

基于改进 GhostFaceNet 的电梯内人脸检测模型研究

张艳红¹ 孙浩波² 丁新其³ 丁玉俊¹

1. 江苏省特种设备安全监督检验研究院泰州分院 江苏泰州 225300

2. 江苏省特种设备安全监督检验研究院盐城分院 江苏盐城 224000

3. 江苏省特种设备安全监督检验研究院 江苏南京 210000

摘要: 为提升系统对多源视觉信息的感知与识别能力, 本文提出一种融合改进型 GhostFaceNet 的多源视觉检测模型 ResSpatial_GhostFaceNet, 针对电梯等狭小场景下的人脸遮挡问题进行算法优化。通过引入多尺度卷积、空间注意力机制、残差结构和 ArcFace 损失函数, 提升了模型在边缘设备上的识别精度与鲁棒性。实验基于真实口罩人脸数据集开展对比与消融测试, 结果表明优化模型在准确率、Loss 与参数控制方面均优于主流轻量模型, 展现了良好的多源视觉识别能力。

关键词: 多源视觉; 遮挡识别; GhostFaceNet; 注意力机制; 人脸识别; 轻量模型

引言

在电梯、人流密集场所与医院等特殊环境中, 视觉检测模型面临多个挑战: 遮挡率高(如口罩佩戴)、光线变化剧烈、设备资源有限。常规深度模型在精度与计算效率之间难以兼顾。因此, 设计高精度、低资源消耗的多源视觉检测模型成为关键任务。GhostFaceNet 作为轻量化人脸识别领域的代表性模型, 已在低功耗场景下展现出优秀性能。本文在其基础上提出改进模型 ResSpatial_GhostFaceNet, 旨在强化对复杂遮挡场景的表征能力, 优化模型结构并引入空间注意力机制, 实现对多源视觉信息的精准识别。

1. GhostFaceNet 结构与局限分析

GhostFaceNet 利用 Ghost Module 生成主特征与廉价冗余特征, 在轻量化的同时保留一定特征表达力。其结构包括: 初始卷积、Ghost Bottleneck 堆叠提取层、GAP 层、全连接层与 Softmax 输出。但其在遮挡情况下存在三大问题: ①低层特征稀疏, 细节捕获不足: Ghost 操作导致低层纹理信息表达能力下降。②残差连接限制在局部模块: 缺乏跨层信息交互。③空间注意力机制缺失: 对局部关键区域响应不强。

2. 模型结构优化设计

GhostFaceNet 的 Ghost Bottleneck 是网络的主干部分(backbone), 负责提取不同层次的特征信息。它以轻量化和高效性为核心, 充分结合了 Ghost Module 模块的创新性结构为人脸识别任务提供了强大的特征提取能力。尽管 GhostFaceNet 的 backbone 设计在轻量化和高效性方面表现较

好, 但在人脸存在遮挡情况下的识别任务中仍存在一定的优化空间, 为进一步提升其性能提供了改进的方向和潜力。第一: 低层特征的提取能力较弱, 由于 Ghost Module 模块通过特征分离的方式来扩展原始的特征图, 这减少了特征之间的冗余, 使模型更倾向于生成稀疏特征。稀疏性通常意味着大量的特征值为零或者接近零, 这一特性很可能导致部分低层纹理信息的丢失, 影响模型对面部细节的捕获能力。为此, 本文通过引入逐点卷积(Pointwise Convolution)、增加多尺度卷积和非线性变换等方式, 加强对低层特征细节的提取, 以增强冗余特征的表达能力, 从而提高特征表达的完整性和细致程度。同时, 底层特征和上层特征的拼接方式也是 Ghost Module 改进的重要方向之一。可在 Ghost Module 中引入通道注意力机制或局部特征增强模块, 例如 SE(Squeeze-and-Excitation)通道注意力机制^[1]。在特征底层特征和上层融合前, 用 SE 模块提取低层特征中的重要信息, 提高特征融合的质量。第二: 残差连接的利用方式存在不足目前残差连接主要局限于 Ghost Module 内部, 缺乏跨层的信息交互, 限制了多层次特征的复用效率。针对这一问题, 改进方向在于设计跨层残差连接结构, 例如类似 ResNet(Residual Networks)风格的跳跃连接机制, 通过连接不同深度的层级特征, 增强模型的特征表达能力, 尤其是在处理遮挡人脸表征时的适应性。第三: 对局部区域特征的关注存在不足, 虽然通过引入 SE 模块能够有效关注通道维度的特征, 但 Ghost Bottleneck 对局部空间信息的捕获能力有限, 可能无法充分表征关键区

