOLOv5模型输出的检测结果图像,图15为改进的YOLOv5模型输出的检测结果图像,可见,传统YOLOv5模型存在漏检情况,而改进的YOLOv5模型可以将待检测图像中所有的检测目标识别出来,在光照环境的影响下,仍能保持较高的检测准确度。

[0128] 此外,本申请还对改进的YOLOv5模型进行消融实验,实验结果如表1所示,其中“传统YOLOv5”为具有四个检测尺度的传统YOLOv5模型;“传统YOLOv5+C3-Ghost”为传统YOLOv5模型的C3模块替换为实施例的C3-Ghost模块的模型;“传统YOLOv5+改进的特征融合”为在具有四个检测尺度的传统YOLOv5模型的特征融合模块增加同一层级的原始输入节点和输出节点之间的横向跳跃连接的模型;“传统YOLOv5+C3-Ghost+改进的特征融合”为传统YOLOv5模型的C3模块替换为实施例的C3-Ghost模块的模型且特征融合模块增加同一层级的原始输入节点和输出节点之间的横向跳跃连接的模型;“传统YOLOv5+Swin Transformer”是使用Swin Transformer作为传统YOLOv5的骨干特征提取网络的模型;“传统YOLOv5+Swin Transformer+C3Ghost”是使用Swin Transformer作为传统YOLOv5的骨干特征提取网络并且将传统YOLOv5模型的C3模块替换为实施例的C3-Ghost模块的模型;“改进的YOLOv5”为实施例中改进的YOLOv5模型;P为模型的准确率,R为模型的召回率,mAP@.5为IoU阈值为0.5的情况下每一类别的AP(Average Precision平均精确率)值的平均;mAP@.5:.95表示IoU阈值从0.5开始、以0.05的步长增长到0.95所对应的平均mAP。

[0129] 表1

	模型	[0130]			
		P	R	mAP@.5	mAP@.5:.95
	传统 YOLOv5	0.904	0.862	0.917	0.529
	传统 YOLOv5+C3-Ghost	0.901	0.857	0.918	0.529
	传统 YOLOv5+改进的特征融合	0.915	0.858	0.922	0.535
	传统 YOLOv5+C3-Ghost+改进的特征融合	0.9	0.861	0.918	0.53
	传统 YOLOv5 + Swin Transformer	0.889	0.834	0.9	0.55
	传统 YOLOv5+Swin Transformer+C3Ghost	0.883	0.834	0.896	0.548
	改进的 YOLOv5	0.893	0.83	0.9	0.552

[0131] 其中传统YOLOv5模型的参数量为 7.17×10^6 ,传统YOLOv5+C3-Ghost的参数量为 6.14×10^6 ,可见,在保持mAP@.5值几乎不变的情况下,传统YOLOv5+C3-Ghost的参数量相比于传统YOLOv5减少了14.4%,证明本实施例的C3-Ghost模块能有效减少模型参数和计算复杂度。传统YOLOv5+改进的特征融合的mAP@.5值相比传统YOLOv5提高了0.5%。传统YOLOv5+Swin Transformer模型相比传统YOLOv5模型,在mAP@.5:.95指标上提升了2.1%,传统YOLOv5+Swin Transformer+C3Ghost模型,在mAP@.5:.95指标上提升了1.9%。改进YOLOv5网络模型由于基于Swin Transformer的特征提取而具有更高的特征提取能力,同时,既有C3-Ghost模块带来的计算轻便性,又有改进的特征融合带来的高准确率,从表1中可以看出本实施例改进YOLOv5模型相比传统YOLOv5模型,在mAP@.5:.95指标上提升了2.3%,即明显

具有更高的检测准确度。

[0132] 本申请可采用在一个或多个其中包含有程序代码的存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。计算机可用存储介质包括永久性和非永久性、可移动和非可移动媒体,可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括但不限于:相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。

[0133] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0134] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0135] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0136] 以上所述实施例仅表达了本发明的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形和改进,则本发明也意图包含这些改动和变形。

