(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116403294 B
(45)授权公告日 2023.10.27

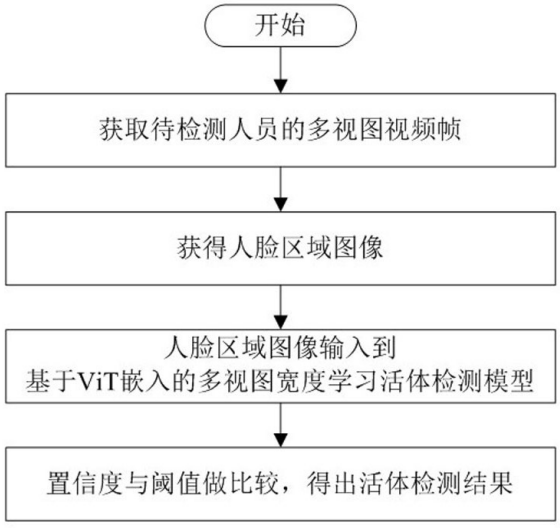
(21)申请号 202310673551.3
(22)申请日 2023.06.08
(65)同一申请的已公布的文献号
申请公布号 CN 116403294 A
(43)申请公布日 2023.07.07
(73)专利权人 华南理工大学
地址 510640 广东省广州市天河区五山路381号
专利权人 人工智能与数字经济广东省实验室(广州)
(72)发明人 陈俊龙 郭继凤 张通 陈业林
(74)专利代理机构 广州市华学知识产权代理有限公司 44245
专利代理师 霍健兰
(51)Int.Cl.
G06F 30/27(2020.01)
G06V 40/40(2022.01)
(56)对比文件
CN 116206373 A,2023.06.02
WO 2023022727 A1,2023.02.23
CN 111461176 A,2020.07.28
CN 110490324 A,2019.11.22
CN 114863497 A,2022.08.05
李佳盈 等.基于ViT的细粒度图像分类.计算机工程与设计.2023,第44卷(第3期),第917-920页.
Zhengang Li 等.Auto-ViT-Acc: An FPGA-Aware Automatic Acceleration Framework for Vision Transformer with Mixed-Scheme Quantization. 2022 32nd International Conference on Field-Programmable Logic and Applications (FPL).2022,全文.
审查员 杨越松
权利要求书2页 说明书6页 附图2页

(54)发明名称

基于Transformer的多视图宽度学习活体检测方法、介质及设备

(57)摘要

本发明涉及活体检测技术领域,具体提供了一种基于Transformer的多视图宽度学习活体检测方法、介质及设备;其中方法为:获取待检测人员的多视图视频帧;多视图视频帧经过人脸区域检测模块获得对应的人脸区域图像;使用基于ViT嵌入的多视图宽度学习活体检测模型对人脸区域图像进行特征提取,并计算输出结果及置信度;将置信度与阈值做比较,得出活体检测结果。该方法能够充分挖掘图像中的人脸关键信息;基于多视图学习技术,能够有效解决因距离、光线等问题造成的不稳定性,不需要刻意面向摄像头进行验证;采用宽度学习方式,能使用比较少的参数实现较高的识别精度和响应速度,具有良好的鲁棒性。



CN 116403294 B

1.一种基于Transformer的多视图宽度学习活体检测方法，其特征在于：包括如下步骤：

S1、获取待检测人员的多视图视频帧；

S2、多视图视频帧经过人脸区域检测模块获得对应的人脸区域图像X_i，$i=1,2,\cdots n$；

S3、使用基于ViT嵌入的多视图宽度学习活体检测模型对人脸区域图像X_i进行特征提取，并计算输出结果及置信度；

S4、将置信度与阈值做比较，得出活体检测结果；

所述S3包括如下分步骤：

S31、对于人脸区域图像X_i，第i个视图ViT模块将对应的人脸区域图像X_i分成若干个Patch，进行线性变换，加上位置编码向量和分类标志位后形成D维的编码序列V_i：

$$V_i = \left[x_c; x_p^1 E; x_p^2 E; \ldots; x_p^N E\right] + E_{pos}$$；

其中，x_c为分类标志位；x_p^k 为Patch压平后的序列数据，$k=1,2,\cdots N$；N为Patch的数量；E为Patch嵌入的全连接层；E_{pos}为位置编码向量；

