DOI:10.11918/202409021

多尺度融合与感受野增强的生成人脸检测算法

王新良[1],王明旭[1],徐志宗[1,2]

(1.河南理工大学 物理与电子信息学院,河南 焦作 454000;2.鹤壁煤业(集团)有限责任公司供电处,河南 鹤壁 458000)

摘 要:为减少滥用生成人脸技术导致的社交媒体欺诈行为及虚假新闻传播问题,本文提出基于YOLOv8的多尺度融合与感受野增强的生成人脸检测算法(MSERF-YOLOv8)。首先,引入逐通道卷积模块替换原网络中的卷积模块,利用其独立卷积核处理通道的特点,保持特征的独立性;其次,在主干网络高层结合扩张式残差模块构建全新的C2f模块,扩大主干网络高层感受野;再次,引入高效多尺度注意力机制,通过跨空间学习方法聚合多个并行子网络的输出特征图,丰富特征聚合;最后,基于ghost module、动态卷积和C2f模块构造新的混合架构,采用分组卷积参数共享的方式重构轻量化检测头。在5种GAN生成人脸自制数据集上开展实验,结果表明:MSERF-YOLOv8的均值平均精度达到了95.14%,相较于YOLOv8提升了1.25%,参数量仅为5.30×10^6,检测速度达到90.10帧/s;与经典目标检测算法和改进算法相比,该算法在特征提取和特征融合方面均有一定的优势。所提算法提高了生成人脸的检测精度、检测速度,同时节约了计算资源。

关键词:生成人脸检测;YOLOv8;轻量化;特征提取;多尺度融合;注意力机制

中图分类号:TP391.4 **文献标志码**:A **文章编号**:0367-6234(2026)09-0093-11

Generative face detection algorithm based on multi-scale fusion and receptive field enhancement

Wang Xinliang[1], Wang Mingxu[1], Xu Zhizong[1,2]

(1. School of Physics & Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, Henan, China;
2. Power Supply Division, Hebi Coal Industry (Group) Co., Ltd., Hebi 458000, Henan, China)

Abstract: In order to reduce the social media fraud and false news dissemination caused by misuse of generative face technology, a generative face detection algorithm is proposed in this paper based on YOLOv8 with multi-scale fusion and receptive field enhancement(MSERF-YOLOv8). First, a depthwise convolution module is introduced to replace the convolution module in the original network, utilizing its independent convolution kernel to process the channels and maintain feature independence; second, a new C2f module is constructed by combining the expansive residual module at the higher level of the backbone network to expand the sensory field at the higher level of the backbone network; third, an efficient multi-scale attention mechanism is introduced to aggregate the output feature maps of multiple parallel sub-networks through a cross-space learning method to enrich the feature aggregation; finally, a new hybrid architecture is constructed based on the ghost module, dynamic convolution and C2f module, the grouped convolutional parameter sharing is utilized to reconstruct the lightweight detection head. Experiments were conducted on five self-made GAN-generated face datasets, and the results show that the mean average accuracy of MSERF-YOLOv8 reaches 95.14%, which is 1.25% higher compared to YOLOv8, with the number of references being only 5.30×10^6 and the detection speed 90.10 frames/s. Compared with the classical target detection algorithm and some improved algorithms, the proposed algorithm has certain advantages in both feature extraction and feature fusion. The proposed algorithm improves the detection accuracy and detection speed of generative faces, while saving computational resources.

Keywords: generative face detection; YOLOv8; lightweight; feature extraction; multi-scale fusion; attention mechanism

收稿日期:2024-09-09;**录用日期**:2024-10-09;**网络首发日期**:2025-06-09
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250606.1512.004
基金项目:河南省高等学校青年骨干教师培养计划(2019GGJS060);河南省高等学校重点科研基金(21B413005)
作者简介:王新良(1980—),男,教授,硕士生导师
通信作者:王新良,wangxl@hpu.edu.cn

随着深度学习技术的飞速发展,生成对抗网络(generative adversarial network,GAN)在计算机视觉领域取得了显著的成果,尤其在生成人脸图像方面表现出色[1]。GAN 可通过生成模型和判别模型之间的对抗训练生成与真实人脸视觉相似度极高的图像[2],通过人眼难以进行区分。该技术为虚拟现实、人脸识别等领域带来了前所未有的创新。然而,GAN 的广泛应用也伴随着安全与伦理挑战。一些不法分子利用该技术生成虚假人脸图像,进而实施包括制造虚假新闻、散播诽谤信息、诈骗等在内的不法行为,这些行为对社会的信任体系和个人隐私构成了严重威胁。因此,对生成人脸的检测显得尤为必要,具有重要的现实意义。

传统的识别方法通常依赖于手工设计的特征,如边缘、纹理和颜色直方图等,这些特征在处理自然图像时具有一定的效果。然而,随着 GAN 的发展,生成的人脸图像在视觉效果上与真实图像几乎没有差异,甚至能够精确模仿复杂的面部细节和光影效果,使得传统识别方法难以有效区分真实人脸和生成图像。近年来,基于深度学习的生成人脸检测方法迅速发展。文献[3]通过编码像素对特征相似性来捕捉像素级关系,并采用多种度量标准测量区域级别的不一致性,以检测伪造人脸,但在处理完全合成的人脸图像时精度不佳;文献[4]提出了结合全局特征和局部特征的生成人脸检测方法,并采用基于 ArcFace 损失的度量技术提取共同特征和判别特征进行分类,但在处理图像压缩问题时精度会降低;文献[5]提出了基于场景理解模型的生成人脸检测方法,该方法通过提取人脸的上下文特征来区分真假人脸,但其提取的上下文特征可能包含与人脸无关的信息,从而引入误导性特征;文献[6]提出了基于光响应非均匀性(photo response non-uniformity,PRNU)的双流 GAN 生成人脸检测算法,该算法融合了 GAN 指纹与 PRNU 特征,通过识别真伪图像在内容及像素层面的细微差异进行生成人脸检测,但其在局部伪造人脸图像上的有效性尚待进一步验证;文献[7]提出了基于空频联合卷积神经网络的 GAN 生成人脸检测算法,该算法通过学习频率域中的伪影和空间域中的图像内容差异并将空频特征融合来鉴别生成人脸图像,但其对模糊、压缩、噪声的人脸图像检测精度略有不足;文献[8]利用双流网络从颜色空间中提取特征,结合注意力特征融合模块和多层次特征融合模块,增强模型对局部生成人脸特征的提取能力,但在该过程中易受背景干扰的影响。上述算法虽然实现了生成人脸的检测,但均存在一些问题。

鉴于此,本文提出基于 YOLOv8(you only look once version 8)的多尺度融合与感受野增强的生成人脸检测算法(generative face detection algorithm based on multi-scale fusion and receptive field enhanced in YOLOv8,MSERF-YOLOv8)。该算法通过构建轻量整体网络结构,加强主干网络高层特征提取,设计多尺度主干网络,组建轻量化颈部网络,重构轻量化检测头,旨在提高对生成人脸的检测精度和速度,为研究生成人脸轻量化检测算法提供参考。

1 MSERF-YOLOv8 算法

1.1 MSERF-YOLOv8 结构

YOLOv8 算法是一种高效、准确的目标检测算法。但生成人脸技术生成的人脸图像能模拟复杂的纹理和细腻的肤质,还能准确再现面部几何特征,进一步提升了逼真度,增加了 YOLOv8 识别的难度,且 YOLOv8 本身结构复杂、参数量较大。

本文基于 YOLOv8 提出了 MSERF-YOLOv8,结构如图 1 所示,其中虚线框标注的模块为改进的部分。MSERF-YOLOv8 主要包括主干网络、颈部网络及检测头网络等。输入端将图像送入主干网络进行特征提取;颈部网络进行特征融合,并将融合后的特征图送入检测头进行目标检测;检测头利用回归分支和分类分支预测目标的边界框和类别;对检测头的输出进行处理,去除冗余的检测框并确定最终检测结果。