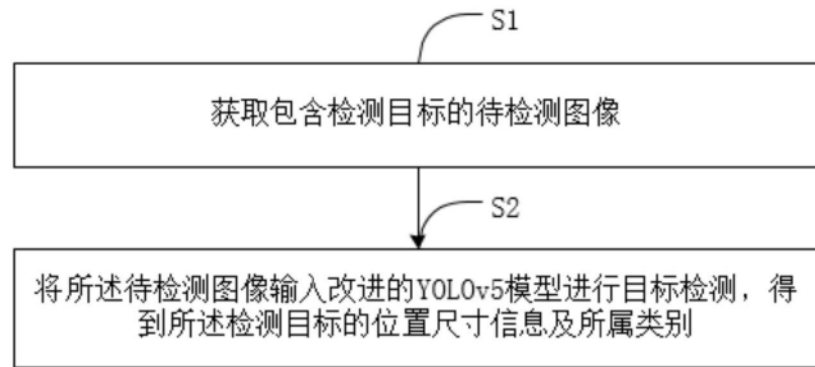


图1

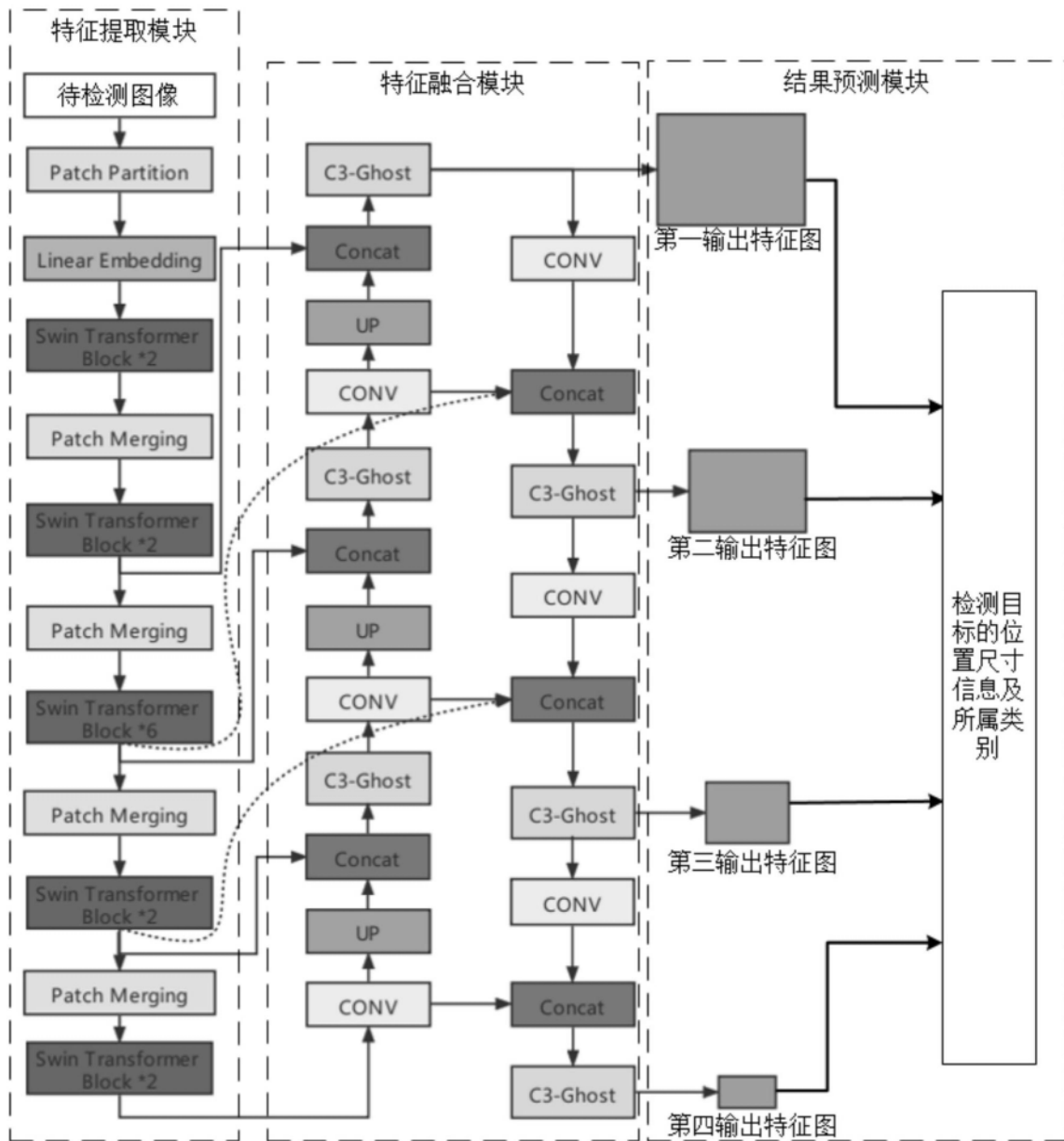


图2

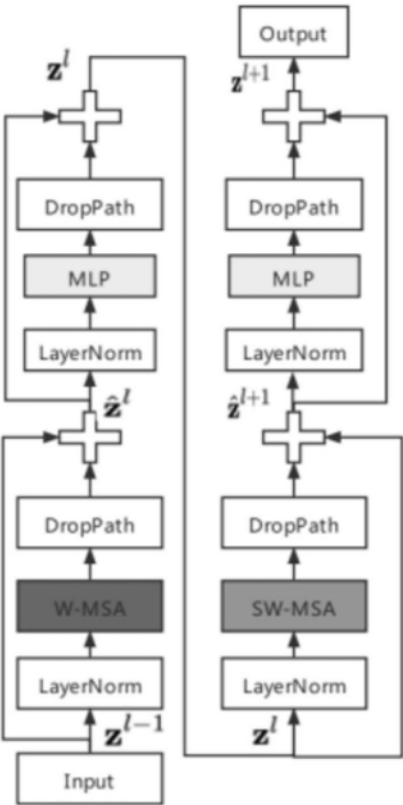


图3