S32、将编码序列V_i输入到视觉transformer中进行全局注意力计算和特征提取：经过多头注意力机制获得编码序列间特征$V_i'_1$，并利用MLP模块的各层依次对特征$V_i'_1$进行特征变换：

$V_i'_1 = MSA(LN(V_{i1-1})) + V_{i1-1}$；

$V_{i1} = MLP(LN(V_i'_1)) + V_i'_1$；

其中，MSA( )为多头注意力对应的转化函数；LN( )为线性标准化；MLP( )为多层感知机对应的映射函数；V_{i1-1}、V_{i1}分别为MLP模块第$1-1$、1层输出的特征；

经过MLP模块的L层计算得到最后一层特征的输出V_{iL}；

S33、采用宽度学习方式计算人脸区域图像X_i的映射特征组Z^n 及增强特征组H，进而得到多视图宽度学习活体检测模型的输出Y；

S34、对输出Y计算置信度Cd_i；

所述S33是指：

采用宽度学习方式计算每个人脸区域图像X_i对应的映射特征Z_i：

$Z_i = LN(V_{iL})$ ；

所有视图数据对应的映射特征组表示为$Z^n = [Z_1, Z_2, \cdots, Z_n]$；映射特征组$Z^n$经过非线性映射函数，形成增强特征组$H^m = [H_1, H_2, \cdots, H_m]$；

其中，第j个增强特征H_j为，$j=1,2,\cdots m$：

$$H_j \equiv \zeta\left(Z^n W_{hj} + \beta_{hj}\right)$$；

其中，W_{hj}和β_{hj}分别是随机产生的连接权重；$\zeta( )$为非线性激活函数；

合并得到的映射特征Z^n和增强特征组H^m，形成新的人脸活体识别特征$A=[Z^n|H^m]$，并连接到输出层，从而得到多视图宽度学习活体检测模型的输出Y为：

$Y = Max(AW)$ ；

其中,W为多视图宽度学习活体检测模型的特征层和输出层的连接权重。

2.根据权利要求1所述的基于Transformer的多视图宽度学习活体检测方法,其特征在于:

所述多视图宽度学习活体检测模型的特征层和输出层的连接权重W的计算方法是:

使用岭回归来求解权值矩阵,通过公式arg min优化问题求解W:

$$\arg\min_{W}: \|AW - Y\|_v^{\sigma_1} + \lambda\|W\|_u^{\sigma_2};$$

使$\sigma_1 = \sigma_2 = v = u = 2$,解得:$W=(\lambda I+ AA^T)^{-1}A^TY$。

3.根据权利要求1所述的基于Transformer的多视图宽度学习活体检测方法,其特征在于:所述置信度Cd_i 的计算方法是:

$$Cd_i = \frac{e^{Y_i}}{\sum_p e^{Y_p}}\ i,p = 1,2,3,\ldots,n\ 。$$

4.一种可读存储介质,其特征在于,其中所述存储介质存储有计算机程序,所述计算机程序当被处理器执行时使所述处理器执行权利要求1-3中任一项所述的基于Transformer的多视图宽度学习活体检测方法。

5.一种计算机设备,包括处理器以及用于存储处理器可执行程序的存储器,其特征在于,所述处理器执行存储器存储的程序时,实现权利要求1-3中任一项所述的基于Transformer的多视图宽度学习活体检测方法。

基于Transformer的多视图宽度学习活体检测方法、介质及设备

技术领域

[0001] 本发明涉及活体检测技术领域，更具体地说，涉及一种基于Transformer的多视图宽度学习活体检测方法。

背景技术

[0002] 在考试、考勤、在线支付等场景中，人脸识别和身份认证技术发挥至关重要的作用。人脸识别技术因其便捷和非接触等优点渗透到各个商业应用中，如互联网金融行业识别开户身份，防诈骗、防损失；移动出行识别司机身份，保证司乘安全；在线考试远程识别学生身份防止替考；电子民生社保认证；在线医疗中挂号拒绝排队难……但单一的人脸识别系统仍无法准确辨别人脸真伪，造成安全性隐患。因此活体检测成为人脸识别迈向更高层次的核心技术，具有很高的研究意义和商业应用价值。