域 (如眼睛、嘴巴) 的细节信息。

2.1 增强 Ghost Module 模块的特征提取能力

尽管 Ghost Module 在轻量化设计方面表现较好, 通过将主特征与冗余特征的生成分离降低了计算成本, 但其在特征提取能力上仍存在一定的不足。本文从冗余特征生成和特征融合方式两个方面对 Ghost Module 模块进行改进, 提出改进后的 E_Ghost Module(Enhanced_Ghost Module) 模块以增强其特征提取能力。一方面, 在冗余特征生成过程中, 引入多尺度卷积和非线性变换, 以增强冗余特征的表达能力。另一方面, 在主特征和冗余特征的融合过程中, 有选择地加入 SE 通道注意力机制以优化拼接方式。通过对特征通道的加权操作, SE 机制能够更充分地挖掘主特征与冗余特征之间的深层次关系, 提升特征融合的有效性和模型的整体表达能力^[2]。

2.2 引入 ResNet 风格的残差结构增强特征表达能力

在 GhostFaceNet 的 Backbone 中残差连接的实现方式主要局限于 Ghost Module 模块内部, 缺乏跨层信息的有效交互。这种限制使得不同深度层次的特征难以实现充分复用, 从而在处理被遮挡人脸特征表征时表现出一定的适应性不足。尤其是在遮挡、光照变化或多角度输入等场景中, 不同层级特征的协同作用对于建模关键细节显得尤为重要。

为克服这一不足, 本文通过引入跨层残差连接结构实现改进, 借鉴 ResNet 风格的连接方式, 将浅层特征与深层特征进行有效融合增强模型对复杂任务的适应能力, ResNet 风格的连接结构见如图 1 所示。

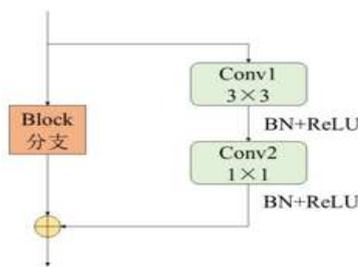


图 1 ResNet 风格的连接方式示意图

3. 改进模型 ResSpatial_GhostFaceNet 结构设计

3.1 引入 Spatial Gate 机制提升空间信息的捕获能力

经过前两点改进后的 Backbone 设计在轻量化和特征表达能力之间取得了一定的平衡, 但其在特征提取过程中存在缺乏空间注意力机制的问题。当前 Backbone 主要通过 E_Ghost Module 和通道注意力机制 (如 SE 模块) 对特征进行优化, 但这些设计主要关注通道维度特征的重要性, 对空间信

息的捕获能力较为有限。在处理遮挡人脸识别任务时, 空间维度的特征分布 (如局部特征的重要性) 对于遮挡、光照变化等复杂场景具有重要作用。因此, 缺乏空间注意力机制可能导致模型对关键空间区域的关注不足, 限制了模型在复杂场景下的鲁棒性与适应能力^[3]。

针对上述不足, 本文提出在 GhostFaceNet 的 Backbone 中进一步引入空间注意力机制, 通过优化特征图的空间加权分布, 增强模型对关键空间区域的关注能力。Spatial Gate 作为一种轻量化的空间注意力模块, 能够通过全局信息聚合、特征融合和加权操作动态调整特征图不同空间位置的权重, 有效提升模型的空间特征表达能力和整体性能。其结构如图 2 所示。

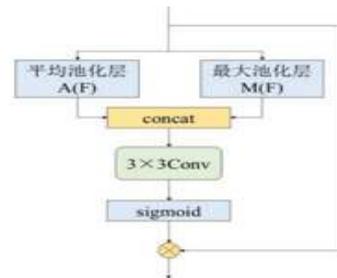


图 2 Spatial Gate 空间注意力机制结构图

其中输入为特征图 $F \in \mathbb{R}^{H \times W \times C}$, 其中 H 和 W 为空间尺寸, C 为通道数。Spatial Gate 通过平均池化和最大池化两条全局信息聚合路径, 提取特征图在通道维度上的全局信息。平均池化对每个通道的特征图进行全局平均池化, 生成空间特征图 $A(F) \in \mathbb{R}^{H \times W \times 1}$ 。最大池化对每个通道的特征图进行全局最大池化, 生成空间特征图 $M(F) \in \mathbb{R}^{H \times W \times 1}$ 。平均池化和最大池化的计算公式分别为:

$$A(F)(i, j) = \frac{1}{C} \sum_{c=1}^C F(i, j, c) \quad (1)$$

$$M(F)(i, j) = \max_{c=1}^C F(i, j, c) \quad (2)$$

接着, 将 $A(F)$ 和 $M(F)$ 在通道维度进行拼接, 生成一个两通道特征图 $S = [A(F); M(F)]$, $S \in \mathbb{R}^{H \times W \times 2}$ 。通过一个 3×3 的卷积操作提取融合特征的空间上下文信息, 生成单通道特征图 $G(F)$, 计算公式为:

$$G(F) = \sigma(\text{Conv } 3 \times 3(S)) \quad (3)$$

式 (3) 中, σ 表示 Sigmoid 激活函数, 用于将输出映射到 $[0, 1]$ 范围内, 作为特征图的空间权重。

最后, 将生成的空间权重 $G(F)$ 与输入特征图 F 逐元素相乘, 完成对输入特征图的加权调整: $F' = F \odot G(F)$ 。其中, \odot 表示逐元素乘法操作, F' 为加权后的输出特征图, 其形状与输入特征图 F 相同。