1.2 构建轻量化整体网络架构

在 YOLOv8 中,传统卷积操作计算时,卷积核同时作用于所有输入通道进行加权和激活操作。在该过程中,每个输出像素的计算均涉及所有通道的数据,尽管该方式能够有效提取特征,但在计算上却相对冗余,尤其是在处理多通道特征图时,计算开销较大。针对此问题,本文引入逐通道卷积替换整体网络中的卷积,构建轻量化网络架构,降低计算成本[9]。逐通道卷积模块结构如图 2 所示。其中,每个通道均由独立的卷积核处理,其输出的通道数与输入的通道数完全相同,因此,不同通道的特征信息在卷积过程中不会相互干扰,这有助于保留更多的原始信息,为后续的处理提供更多的灵活性,还可降低卷积层的参数量,从而减少算法的计算开销。

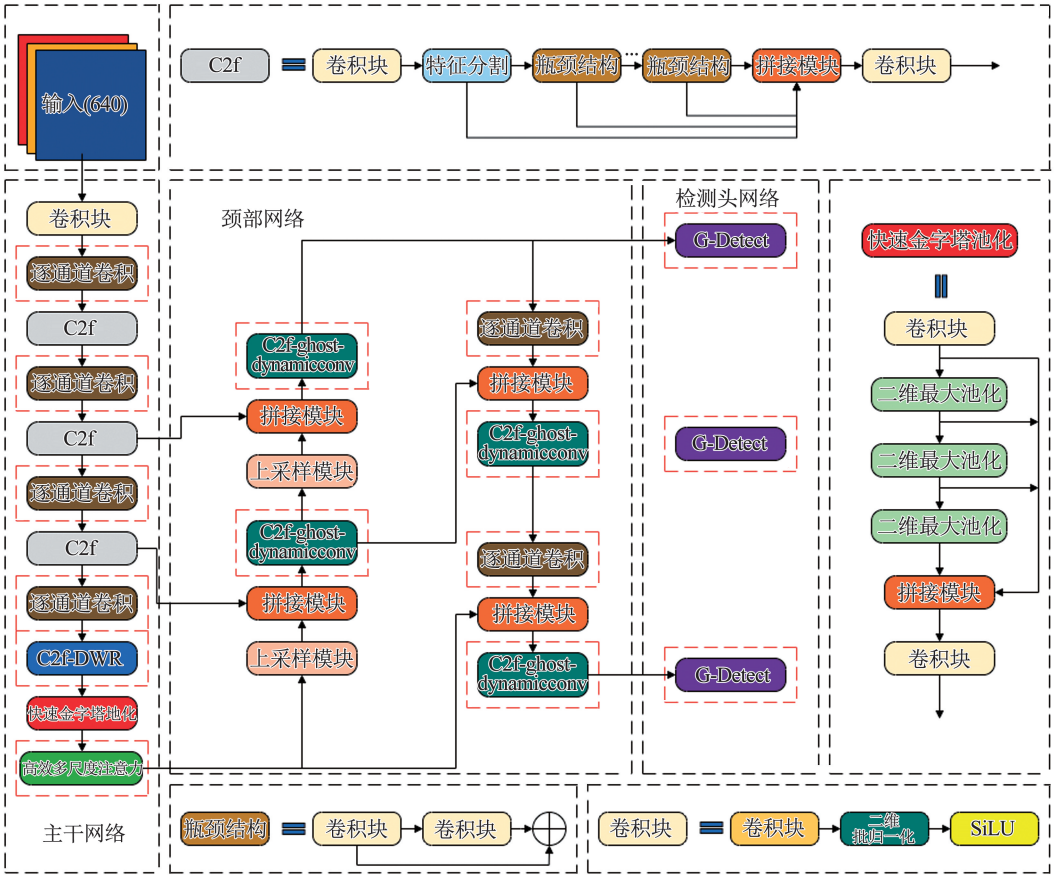


图 1　MSERF-YOLOv8 算法结构

Fig. 1　MSERF-YOLOv8 algorithm structure

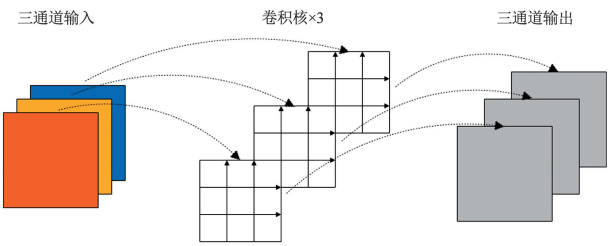


图 2　逐通道卷积模块结构

Fig. 2　Depthwise convolution module structure

在逐通道卷积模块中，卷积核为单通道模式，需要对输入的每一个通道进行卷积，一个通道只被一个卷积核卷积。输入特征图先经过卷积运算，卷积核的数量与上一层的通道数相同，所以输入特征图经过卷积运算后生成与输入特征图通道数一致的输出特征图。卷积过程表达式为

$$O(h,w,c)=\sum_{i=0}^{h_k-1}\sum_{j=0}^{w_k-1}I(h+i,w+j,c)\times \boldsymbol{K}(i,j,c) \tag{1}$$

式中：h、w 分别为输出特征图的高度、宽度索引，c 为通道索引，h_k、w_k 分别为卷积核的高度、宽度，$I(h+i,w+j,c)$ 为输入特征图在位置 $(h+i,w+j,c)$ 的值，$\boldsymbol{K}(i,j,c)$ 为卷积核 $\boldsymbol{K}$ 在位置 (i,j,c) 的权重值。

1.3　加强主干网络高层特征提取

YOLOv8 主干网络中的 C2f（csp layer with 2 convolutions）模块能够融合不同层级的特征，但在处理多尺度人脸目标时缺乏一定的多尺度特征提取能力，无法从目标中高效提取特征信息，并且其感受野会受到限制。此外，C2f 模块的计算开销相对较高，难以满足生成人脸检测任务对高精度和轻量化的需求。针对此问题，本文在主干网络高层中结合扩张式残差（dilation-wise residual，DWR）模块和 C2f 模块构建 C2f-DWR 模块，加强从主干网络高层可扩展感受野中提取特征的能力并降低参数量[10]。C2f-DWR 模块结构如图 3 所示。其中，DWR 模块采用残差方式设计，能高效提取多尺度的上下文信息，融合多尺度感受野生成的特征图，同时还能有效融合不同层次的特征，提高多尺度信息捕获的效率。此外，DWR 模块利用多速率扩展深度卷积扩展感受野，捕捉更多上下文信息，可提升算法网络高层特征的表达能力。