[0003] 在金融支付，门禁等应用场景，活体检测一般是嵌套在人脸检测与人脸验证中的模块，用来验证是否用户真实本人，可以防止照片、视频、面具等攻击手段用于身份认证，杜绝在考勤、签到、考试等场景的顶替、作弊行为，保证真人通过率，实现人脸识别的安全保障。因此，在进行人脸识别前要首先判断捕捉的人脸是否是一个真实的脸部，之后再进行身份验证，这样有助于杜绝顶替、作弊行为，保证真人通过率。

[0004] 目前，活体检测技术主要分为基于手工设计特征的方法和基于深度学习的方法。人工设计的特征针对图像采集时的信息损失和噪声引入，对比图像的纹理差异，如局部高光、阴影变化、模糊程度和高频分量信息损失等实现识别目的。这种基于纹理信息的方法简单，实时性高，成本低，但随着高清摄像机和高清3D面具的应用，其不足之处日益凸显。基于运动信息的检测方法是常见识别率较高的人脸认证技术，但它需要认证人员的高度配合，检测过程不友好，且耗时较长。其他如基于深度信息、热红外成像分析、心率检测分析的方法需要较高的底层硬件支持以获得所需人工设计特征。总体来讲，这类方法虽识别率较高，但严重依赖于特征表达(需要解决细节损失、颜色失真、阴影模糊和图像高光等问题)和硬件支持，在视频回放、3D面具等逼真的伪信息下，鲁棒性和泛化能力有限。

[0005] 相较于基于手工设计特征的方法，基于深度学习的活体检测方式具有无法比拟的优势，适用于各种欺骗手段。如针对照片和视频攻击的双流CNN的人脸反欺骗方法；反3D 面具欺骗方法；使用Inception 和ResNet架构在不同的环境下的人脸欺骗检测等。然而该类方法成本高、体量大，轻量化部署难度高，不能满足在线实时处理需求。因此，如何设计出一种检测精度高、耗时短、鲁棒性强、实时响应的人脸活体检测技术至关重要。

发明内容

[0006] 为克服现有技术中的缺点与不足，本发明的目的在于提供一种基于Transformer的多视图宽度学习活体检测方法、介质及设备；该方法将视觉Transformer机制嵌入到宽度学习的映射特征节点层，能够充分挖掘图像中的人脸关键信息；基于多视图学习技术，能够

有效解决因距离、光线等问题造成的不稳定性，不需要刻意面向摄像头进行验证；采用宽度学习方式，能使用比较少的参数实现较高的识别精度和响应速度，具有良好的鲁棒性。

[0007] 为了达到上述目的，本发明通过下述技术方案予以实现：一种基于Transformer的多视图宽度学习活体检测方法，包括如下步骤：

[0008] S1、获取待检测人员的多视图视频帧；

[0009] S2、多视图视频帧经过人脸区域检测模块获得对应的人脸区域图像X_i，$i=1,2,\cdots n$；

[0010] S3、使用基于ViT嵌入的多视图宽度学习活体检测模型对人脸区域图像X_i进行特征提取，并计算输出结果及置信度；

[0011] S4、将置信度与阈值做比较，得出活体检测结果；

[0012] 所述S3包括如下分步骤：

[0013] S31、对于人脸区域图像X_i，第i个视图ViT模块将对应的人脸区域图像X_i分成若干个Patch，进行线性变换，加上位置编码向量和分类标志位后形成D维的编码序列V_i：

[0014] $$V_i = \left[x_c; x_p^1 E; x_p^2 E; \ldots; x_p^N E\right] + E_{pos}$$；

[0015] 其中，x_c为分类标志位；x_p^k为Patch压平后的序列数据，$k=1,2,\cdots N$；N为Patch的数量；E为Patch嵌入的全连接层；E_{pos}为位置编码向量；

[0016] S32、将编码序列V_i输入到视觉transformer中进行全局注意力计算和特征提取：经过多头注意力机制获得编码序列间特征$V_i'_1$，并利用MLP模块的各层依次对特征$V_i'_1$进行特征变换：

[0017] $V_i'_1 = MSA(LN(V_{il-1})) + V_{il-1}$；

[0018] $V_{il} = MLP(LN(V_i'_1)) + V_i'_1$；

[0019] 其中，MSA( )为多头注意力对应的转化函数；LN( )为线性标准化；MLP( )为多层感知机对应的映射函数；V_{il-1}、V_{il}分别为MLP模块第$l-1$、l层输出的特征；