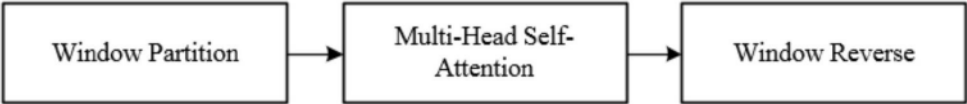


图4

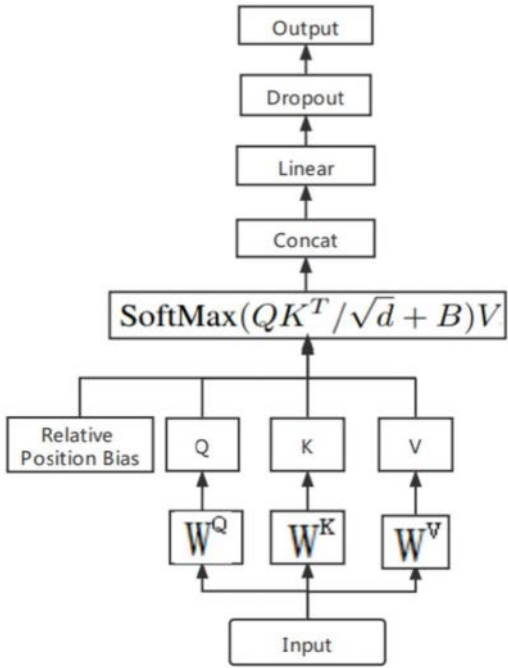


图5

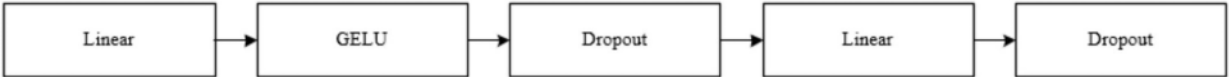


图6

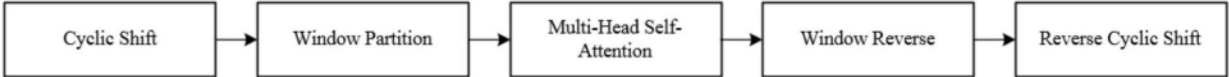