为捕捉更丰富的上下文信息，DWR 模块采用两步法进行上下文信息的获取，结构如图 4 所示。

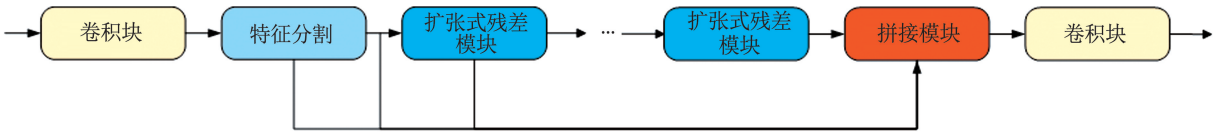


图 3 C2f-DWR 模块结构

Fig. 3 C2f-DWR module structure

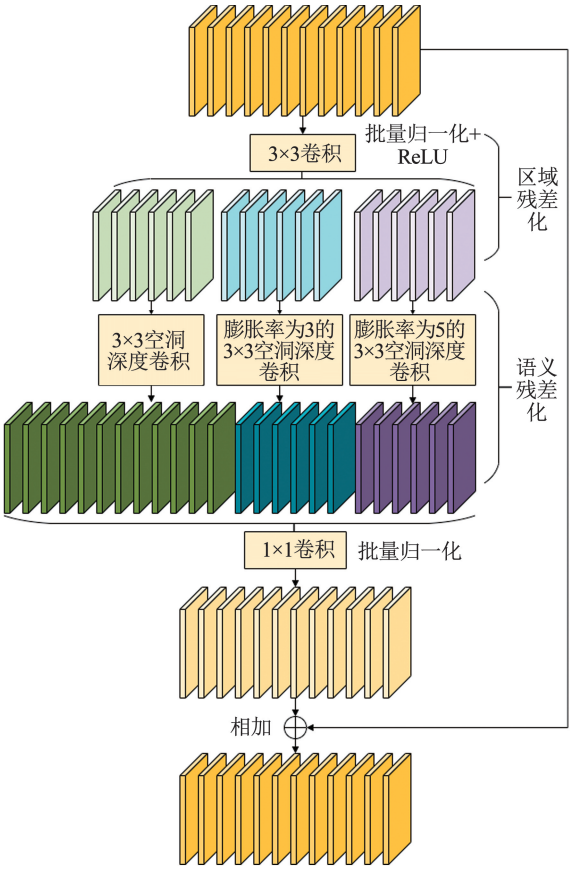


图 4 DWR 模块结构

Fig. 4 DWR module structure

第 1 步进行区域残差化。通过 1 个 3 × 3 卷积层,结合批量归一化和 ReLU 激活函数,完成输入特征的初步提取与空间编码,生成多尺度特征图。第 2 步进行语义残差化。对分组后的特征图施加不同膨胀率的空洞深度卷积,实现多尺度形态学滤波,确保每个通道仅对应单一感受野,避免特征冗余。通过两步法,将多速率扩展深度卷积简化为对各特征图进行形态学滤波。区域特征映射将其转化为形态滤波,高效提取多尺度上下文信息;随后聚合多个输出,对映射执行批量归一化,并通过逐点卷积融合特征,生成残差并叠加至输入,实现更全面的特征表示。

1.4 设计多尺度主干网络

尽管 YOLOv8 在多尺度检测上有所优化,但其主干网络在提取特征时经过多次卷积和池化,可能导致不同尺度人脸的特征信息丢失,影响算法的检测能力。因此,本文在主干网络中的快速金字塔池化(spatial pyramid pooling-fast,SPPF)模块后引入高效多尺度注意力(efficient multi-scale attention,EMA)模块,设计高效的多尺度主干网络,以提高算法对多尺度人脸的检测能力[11]。EMA 结构如图 5 所示。

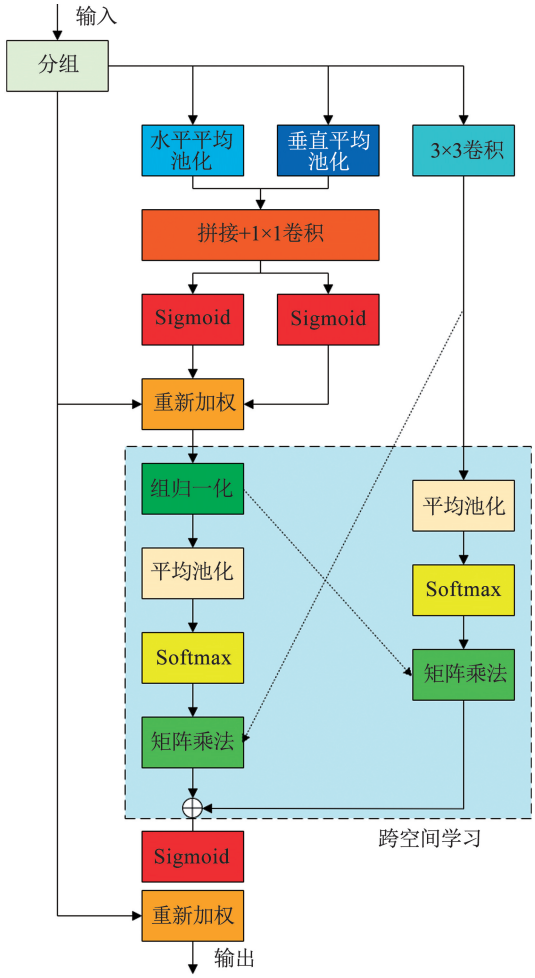


图 5 EMA 结构

Fig. 5 EMA structure

EMA 主要包括特征分组、并行分支和跨空间信息聚合 3 个操作。首先,将输入特征图 $\boldsymbol{X}$ 分为 G 个子特征组进行多语义学习。随后,通过两个 1 × 1 分支和 1 个 3 × 3 分支提取注意力权重。其中,1 × 1 分支采用空间方向编码和一维全局平均池化,而 3 × 3 分支通过单层卷积捕获多尺度特征。EMA 将编码特征拼接后共享 1 × 1 卷积层以保留丰富的信息,并将其输出分解为两个向量,通过 Sigmoid 函数拟合二维二项分布。最终通过乘法聚合通道注意力图实现跨通道交互。同时,3 × 3 分支利用卷积增强局部跨通道交互扩展特征空间。此外,EMA 引入跨空间聚合,对 1 × 1 分支输出进行二维全局平均池化

编码，结合 3 ×3 分支输出实现多维度特征聚合。该过程中池化操作表示为

$$Z_c = \frac{1}{H \times W}\sum_{b}^{H}\sum_{a}^{W} x_q(a,b) \tag{2}$$

式中：H、W 分别为特征图的高、宽，$x_q(a,b)$ 为第 q 个通道上位置为 (a,b) 的元素值。

EMA 利用二维全局平均池化和基于高斯映射的 Softmax 进行线性变换，其输出与 3 ×3 分支点积生成首个空间注意力图，聚焦精细空间依赖；同时，对 3 ×3 分支执行全局平均池化，与 1 ×1 分支点积得到第 2 个空间注意力图，保留位置信息。二者权重聚合后重加权输入特征图以提升效率。EMA 将特征分组，通过并行分支提取多尺度人脸特征，并全局重校准通道权重，结合跨维度交互建模像素级关系，融合通道与空间特征，可缓解主干网络在多尺度人脸检测中的不足。

1.5　组建轻量化颈部网络

YOLOv8 中颈部网络涉及多个卷积层和上采样层，复杂的特征融合和上采样操作会增加算法的计算开销。为此，本文在颈部网络中结合 ghost module[12]、动态卷积（dynamic convolution）[13]和 C2f 模块构造混合模块（C2f-ghost-dynamicconv）架构（图 6），并基于此模块组建轻量化的颈部网络，以降低算法的计算开销。

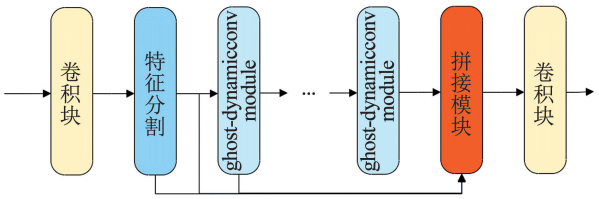


图 6　C2f-ghost-dynamicconv 模块结构

Fig. 6　C2f-ghost-dynamicconv module structure

ghost module 通过减少卷积核数量并利用线性变换生成更多特征图，降低计算复杂度；动态卷积模块则动态调整卷积核参数以适应输入特征。ghost module 保持轻量化但牺牲部分性能，而动态卷积模块可补偿其性能损失。两者结合，可以在维持精度的同时进一步减少计算开销。

ghost module 的输出特征图有两部分：一部分直接由卷积获得，另一部分则由上部分特征图经过线性变换得到，然后将两部分拼接形成最终的输出特征图。对 t 个基于普通卷积得到的原始特征图 $\boldsymbol{Y}'$，其表达式为

$$\boldsymbol{Y}' = \boldsymbol{X} \times \boldsymbol{K} \tag{3}$$

式中：$\boldsymbol{X}$ 为输入特征图，$\boldsymbol{K}$ 为卷积核。

为获得所需的 n 个特征图，对 $\boldsymbol{Y}'$ 中的每个原始特征图进行廉价线性变换，生成 s 个幻影特征图。该过程表达式为

$$\boldsymbol{y}_p = \phi_p(\boldsymbol{y}'_o), \forall o = 1,\cdots,t; p = 1,\cdots,s \tag{4}$$