[0020] 经过MLP模块的L层计算得到最后一层特征的输出V_{iL}；

[0021] S33、采用宽度学习方式计算人脸区域图像X_i的映射特征组Z^n及增强特征组H，进而得到多视图宽度学习活体检测模型的输出Y；

[0022] S34、对输出Y计算置信度Cd_i。

[0023] 优选地，所述S33是指：

[0024] 采用宽度学习方式计算每个人脸区域图像X_i对应的映射特征Z_i：

[0025] $Z_i = LN(V_{iL})$；

[0026] 所有视图数据对应的映射特征组表示为$Z^n=[Z_1, Z_2, \cdots, Z_n]$；映射特征组$Z^n$经过非线性映射函数，形成增强特征组$H^m=[H_1, H_2, \cdots, H_m]$；

[0027] 其中，第j个增强特征H_j为，$j=1,2,\cdots m$：

[0028] $$H_j \equiv \zeta\left(Z^n W_{hj} + \beta_{hj}\right)$$；

[0029] 其中，W_{hj}和β_{hj}分别是随机产生的连接权重；$\zeta( )$为非线性激活函数；

[0030] 合并得到的映射特征Z^n和增强特征组H^m，形成新的人脸活体识别特征$A=[Z^n|H^m]$，

并连接到输出层，从而得到多视图宽度学习活体检测模型的输出Y为：

[0031] $Y=Max(AW)$；

[0032] 其中，W为多视图宽度学习活体检测模型的特征层和输出层的连接权重。

[0033] 优选地，所述多视图宽度学习活体检测模型的特征层和输出层的连接权重W的计算方法是：

[0034] 使用岭回归来求解权值矩阵，通过公式arg min优化问题求解W：

[0035] $$\arg\min_{W}: \|AW-Y\|_{v}^{\sigma_1}+\lambda\|W\|_{u}^{\sigma_2}$$；

[0036] 使$\sigma_1=\sigma_2=v=u=2$，解得：$W=(\lambda I+AA^T)^{-1}A^TY$。

[0037] 优选地，所述置信度Cd_i的计算方法是：

[0038] $$Cd_i=\frac{e^{Y_i}}{\sum_p e^{Y_p}}\ i,p=1,2,3,\ldots,n$$。

[0039] 一种可读存储介质，其中所述存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行上述基于Transformer的多视图宽度学习活体检测方法。

[0040] 一种计算机设备，包括处理器以及用于存储处理器可执行程序的存储器，所述处理器执行存储器存储的程序时，实现上述基于Transformer的多视图宽度学习活体检测方法。

[0041] 与现有技术相比，本发明具有如下优点与有益效果：

[0042] 1、针对主流的基于深度学习的检测方法存在的消耗大，训练时间长等问题，本发明以宽度学习为基础循序渐进设计高效的轻量级检测方法；将视觉transformer嵌入宽度学习生成映射特征组，提高了宽度学习的特征提取能力，并通过增强层加强和融合特征，为后期人脸识别模块提供有效的人脸特征；此外，由于宽度学习的特点，该框架能使用比较少的参数实现较高的识别精度和响应速度；

[0043] 2、本发明采用的多视图方式能够有效解决由于距离、角度和光照等原因造成的不稳定性，不需要刻意面向摄像头进行验证，对环境等因素具有较高的容忍性；

[0044] 3、本发明研发基于宽度学习的轻量化人脸活体检测技术，可提高识别精度，并解决基于深度学习的方法时间消耗和资源占用问题，有助于模型的快速开发和技术部署；本发明中使用的增量学习方式能够方便利用新增加的各种类型攻击手段数据，无需模型重建，使用最低成本提高模型鲁棒性。