图7

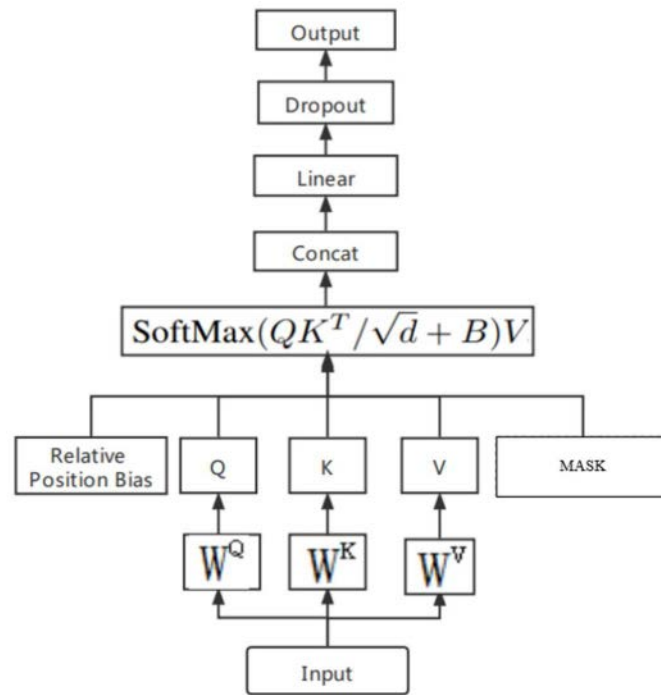


图8

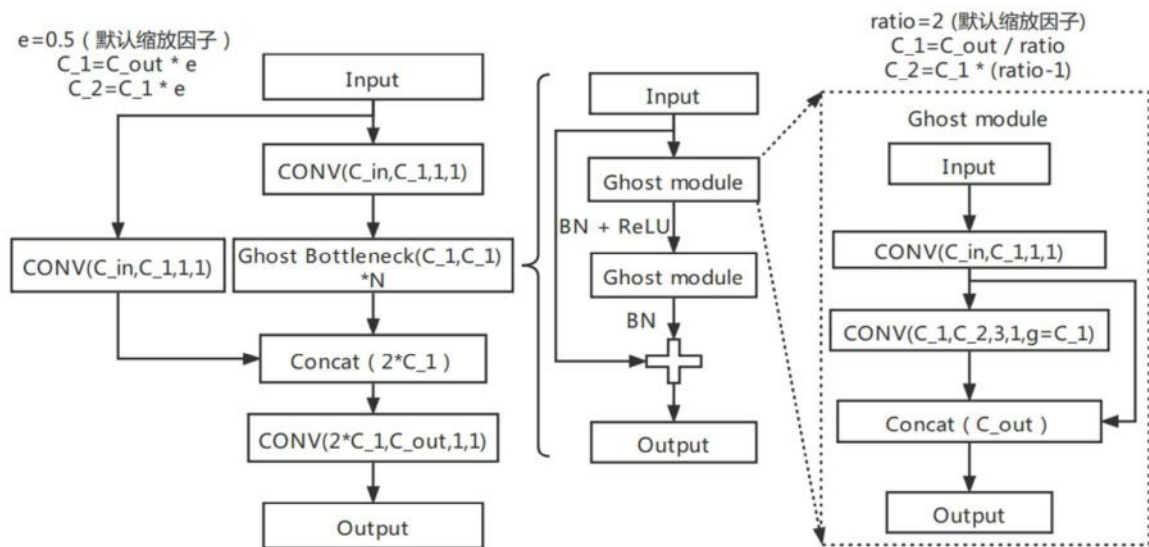


图9



图10



图11



图12



图13



图14



图15