式中：$\boldsymbol{y}'_o$ 为 $\boldsymbol{Y}'$ 中第 o 个原始特征图，ϕ_p 为线性运算。

动态卷积模块利用多个专家（expert）处理输入特征，并根据输入特征动态调整各 expert 的权重，使算法自适应地选择最合适的卷积核组合，在增加参数的同时保持较低的计算量。其具体实现过程如下：

首先，对输入特征进行全局平均池化；其次，通过全连接层和 Softmax 操作生成注意力权重；最后，将 expert 的卷积核按权重加权聚合为动态卷积核并将其应用于输入特征得到输出。该过程中具有 M 个 expert 的动态卷积表达式为

$$\begin{cases} \boldsymbol{X}' = \rho(\boldsymbol{X}) \\ \alpha = \mathrm{Softmax}(\mu(\boldsymbol{X}')) \\ \boldsymbol{W}_{\mathrm{D}} = \sum_{m=1}^{M} \alpha_m \cdot \boldsymbol{W}_m \\ \boldsymbol{Y} = \boldsymbol{X} \times \boldsymbol{W}_{\mathrm{D}} + v \end{cases} \tag{5}$$

式中：$\boldsymbol{X}'$ 为全局特征，ρ 为全局平均池化过程，α 为注意力权重，μ 为全连接层，$\boldsymbol{W}_{\mathrm{D}}$ 为动态卷积核，α_m 为第 m 个专家的注意力权重，$\boldsymbol{W}_m$ 为第 m 个专家的卷积核，v 为偏置项。

将动态卷积和 ghost module 进行融合，得到 ghost-dynamicconv 模块，结构如图 7 所示。

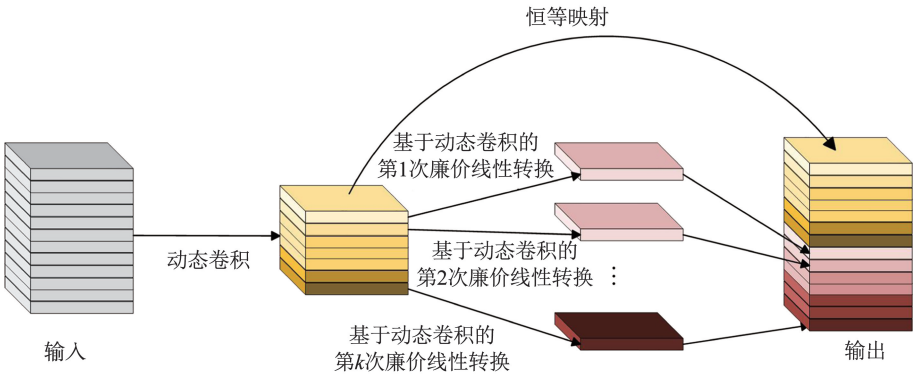


图 7　Ghost-dynamicconv 模块结构

Fig. 7　Ghost-dynamicconv module structure

1.6 重构轻量化检测头

YOLOv8 采用解耦头结构,将分类和边界框回归两个任务分离到不同的分支中,每个检测头包括两个分支,每个分支中有 2 个 3×3 卷积和 1 个 1×1 卷积,该设计有助于提高算法的灵活性和精度,但会增加计算开销。针对此问题,本文采用分组卷积共享参数信息的方式重新构建轻量化检测头[14],其结构如图 8 所示。

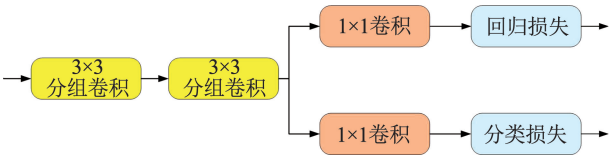


图 8 轻量化检测头模块结构

Fig. 8 Lightweight detection head module structure

轻量化检测头中的分组卷积将输入特征图分为多组,每组内部独立进行常规卷积操作。同时,采用参数共享的方式减少每个检测头中的计算开销。分组卷积的结构如图 9 所示。

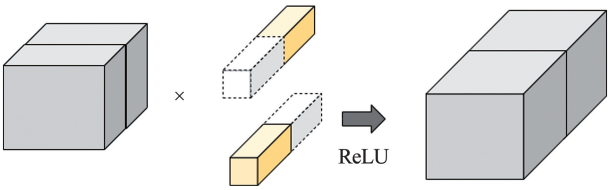


图 9 分组卷积模块结构

Fig. 9 Group convolution module structure

分组卷积将输入的 N 个特征图依据通道维度划分为 g 组。划分过程可表示为

$$\boldsymbol{X} \in [\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{g-1}] \tag{6}$$

然后,在每组内进行独立的卷积操作。分组卷积的参数量可表示为

$$P = \frac{C_0 \times C_1 \times e^2}{g} \tag{7}$$

式中:C_0 和 C_1 分别为输入、输出特征图的通道数,e 为卷积核的大小。

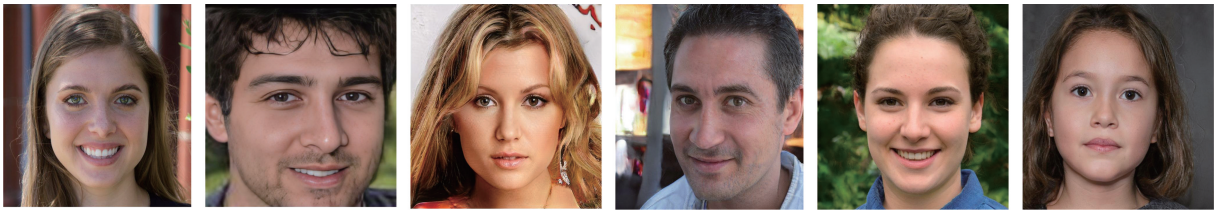
轻量化检测头中,标准卷积转变为两个 3×3 分组卷积。由于分组卷积不同组的卷积核专注于输入特征图的不同部分,使得算法能够学习更加多样化的特征表示,可以更好地捕捉人脸目标的特征。

2 实验结果及分析

2.1 数据集的制作与评价指标

本研究实验所用数据集基于英伟达的 PGGAN、StyleGAN 系列模型以及微软的 DiscoFaceGAN 模型生成[15-19]。PGGAN 通过渐进式训练生成高分辨率的图像,同时提高训练过程的稳定性和效率。StyleGAN 系列网络凭借渐进式生成策略、解耦的特征表示以及精细的层次化特征控制技术,能够实现高分辨率且逼真的人脸图像生成。DiscoFaceGAN 则通过模仿学习、对比学习和 3D 先验知识实现高度解缠和精确可控的人脸图像生成。真实人脸与生成人脸示例如图 10 所示。