附图说明

[0045] 图1是本发明基于Transformer的多视图宽度学习活体检测方法的流程图；

[0046] 图2是本发明多视图宽度学习活体检测模型的结构框图；

[0047] 图3是本发明多视图宽度学习活体检测模型的视图ViT模块。

具体实施方式

[0048] 下面结合附图与具体实施方式对本发明作进一步详细的描述。

[0049] 实施例

[0050] 本实施例一种基于Transformer的多视图宽度学习活体检测方法，如图1所示，包括如下步骤：

[0051] S1、获取待检测人员的多视图视频帧。

[0052] S2、多视图视频帧经过人脸区域检测模块获得对应的人脸区域图像X_i，$i=1,2,\cdots n$。人脸区域检测模块可采用现有技术。

[0053] S3、使用基于ViT嵌入的多视图宽度学习活体检测模型对人脸区域图像X_i进行特征提取，多视图宽度学习活体检测模型，并计算输出结果及置信度。

[0054] 多视图宽度学习活体检测模型，如图2。多视图的人脸区域图像X_i通过多组轻量化预训练视图ViT模块分别提取不同视图数据下的初步特征，然后通过非线性函数将其映射，得到增强节点，该部分特征能够有效结合多视图数据的互补信息等。最后，单视图特征与融合特征，即最终人脸活体多粒度识别特征连接到输出层。输出层输出置信度。根据置信度，判断该区域人脸是否是活体。值得注意的是，该模型也可采用增量学习的方式进行训练。这样既能缓解计算压力，又能在动态环境中不断更新数据和模型。

[0055] 视图ViT模块，如图3所示，包括：

[0056] Embedding层，用于将输入人脸区域图像X_i格式[H, W, C]转化为Transformer编码器设定的向量序列；

[0057] Transformer 编码器，包含重复堆叠的L次编码器模块，包括Layer Norm、多头注意力部分、Dropout和MLP Block；其中多头注意力部分用于获得关注部分；MLP 层用于得到视图ViT模块最后一层的特征输出。

[0058] S3具体包括如下分步骤：

[0059] S31、对于人脸区域图像X_i，第i个视图ViT模块将对应的人脸区域图像X_i分成若干个Patch，进行线性变换，加上位置编码向量和分类标志位后形成D维的编码序列V_i：

[0060] $$V_i = \left[x_c; x_p^1 E; x_p^2 E; \ldots; x_p^N E\right] + E_{pos}$$；

[0061] 其中，x_c为分类标志位，该向量是人为设定的，可学习的嵌入向量，用于Transformer训练过程中的类别信息学习；x_p^k为Patch压平后的序列数据，$k=1,2,\cdots N$；N为Patch的数量；E为Patch嵌入的全连接层；E_{pos}为位置编码向量，为了保持输入图像patch之间的空间位置信息；其中位置编码向量由以下公式计算：

[0062] $$E_t^{(j)} = \begin{cases} \sin(\omega_j, t) \\ \cos(\omega_j, t) \end{cases}, \quad \omega_j = \frac{1}{10000^{2j/d}}, \quad j = 0,1,2,3,\ldots\ldots,\frac{d}{2}-1,$$；

[0063] 其中，t为向量在序列中的实际位置(例如第一个向量为1，第二个向量为2…)；d为向量对应的维度；

[0064] S32、将编码序列V_i输入到视觉transformer中进行全局注意力计算和特征提取：经过多头注意力机制获得编码序列间特征$V_i'_1$，并利用MLP模块的各层依次对特征$V_i'_1$进行特征变换：

[0065] $V_i'_1 = MSA(LN(V_{i1-1})) + V_{i1-1}$；

[0066] $V_{i1} = MLP(LN(V_i'_1)) + V_i'_1$；

[0067] 其中，MSA( )为多头注意力对应的转化函数；LN( )为线性标准化；MLP( )为多层感知机对应的映射函数；V_{i1-1}、V_{i1}分别为MLP模块第*1-1*、*1*层输出的特征；

[0068] 这里MSA模块计算过程与Self-Attention中的计算过程一样；

[0069] 经过MLP模块的L层计算得到最后一层特征的输出V_{iL}。

[0070] S33、采用宽度学习方式计算人脸区域图像X_i的映射特征组Z^n及增强特征组H，进而得到多视图宽度学习活体检测模型的输出*Y*；

[0071] 具体地说，采用宽度学习方式计算每个人脸区域图像X_i对应的映射特征Z_i：

[0072] $Z_i=LN(V_{iL})$；

[0073] 所有视图数据对应的映射特征组表示为$Z^n=[Z_1,Z_2,\cdots,Z_n]$；映射特征组Z^n经过非线性映射函数，形成增强特征组$H^m=[H_1,H_2,\cdots,H_m]$；增强特征能实现单视图信息的融合和互补；