在特定位置的插入对注意力机制的效果至关重要。第一, 在残差支路中插入 Spatial Gate 机制, 可以针对通过跳跃连接传递的特征动态分配空间权重, 使得残差路径的信息更加贴合主干路径的空间需求。这种调整能够在保留深层特征重用优势的同时, 进一步优化信息流的有效性和梯度传播的稳定性。第二, 在 Ghost Module 模块之间插入 Spatial Gate, 则可以对生成的特征图进行筛选, 尤其是通过廉价操作生成的冗余特征中, 突出有助于网络表达的关键部分。这种选择不仅能够提升模块间的特征关联性, 还能减少不必要的计算开销, 进一步优化轻量化设计的效率。

综合来看, Spatial Gate 的引入位置直接决定了特征选择的有效性和模型性能的提升。结合网络的不同需求, 将其合理分布于网络的残差支路和模块之间, 可以达到局部特征增强和全局特征优化的双重效果。

3.2 改进后的 GhostFaceNet 特征提取网络结构总体设计

为了设计高效且结构优化的 backbone 结构, 需要在不同层次选择具有针对性的特征提取模块, 以在计算复杂度与特征表达能力之间取得最佳平衡。本文根据 backbone 各层次的功能, 对模块的选取和注意力机制的引入进行了详细设计, 从浅层、中间层到深层逐步优化 backbone 的特征提取能力, 并通过跳跃式残差连接增强全局特征流通性。

浅层是网络的初始阶段, 负责从输入数据中提取基本特征。由于浅层特征图的空间分辨率较高, 其计算开销主要集中在卷积操作上。因此, 在该阶段应优先考虑计算效率较高的模块以降低计算成本^[4]。为此, 本文在浅层使用原始 Ghost Module 提取初步特征。原始 Ghost Module 利用主卷积生成基础特征, 同时通过廉价操作生成冗余特征, 其结构简单、计算开销小, 适合浅层特征提取任务。

中间层是特征提取和表达的关键阶段, 其主要任务是通过更复杂的模块对初步特征进行深度处理与增强。在中

间层, 特征图的空间分辨率逐步降低, 但通道数量逐步增加, 因此需要在特征表达能力与计算效率之间进行权衡。本文在中间层采用增强后的 E_Ghost Module。该模块通过多尺度卷积和非线性变换增强特征表达能力, 可以捕获更丰富的特征。

深层是网络的高阶特征提取阶段, 其主要任务是捕获更加抽象和语义化的特征。此阶段的特征图具有更小的空间分辨率和更高的通道数, 因此模块的设计需要关注通道间的重要性分配。在深层, 本文使用增强的 E_Ghost Module, 并进一步结合 SE 通道注意力机制以优化通道间特征权重分布。SE 模块通过显式建模通道间的依赖关系, 能够突出具有语义信息的通道特征, 提高网络的表示能力。

4. 结语

本文围绕多源视觉场景下的识别挑战, 以 GhostFaceNet 为基础提出改进型 ResSpatial_GhostFaceNet 网络。模型引入 E_Ghost 模块、空间注意力、残差结构及 ArcFace 损失函数, 在保持轻量化的基础上, 显著提升遮挡识别精度与鲁棒性。实验验证显示, 本模型在准确率、模型大小、帧率等方面优于多种主流轻量模型, 适合部署于嵌入式系统。未来工作可拓展至多模态融合与跨域泛化方向, 进一步推动多源视觉识别系统的发展。

参考文献:

- [1] 韩凯, 王昱, 田启阳, 等. GhostNet: 基于廉价操作的高效特征网络 [C]// 计算机视觉与模式识别会议 (CVPR). 美国: IEEE, 2020: 1580–1589.
- [2] 邓嘉, 郭健, 薛楠, 等. ArcFace: 用于人脸识别的加性角度间隔损失函数 [C]// 计算机视觉与模式识别会议 (CVPR). 美国: IEEE, 2019: 4690–4699.
- [3] Howard A, Sandler M, Chu G, 等. MobileNetV3: 轻量级神经网络架构搜索 [C]// 国际计算机视觉大会 (ICCV). 韩国: IEEE, 2019: 1314–1324.
- [4] 胡杰, 沈磊, 孙健. Squeeze-and-Excitation Networks [C]// 计算机视觉与模式识别会议 (CVPR). 美国: IEEE, 2018: 7132–7141.