由图 10 可以看出,DiscoFaceGAN 在控制表情、姿势等 3D 人脸参数方面具有较高的精确度,能够通过 3D 参数精确控制面部表情,但图像的逼真度稍显不足,面部结构、皮肤纹理、毛发以及人脸周围环境的处理细节相对模糊。PGGAN 生成的人脸图像质量较高,面部结构清晰,但仍存在一些问题,如生成的人脸有时会出现不规则形状或扭曲,面部边缘与背景的过渡不够自然,皮肤和头发纹理较为模糊,发丝细节不清晰。StyleGAN 使面部特征更加灵活自然,面部轮廓和五官比例更为协调,但在极端视角下偶尔会出现形变;皮肤细节和光影处理更加细腻,但在极端光照条件下可能出现不协调现象;头发的细节有所提升,发丝纹理和光泽感较为自然,但在处理复杂发型时仍有一定模糊感。StyleGAN2 解决了伪影问题,生成的人脸图像更加稳定和逼真,面部结构的精确度更高,轮廓平滑,细节自然,皮肤质感更为逼真,光影过渡流畅,反射效果接近真实,头发的质感和细节也显著提升,发丝分离效果更好,光泽感更为逼真。StyleGAN3 进一步增强了生成人脸时的空间一致性,即使在不同视角和动态变化下,生成的图像依然连贯一致,几何一致性进一步提升,面部在不同视角下保持稳定连贯,光照和皮肤纹理在不同角度和动态变化中表现出更高的一致性,头发细节在多视角和动态变化下也保持一致,复杂发型如卷发的表现更加自然。



(a) real[20] (b) DiscoFaceGAN[19] (c) PGGAN[15] (d) StyleGAN[16] (e) StyleGAN2[17] (f) StyleGAN3[18]

图 10 真实人脸及生成人脸示例

Fig. 10 Examples of real faces and generative faces

综上所述，本研究的数据集中数据来源分为 6 部分。其中，真实人脸来源于 FFHQ 人脸数据集[20]。生成人脸数据集分别由 PGGAN、StyleGAN、StyleGAN2、StyleGAN3 以及 DiscoFaceGAN 生成。经过手动筛选和数据清洗后，得到 12 000 张图像，每一类 2 000 张。为了减少算法的训练时间成本，将每两张人脸图像拼接成一张图像，最终得到 6 000 张图像。然后对数据进行标注，标注为 PASCAL VOC 格式，标注类别为真实人脸（real）、DiscoFaceGAN 生成人脸（fake_DiscoFaceGAN，fDFG）、PGGAN 生成人脸（fake_PGGAN，fPG）、StyleGAN 生成人脸（fake_StyleGAN，fSG）、StyleGAN2 生成人脸（fake_StyleGAN2，fSG2）和 StyleGAN3 生成人脸（fake_StyleGAN3，fSG3）。样本集中人脸目标数目分别为 2 212、2 110、2 071、2 012、2 040 和 2 010 个，并以 8∶1∶1的比例划分为训练集、验证集及测试集。

以平均精度（P_A）、均值平均精度（P_{MA}）、参数量（P_{AR}）、帧率（P_{FPS}）、浮点运算量（P_{FPO}）作为评价指标。

2.2　实验环境与参数设置

本研究基于 Ubuntu 22.04 操作系统，利用 PyTorch 深度学习框架及 Python 编程语言开展全部实验。GPU 为 NVIDIA GeForce RTX 3090，CUDA 版本为 11.6，PyTorch 版本为 1.13.0。输入图像尺寸设置为 640×640 像素，批量大小为 8，最大迭代次数为 250。

2.3　消融实验

为验证本文提出 5 种改进方法的有效性，以 YOLOv8 为基准进行消融实验。实验设计及结果见表 1 和表 2。

表 1　消融实验设计

Tab. 1　Ablation experimental design

分组	基线	逐通道卷积	C2f-DWR	EMA	C2f-ghost-dynamicconv	轻量化检测头
1	√	×	×	×	×	×
2	√	√	×	×	×	×
3	√	√	√	×	×	×
4	√	√	√	√	×	×
5	√	√	√	√	√	×
6	√	√	√	√	√	√

注：√表示采用该改进方法，×表示未采取该改进方法。

表 2　消融实验结果

Tab. 2　Ablation experimental results

分组	P_A/%						P_{MA}/%	$P_{AR}/10^6$	P_{FPS}/(帧·s^{-1})	$P_{FPO}/10^9$
	real	fDFG	fPG	fSG	fSG2	fSG3				
1	94.72	98.35	97.04	91.65	89.85	91.76	93.89	11.13	87.72	28.60
2	94.97	96.41	96.98	91.98	88.66	91.00	93.33	8.85	98.04	24.00
3	98.52	98.56	97.65	92.98	91.39	91.46	95.09	8.68	82.64	23.90
4	97.76	98.58	97.24	93.93	91.52	93.65	95.45	8.72	84.75	24.20
5	96.39	98.47	97.35	94.47	90.10	93.29	95.01	7.10	88.50	20.40
6	97.63	98.08	97.76	93.70	90.90	92.76	95.14	5.30	90.10	13.30

由表 2 可知：在整体网络中引入逐通道卷积替换原有卷积模块后，算法的 P_{MA} 降低了 0.56%，P_{AR} 减少了 2.28×10^6，P_{FPS} 提升了 10.32 帧/s，原因是逐通道卷积可以保持特征独立性的同时降低算法的计算开销，减少了卷积运算的复杂度，削弱了算法的特征提取能力，降低了检测精度；引入 C2f-DWR 模块后，算法的 P_{MA} 提高了 1.76%，P_{AR} 降低了 0.17×10^6，P_{FPS} 降低了 15.40 帧/s，原因是 DWR 模块能够获取多尺度上下文信息以及扩展感受野，提高了算法细节特征的表达能力，但其引入了计算方式更复杂的多速率扩展深度卷积，降低了检测速度；引入 EMA 后，算法的 P_{MA} 提高了 0.36%，P_{AR} 增加了 0.04×10^6，P_{FPS} 提升了 2.11 帧/s，原因是 EMA 的特征分组和多尺度结构能够有效建立短程和长程依赖关系，增强了算法的特征表示能力，并且基于其高效的设计，有效地减少了计算开销，提升了算法的检测速度；引入基于 ghost module 和动态卷积模块构建的混合架构后，算法的 P_{MA} 降低了 0.44%，P_{AR} 减少了 1.62×10^6，P_{FPS} 提升了 3.75 帧/s，原因是结合 ghost module 减少卷积核数量和动态卷积动态调整卷积核参数的特点，使得算法在降低参数量的同时，保持了检测精度；采用分组卷积共享参数信息的方式重构检测头，算法的 P_{MA} 提高了 0.13%，P_{AR} 减少了 1.80×10^6，P_{FPS} 提升了 1.60 帧/s，说明本文设计的轻量化检测头减少了算法的参数量，同时提升了算法的检测精度。相较于基线算法，MSERF-YOLOv8 算法的 P_{MA} 提升了 1.25%，P_{AR} 减少了 5.83×10^6，P_{FPS} 提升了 2.38 帧/s，在检测精度提升的同时将参数量压缩至原算法的 47.62%，说明本文算法兼顾了检测速度与检测精度的平衡，能够满足实际的生成人脸检测需求。

2.4 对比实验

2.4.1 不同注意力机制模块对比

为验证所引入注意力机制模块的有效性,设置了对比实验。以引入逐通道卷积和构建 C2f-DWR 模块的网络为基准,选择简单无参卷积神经网络注意力模块(SimAM)、全局注意力机制(GAM)模块、坐标注意力(CA)模块、压缩-激励注意力(SE)模块、基于归一化的注意力模块(NAM)与 EMA 在相同条件下进行对比,结果见表 3。