[0074] 其中，第*j*个增强特征H_j为，*j=1,2,…m*：

[0075] $$H_j \equiv \zeta\left(Z^n W_{hj} + \beta_{hj}\right)$$；

[0076] 其中，W_{hj}和β_{hj}分别是随机产生的连接权重和偏置；$\zeta()$为非线性激活函数；

[0077] 合并得到的映射特征Z^n和增强特征组H^m，形成新的人脸活体识别特征$A=[Z^n|H^m]$，并连接到输出层；由于每一帧图像对应的人脸真伪信息已知，只需计算特征层和输出层的连接权重*W*即可；

[0078] 多视图宽度学习活体检测模型的特征层和输出层的连接权重*W*的计算方法是：

[0079] $W=A^{-1}Y$；

[0080] 使用岭回归来求解权值矩阵，通过公式arg min优化问题求解*W*：

[0081] $$\arg\min_{W}: \|AW-Y\|_v^{\sigma_1} + \lambda\|W\|_u^{\sigma_2}$$；

[0082] 使$\sigma_1=\sigma_2=v=u=2$，解得：$W=(\lambda I+AA^T)^{-1}A^TY$。

[0083] 从而得到多视图宽度学习活体检测模型的输出*Y*为：

[0084] $Y=Max(AW)$。

[0085] S34、对输出Y计算置信度Cd_i：

[0086] $$Cd_i=\frac{e^{Y_i}}{\sum_p e^{Y_p}}\ i,p=1,2,3,\ldots,n$$。

[0087] S4、将置信度与阈值做比较，得出活体检测结果。

[0088] S4之后还可以包括S5：根据活体检测结果，决定是否进行人脸识别。若为假体，则给出警报和提示；若为真人，则进行人脸识别。

[0089] 实施例二

[0090] 本实施例一种可读存储介质，其中所述可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行实施例一所述的基于Transformer的多视图宽度学习活体检测方法。

[0091] 实施例三

[0092] 本实施例一种计算机设备,包括处理器以及用于存储处理器可执行程序的存储器,所述处理器执行存储器存储的程序时,实现实施例一所述的基于Transformer的多视图宽度学习活体检测方法。