表 3 不同注意力机制对比结果

Tab. 3 Comparison results of different attention mechanisms

注意力机制模块	P_A/%						P_{MA}/%	$P_{AR}/10^6$
	real	fDFG	fPG	fSG	fSG2	fSG3		
SimAM	96.54	98.18	96.70	93.06	90.71	91.83	94.51	**8.68**
GAM	**98.47**	98.26	**97.68**	92.55	91.38	92.51	95.14	10.42
CA	95.45	98.10	96.23	92.25	91.68	92.61	94.39	**8.68**
SE	98.27	98.48	97.37	93.71	91.69	91.36	95.15	**8.68**
NAM	97.44	98.23	97.21	93.36	**91.99**	92.67	95.15	**8.68**
EMA	97.76	**98.58**	97.24	**93.93**	91.52	**93.65**	**95.45**	8.72

由表 3 可知,相比于 SimAM、GAM、CA、SE 和 NAM,添加 EMA 后,算法的 P_{MA} 最高,表明 EMA 能够更有效地提升算法的检测精度。在生成人脸检测任务中,SimAM 基于局部自相似性生成注意力权重,在处理多尺度信息方面相对较弱;GAM 关注全局信息整合,但在处理人脸多尺度特征方面略显不足,且计算开销较大;CA 虽结合通道注意力和坐标信息,但在处理多尺度特征方面不及 EMA 全面;SE 主要关注通道间的特征重标定,并不涉及多尺度特征融合,因此,在生成人脸检测方面不如 EMA;NAM 主要通过归一化的方式提升检测效果,在处理多尺度人脸方面稍有欠缺;而 EMA 注意力机制通过并行子网络能够有效捕捉和处理人脸图像中的多尺度特征,使得算法能够更全面地理解图像信息,进而提高对生成人脸的检测精度。综上所述,本文选择在主干网络 SPPF 后引入 EMA,提升了算法整体的检测能力。

2.4.2 不同算法对比

为进一步验证 MSERF-YOLOv8 算法的有效性,将其与更快速的区域卷积神经网络(Fast R-CNN)[21]、单发多框检测器(SSD)[22]、焦点损失检测网络(Retina-Net)[23]、中心点网络(Center-Net)[24]、全卷积单阶段目标检测器(FCOS)[25]、高效检测器(EfficientDet)[26]、检测 Transformer(DETR)[27]、YOLOv3[28]、YOLOv4[29]、YOLOv5[30]、YOLOv7[31]、YOLOv8[32]、YOLOX[33]、Efficientnet-YOLOv3、YOLOv4-tiny、Mobilenet-YOLOv4、YOLOv7-tiny 以及 3 种改进算法[34-36]进行对比,实验结果见表 4。

表 4 MSERF-YOLOv8 算法与不同目标检测算法对比实验结果

Tab. 4 Experimental results of MSERF-YOLOv8 algorithm compared with different target detection algorithms

算法	P_A/%						P_{MA}/%	$P_{AR}/10^6$	P_{FPS}/(帧·s^{-1})	$P_{FPO}/10^9$
	real	fDFG	fPG	fSG	fSG2	fSG3				
Fast R-CNN	73.46	98.01	89.14	69.19	70.23	82.56	80.52	136.75	25.85	401.76
SSD	88.22	98.52	92.16	83.22	84.83	86.73	88.95	24.01	60.89	274.49
Retina-Net	83.77	98.52	91.60	79.17	76.35	83.63	85.51	36.35	37.40	164.55
Center-Net	75.08	98.52	87.73	72.81	72.05	70.60	79.47	32.67	67.46	109.71
FCOS	92.42	99.01	93.89	79.82	88.45	86.82	90.07	32.11	41.59	161.25
EfficientDet	73.14	98.51	92.40	84.06	90.14	91.00	88.21	**3.83**	16.59	**4.76**
DETR	86.15	98.60	92.73	88.84	82.75	92.69	90.30	36.74	23.17	73.62
YOLOv3	65.71	99.98	86.14	61.94	79.03	81.87	79.11	61.54	46.80	155.31
YOLOv4	39.54	98.49	77.71	25.89	33.96	70.96	57.76	63.94	29.35	141.95
YOLOv5	81.80	99.01	92.47	74.31	79.98	80.05	84.60	7.24	52.85	16.50
YOLOv7	82.92	99.01	95.41	78.44	72.35	80.00	84.69	37.20	43.06	105.17
YOLOv8	94.72	98.35	97.04	91.65	89.85	91.76	93.89	11.13	87.22	28.60
YOLOX	94.09	**99.46**	96.69	90.67	89.61	91.39	93.65	8.94	43.63	16.80
Efficientnet-YOLOv3	59.21	98.52	82.64	77.79	62.87	84.04	77.51	7.01	28.33	**3.83**
YOLOv4-tiny	78.03	98.52	91.52	69.86	68.61	70.42	79.49	5.88	**135.31**	16.19
Mobilenet-YOLOv4	66.98	98.51	84.45	55.89	63.44	71.02	73.38	12.28	53.77	23.91
YOLOv7-tiny	82.22	99.01	90.45	78.46	78.27	86.59	85.83	6.03	74.18	13.22
改进 YOLOv5s	82.48	98.59	92.57	76.23	80.22	84.35	85.74	**2.42**	44.29	**4.20**

表 4(续)

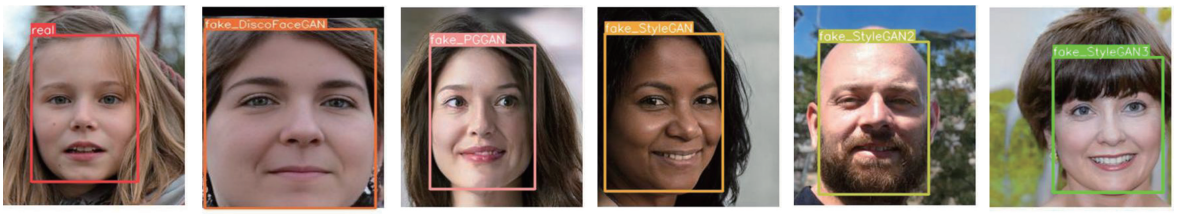
算法	P_A/%						P_{MA}/%	$P_{AR}/10^6$	P_{FPS}/(帧·s^{-1})	$P_{FPO}/10^9$
	real	fDFG	fPG	fSG	fSG2	fSG3				
改进 YOLOv7	84.59	98.55	95.12	82.11	75.98	84.99	86.89	34.90	52.74	96.50
改进 YOLOv8s	96.02	98.39	97.55	92.01	90.04	92.09	94.35	11.13	80.14	28.60
MSERF-YOLOv8	**97.63**	98.08	**97.76**	**93.70**	**90.90**	**92.76**	**95.14**	5.30	90.10	13.30

由表 4 可知：与二阶段检测算法 Faster R-CNN 相比，MSERF-YOLOv8 的 P_{MA}、P_{FPS} 分别提高了 14.62%、64.25 帧/s，P_{AR} 降低了 131.45 $\times 10^6$，P_{FPO} 减少了 388.46 $\times 10^9$；与 Transformer 系列检测算法中的 DETR 相比，MSERF-YOLOv8 的 P_{MA}、P_{FPS} 分别提高了 4.84%、66.93 帧/s，P_{AR} 降低了 31.44 $\times 10^6$，P_{FPO} 减少了 60.32 $\times 10^9$；与一阶段有锚框检测算法中 P_{MA} 最高的 YOLOv8 相比，MSERF-YOLOv8 的 P_{MA}、P_{FPS} 分别提高了 1.25%、2.88 帧/s，P_{AR} 降低了 5.83 $\times 10^6$，P_{FPO} 减少了 15.30 $\times 10^9$；与一阶段无锚框检测算法中 P_{MA} 最高的 YOLOX 相比，MSERF-YOLOv8 的 P_{MA}、P_{FPS} 分别提高了 1.49%、46.47 帧/s，P_{AR} 降低了 3.64 $\times 10^6$，P_{FPO} 减少了 3.50 $\times 10^9$。EfficientDet 的 P_{AR}、P_{FPO} 和 YOLOv4-tiny 的 P_{FPS} 与上述算法相比有较大的优势，但两者的 P_{MA} 损失较大；与 3 种改进算法相比，MSERF-YOLOv8 的 P_{MA} 和 P_{FPS} 表现均最优，P_{AR} 和 P_{FPO} 仅次于改进 YOLOv5 算法，但其他指标优于改进 YOLOv5 算法。综合考虑所有指标，MSERF-YOLOv8 表现最佳，不仅降低了计算开销，同时提升了检测精度和速度。

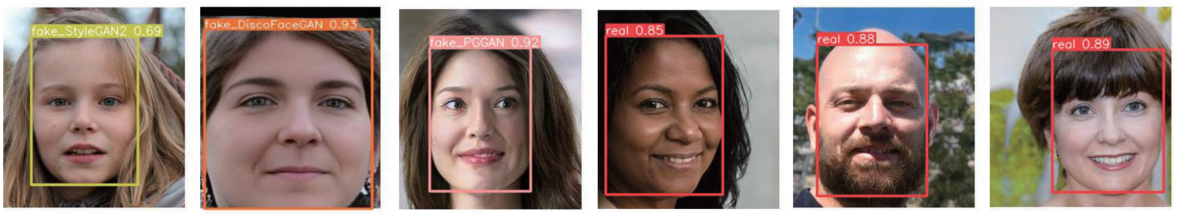
2.5　检测效果分析

为验证改进后算法相较于原算法在生成人脸检测方面的优势，分别选择基线算法和改进后的算法进行生成人脸检测并对比。检测时，采用矩形框标注目标，标签旁的数字为检测框的置信度。检测结果和检测精度如图 11 和图 12 所示。其中，图 11(a)和图 12(a)是通过 LabelImg 标注的原始图像及其对应标签。

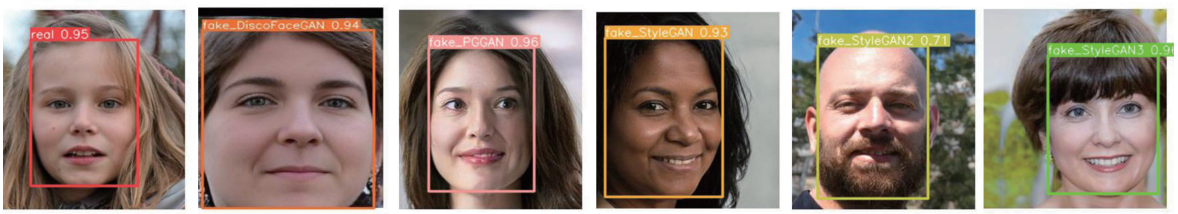
由图 11 可知，YOLOv8 在检测真实人脸和生成人脸过程中存在误检现象，而 MSERF-YOLOv8 对真实人脸和生成人脸的检测结果是准确的。相比于 YOLOv8，MSERF-YOLOv8 改善了误检现象，提升了生成人脸检测的准确度。由图 12 可知，YOLOv8 对一些难以辨识的生成人脸的检测精度较低，而 MSERF-YOLOv8 基于 YOLOv8 进行了改进，其对生成人脸的检测精度普遍高于 YOLOv8。



(a) Original



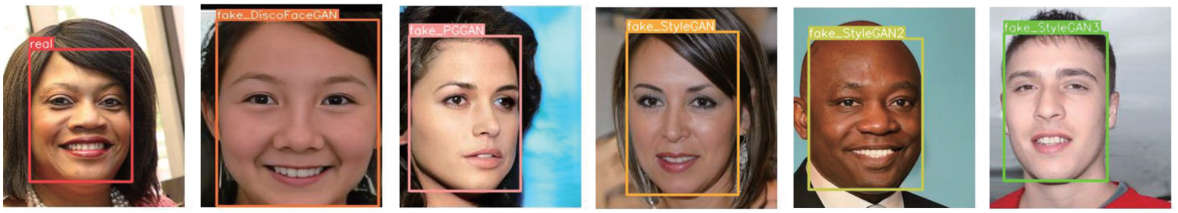
(b) YOLOv8



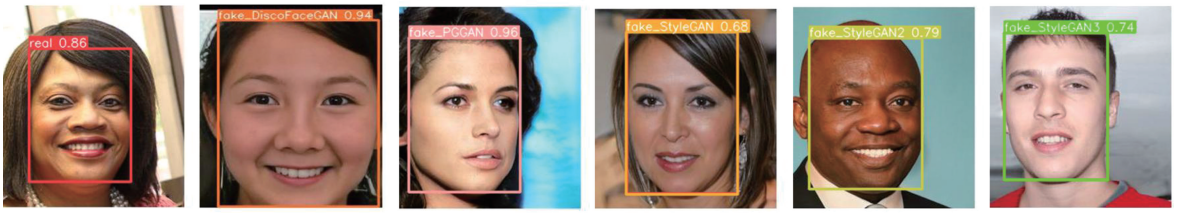
(c) MSERF-YOLOv8

图 11　算法改进前后检测结果对比

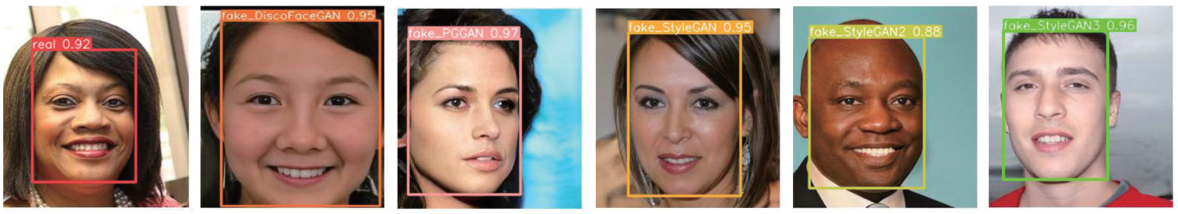
Fig. 11　Comparison of detection results before and after algorithm improvement



(a) Original



(b) YOLOv8



(c) MSERF-YOLOv8

图 12　算法改进前后检测精度对比

Fig. 12　Comparison of detection accuracy before and after algorithm improvement

3　结　论

1)基于 YOLOv8 提出了一种多尺度融合与感受野增强的轻量化人脸检测算法 MSERF-YOLOv8,该算法构建了轻量化整体网络架构,加强了主干网络高层特征提取,设计了多尺度主干网络,组建了轻量化颈部网络以及重构了轻量化检测头。

2)本文所提算法对真实人脸和生成人脸检测的 P_{MA} 达到了 95.14%,相较于原算法提高了 1.25%,P_{AR} 减少了 5.83×10^6,P_{FPS} 提升了 2.88 帧/s。相较于经典目标检测算法和改进的算法,综合考虑所有评价指标,本算法表现最佳,不仅提高了特征提取能力,降低了计算量,还提升了检测精度和速度,能够满足在性能较差的边缘设备上生成人脸检测任务的需求,具有实际的应用价值。

3)所提算法能够实现对生成人脸目标快速且精准的检测,在训练过程中,其对 DiscoFaceGAN 生成的人脸 P_A 要低于部分检测算法;同时 P_{AR}、P_{FPS}、P_{FPO} 要略逊于一些算法。为此,下一步将致力于优化算法的 P_{AR}、P_{FPS} 及 P_{FPO},使其更好地满足生成人脸检测任务的需求。

参考文献

[1]王鹏. 基于卷积神经网络的局部 GAN 生成人脸检测算法研究[D]. 南京: 南京信息工程大学, 2023: 1
Wang Peng. Research on local GAN generation face detection algorithm based on convolutional neural network [D]. Nanjing: Nanjing University of Information Science & Technology, 2023: 1