[0093] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。

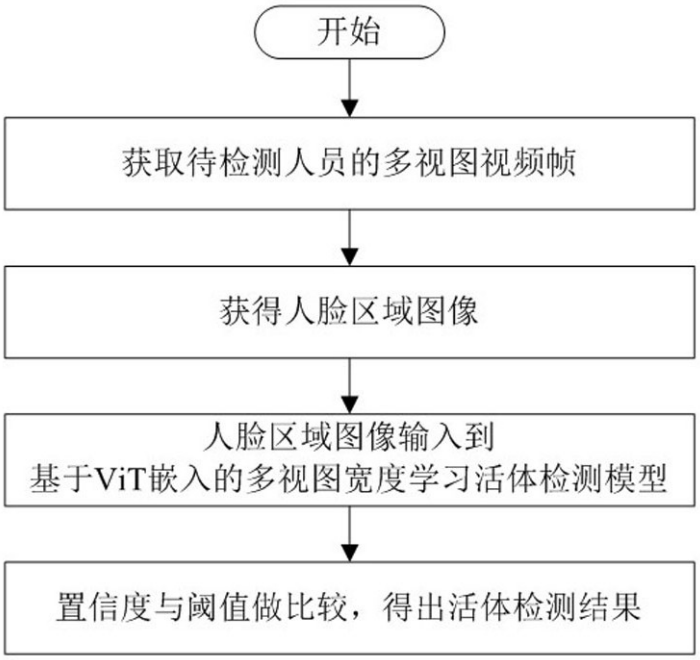

开始
获取待检测人员的多视图视频帧
获得人脸区域图像
人脸区域图像输入到
基于ViT嵌入的多视图宽度学习活体检测模型
置信度与阈值做比较，得出活体检测结果

图1

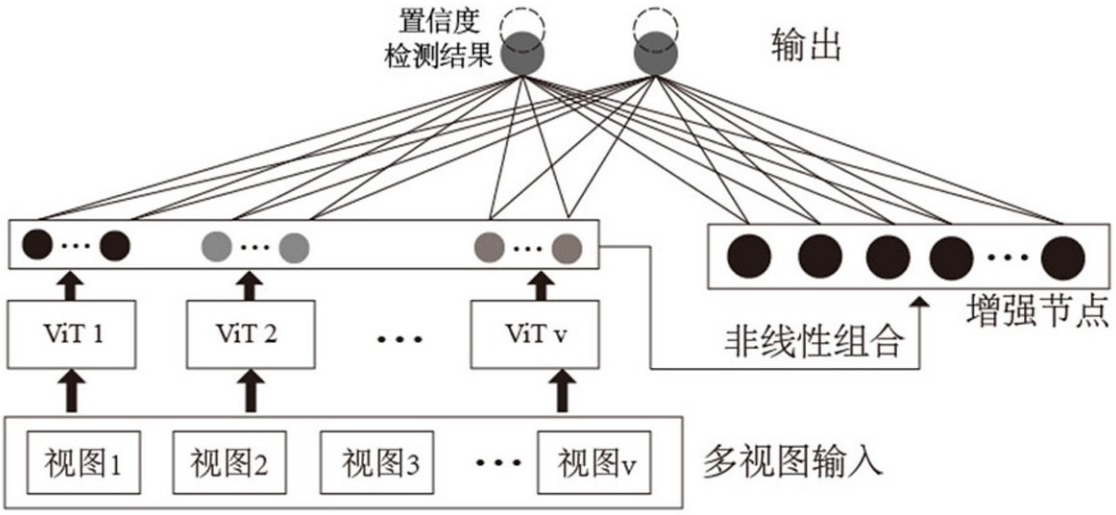

置信度
检测结果
输出
ViT 1
ViT 2
ViT v
非线性组合
增强节点
视图1
视图2
视图3
视图v
多视图输入

图2

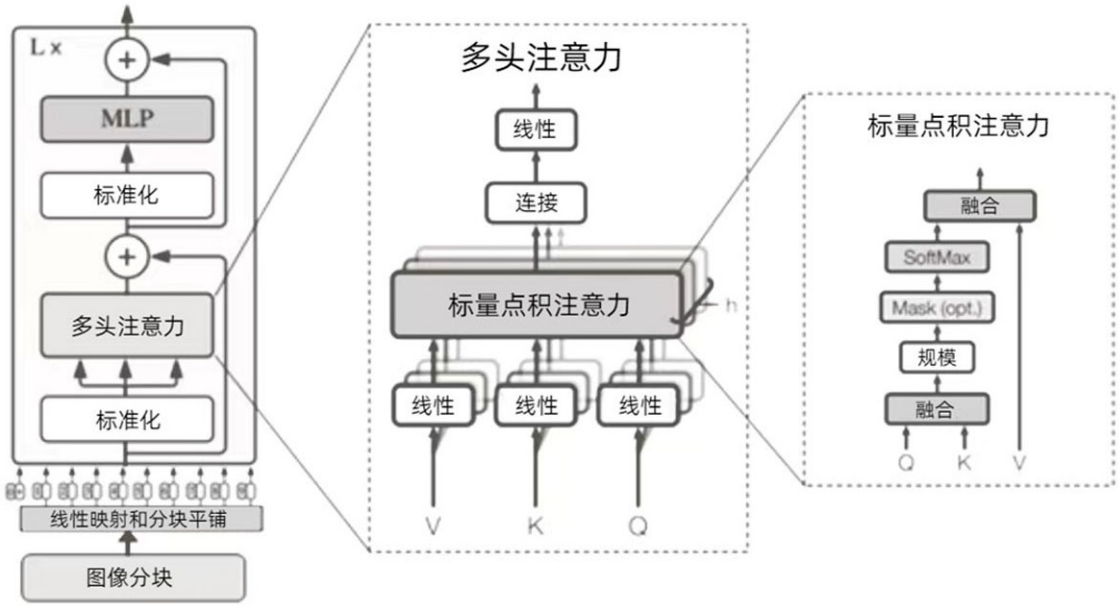

L x
MLP
标准化
多头注意力
标准化
线性映射和分块平铺
图像分块
多头注意力
线性
连接
标量点积注意力
h
线性
线性
线性
V
K
Q
标量点积注意力
融合
SoftMax
Mask (opt.)
规模
融合
Q
K
V

图3