[2]王富平, 段冠庄, 李藕, 等. 融合边缘和关键点的遮挡人脸修复网络[J/OL]. 哈尔滨工业大学学报, 2024-06-28. https://link.cnki.net/urlid/23.1235.T.20240627.1342.002
Wang Fuping, Duan Guanzhuang, Li Ou, et al. Fusion of edge and landmarks occlusion face inpainting network [J/OL]. Journal of Harbin Institute of Technology, 2024-06-28. https://link.cnki.net/urlid/23.1235.T.20240627.1342.002

[3]Shang Zhihua, Xie Hongtao, Zha Zhengjun, et al. PRRNet: pixel-region relation network for face forgery detection [J]. Pattern Recognition, 2021, 116: 107950. DOI: 10.1016/j.patcog.2021.107950

[4]Chen Beijing, Tan Weijin, Wang Yiting, et al. Distinguishing between natural and GAN-generated face images by combining global and local features [J]. Chinese Journal of Electronics, 2022, 31(1): 59. DOI: 10.1049/cje.2020.00.372

[5]Mansourifar H, Shi Weidong. One-shot gan generated fake face detection[PP/OL]. arXiv (2020-05-27) [2024-09-01]. https://doi.org/10.48550/arXiv.2003.12244

[6]曾可慧. 基于卷积神经网络的 GAN 生成人脸检测算法研究[D]. 南京: 南京信息工程大学, 2022: 16
Zeng Kehui. Research on GAN-generated face detection algorithm based on convolutional neural network [D]. Nanjing: Nanjing University of Information Science & Technology, 2022: 16

[7]王金伟, 曾可慧, 张家伟, 等. 基于空频联合卷积神经网络的 GAN 生成人脸检测[J]. 计算机科学, 2023, 50(6): 216
Wang Jinwei, Zeng Kehui, Zhang Jiawei, et al. GAN-generated face detection based on space-frequency convolutional neural network[J]. Computer Science, 2023, 50(6): 216. DOI: 10.11896/jsjkx.220400268

[8]陈北京, 王鹏, 喻乐延, 等. 注意力融合双流特征的局部 GAN

生成人脸检测算法[J]. 东南大学学报(自然科学版), 2023, 53(3): 543
Chen Beijing, Wang Peng, Yu Leyan, et al. Locally GAN-generated face detection algorithm based on dual-stream features fused by attention[J]. Journal of Southeast University (Natural Science Edition), 2023, 53(3): 543. DOI: 10.3969/j.issn.1001-0505.2023.03.020

[9] Howard A G, Zhu Menglong, Chen Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[PP/OL]. V1. arXiv (2017-04-17) [2024-09-01]. https://doi.org/10.48550/arXiv.1704.04861

[10] Wei Haoran, Liu Xu, Xu Shouchun, et al. DWRSeg: rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation[PP/OL]. V3. arXiv (2023-09-13) [2024-09-01]. https://doi.org/10.48550/arXiv.2212.01173

[11] Ouyang Daliang, He Su, Zhang Guozhong, et al. Efficient multi-scale attention module with cross-spatial learning[PP/OL]. V1. arXiv (2023-05-23) [2024-09-01]. https://doi.org/10.48550/arXiv.2305.13563

[12] Han Kai, Wang Yunhe, Tian Qi, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 1580. DOI: 10.1109/CVPR42600.2020.00165

[13] Han Kai, Wang Yunhe, Guo Jianyuan, et al. ParameterNet: parameters are all you need for large-scale visual pretraining of mobile networks[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 15751. DOI: 10.1109/CVPR52733.2024.01491

[14] Xie Saining, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5987. DOI: 10.1109/CVPR.2017.634

[15] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variati[PP/OL]. V3. arXiv (2018-02-26) [2024-09-01]. https://doi.org/10.48550/arXiv.1710.10196

[16] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4401. DOI: 10.1109/TPAMI.2020.2970919

[17] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of stylegan[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 8110. DOI: 10.1109/CVPR42600.2020.00813

[18] Karras T, Aittala M, Laine S, et al. Alias-free generative adversarial networks[PP/OL]. V4. arXiv (2021-10-18) [2024-09-01]. https://doi.org/10.48550/arXiv.2106.12423

[19] Deng Yu, Yang Jiaolong, Chen Dong, et al. Disentangled and controllable face image generation via 3D imitative-contrastive learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5154. DOI: 10.1109/CVPR42600.2020.00520

[20] Bai Haoran, Kang Di, Zhang Haoxian, et al. FFHQ-UV: normalized facial UV-texture dataset for 3D face reconstruction[PP/OL]. V2. arXiv (2023-05-24) [2024-09-01]. https://doi.org/10.48550/arXiv.2211.13874

[21] Girshick R. Fast R-CNN[C]//15th IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440. DOI: 10.1109/ICCV.2015.169

[22] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21. DOI: 10.1007/978-3-319-46448-0_2

[23] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//16th IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999. DOI: 10.1109/ICCV.2017.324

[24] Duan Kaiwen, Bai Song, Xie Lingxi, et al. CenterNet: keypoint triplets for object detection[C]//17th IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6568. DOI: 10.1109/ICCV.2019.00667

[25] Tian Zhi, Shen Chunhua, Chen Hao, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922. DOI: 10.1109/TPAMI.2020.3032166

[26] Tan Mingxing, Pang Ruoming, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781. DOI: 10.1109/CVPR42600.2020.01079

[27] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//16th European Conference on Computer Vision. Glasgow: Springer, 2020: 213. DOI: 10.1007/978-3-030-58452-8_13

[28] Redmon J, Farhadi A. YOLOv3: an incremental improvement[PP/OL]. V1. arXiv (2018-04-08) [2024-09-01]. https://doi.org/10.48550/arXiv.1804.02767

[29] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[PP/OL]. V1. arXiv (2020-04-23) [2024-09-01]. https://doi.org/10.48550/arXiv.2004.10934

[30] Yan Bin, Fan Pan, Lei Xiaoyan, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021, 13(9): 1619. DOI: 10.3390/rs13091619

[31] 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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464. DOI: 10.1109/CVPR52729.2023.00721

[32] Kang C H, Kim S Y. Real-time object detection and segmentation technology: an analysis of the YOLO algorithm[J]. JMST Advances, 2023, 5(2): 69. DOI: 10.1007/s42791-023-00049-7

[33] Ge Zheng, Liu Songtao, Wang Feng, et al. YOLOX: exceeding YOLO series in 2021[PP/OL]. V1. arXiv (2021-07-18) [2024-09-01]. https://doi.org/10.48550/arXiv.2107.08430

[34] 贺帅, 李锐. 基于轻量化改进 Yolov5s 的驾驶员人脸检测算法[J]. 长江信息通信, 2023, 36(8): 43
He Shuai, Li Rui. Driver face detection algorithm based on lightweight improved Yolov5s[J]. Changjiang Information & Communications, 2023, 36(8): 43. DOI: 10.3969/j.issn.1673-1131.2023.08.014

[35] 戴莹, 叶贵. 面向智慧养老的改进 YOLOv7 人脸识别算法[J]. 信息工程大学学报, 2024, 25(2): 175
Dai Ying, Ye Gui. Improved YOLOv7 face recognition algorithm for smart pension[J]. Journal of Information Engineering University, 2024, 25(2): 175. DOI: 10.3969/j.issn.1671-0673.2024.02.008

[36] 孙涵, 田野, 孙春凤. 基于改进 YOLOv8s 的密集多人脸检测[J]. 软件, 2024, 45(4): 142
Sun Han, Tian Ye, Sun Chunfeng. A dense multi-face detection based on improved YOLOv8s[J]. Software, 2024, 45(4): 142. DOI: 10.3969/j.issn.1003-6970.2024.04.040