复杂背景下小尺寸多角度人脸检测方法研究*

黄　杰，刘　芬

（天津职业技术师范大学　电子工程学院，天津　300222）

摘　要：为了提升复杂背景下小尺寸人脸检测精度，提出了一种人脸检测方法 GhostNet-MTCNN。在多任务级联卷积神经网络（MTCNN）主干网络上，将占用计算资源的普通卷积进行舍弃，利用 GhostNet 网络中计算量更低的 Ghost bottleneck 模组替代卷积的作用，重新构建网络特征提取功能，从而搭建一个新的模型。实验结果表明，该方法可以有效平衡参数量和精度。在 Easy、Medium、Hard 三种验证集上，与 MTCNN 相比在参数量仅增加 0.62M 的前提下精度分别提升了 5.6%、6.6%、7.8%，与 MobileNetV3-MTCNN 相比在参数量减少 1.27M 的同时精度又分别提升了 1.6%、0.8%、0.5%。该研究能够在复杂场景下提高模型对小尺寸、多角度人脸检测精度，同时也能够有效平衡参数量和检测精度使其成为在边缘设备部署中更优的选择。

关键词：人脸检测；多任务级联卷积神经网络；轻量化网络；边缘设备

中图分类号：TP18　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2024.04.008

引用格式：黄杰，刘芬. 复杂背景下小尺寸多角度人脸检测方法研究［J］. 网络安全与数据治理，2024，43(4)：46-52.

Research on small-scale, multi-angle face detection methods in complex backgrounds

Huang Jie, Liu Fen

(School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

Abstract: A face detection approach which is named GhostNet-MTCNN was proposed to enhance the precision of small sized face detection in complex backgrounds. On the backbone of MTCNN, this approach uses the lower computational Ghost bottleneck module which is in the GhostNet to replace the convolutional function, and discards the common convolution which occupies computer resources to configure the network's feature extraction function. Through the process, a new module will be set up. The experimental results showed that the approach can effectively balance parameter quantity and precision. Across three validation sets categorized as Easy, Medium and Hard, compared to the original MTCNN, the proposed GhostNet-MTCNN achieves notable improvements in accuracy respectively 5.6%, 6.6% and 7.8%, while the parameter quantity only with a minimal increase of 0.62M. Furthermore, compared to MobileNetV3-MTCNN, GhostNet-MTCNN outperforms by enhancing accuracy by 1.6%, 0.8% and 0.5%, meanwhile a reduction in parameter quantity by 1.27M. The study can not only enhance the precision of the module to detect the small-sized and multi-angle faces in complex backgrounds but also can effectively balance parameter quantity and detection precision, which will make it a superior choice for edge deployment devices.

Key words: face detection; multi-task cascaded convolutional networks; lightweight network; edge devices

0　引言

人脸检测技术广泛应用于考勤、解锁设备、身份验证、监控场所、自动驾驶等场合[1-3]。在当前的人脸检测领域，通常采用深度神经网络架构。2014 年 Girshick 等人提出的 R-CNN[4] 目标检测算法模型成功地将深度学习应用到目标检测领域，这种目标检测算法使用的是基于候选区域的检测方法。Ren 等人在 FastR-CNN 基础上进行改进，提出了 FasterR-CNN[5]，该模型提出了专门的候选区域生成网络。除了以上两种目标检测网络模型外，还有基于单次目标检测的网络模型，如 YOLO[6-8] 和 SSD[9]。这类方法优势在于检测速度快，但对小目标的检测效果不佳。这些深度神经网络在边缘设备部署十分消

*基金项目：教育部产学合作协同育人项目（202002050030）

耗资源，对于硬件的计算能力和能耗的要求很高，很难应用到实际场景中。多任务级联卷积神经网络（Multi-task Cascaded Convolutional Networks，MTCNN）[10] 作为一种经典的人脸检测方法，以其高效的性能、模型复杂度低而闻名，更适合边缘设备的应用。但随着人脸检测任务的不断复杂化，MTCNN 也面临一系列挑战，例如在小尺寸、遮挡、多角度和光照变化等情况下的检测效果下降。文献［11］中将 MTCNN 与 VGGNet 相结合，提升了模型检测精度，但是相对应的模型计算量也变多了。文献［12］将 MobileNet 与 MTCNN 相结合提出 MobileMTCNN，该方案虽然降低了网络所需浮点数的运算，但同时也会导致模型检测精度的下降。

针对以上问题，本文提出一种新的模型 GhostNet-MTCNN，以 MTCNN 为主干网络使用 GhostNet[13] 中 Gbneck 块重构特征提取网络。通过这些改进，该模型在复杂场景下的人脸检测任务中展现出优于 MTCNN 的性能，有效提升了人脸检测的准确性和鲁棒性。此外，本文模型在保证精度的同时，通过优化模型结构和参数配置，实现了模型参数量的显著降低，使其更加适合在嵌入式系统、移动设备等资源受限的环境中运行。这一特性使得本文方法在地铁、商场等人员流动密集的场景中具有广阔的应用前景，有助于推动神经网络技术在现实生活中的广泛部署和实际应用。

1　方法介绍

1.1　MTCNN 算法原理

在深度学习中，MTCNN 是一种常用的人脸检测模型，其主要设计理念是通过级联多个卷积神经网络逐步实现人脸检测的任务，模型检测流程如图 1 所示。MTCNN 由三个网络组成，分别为 P-Net、R-Net、O-Net，网络输入采用图像金字塔的形式，将原始输入图片缩放成一系列不同尺寸的图像，然后送入 P-Net 进行处理产生大量的候选框。将 P-Net 生成的人脸候选框在原图上进行剪裁，然后再将其送入 R-Net 进行处理，R-Net 网络对大量的候选框进行精简，删除非人脸的候选框，修正由 P-Net 网络生成的人脸框坐标和置信度。最后将 R-Net 网络生成结果送入 O-Net 网络处理，输出整个模型的检测结果。

1.2　GhostNet 网络

GhostNet 是一种新型轻量级神经网络架构，能更好地在边缘设备上进行高效计算，通过廉价的操作生成更多的特征图。GhostNet 通过 Ghost 模块进行构建，如图 2 所示。Ghost 模块将原始特征图通过少量的卷积核生成部分特征图，将计算得到的部分特征图经过一系列的线性变换得到新的特征图，将卷积操作产生的特征图和线性变化生成的特征图进行叠加得到最终的输出特征图。

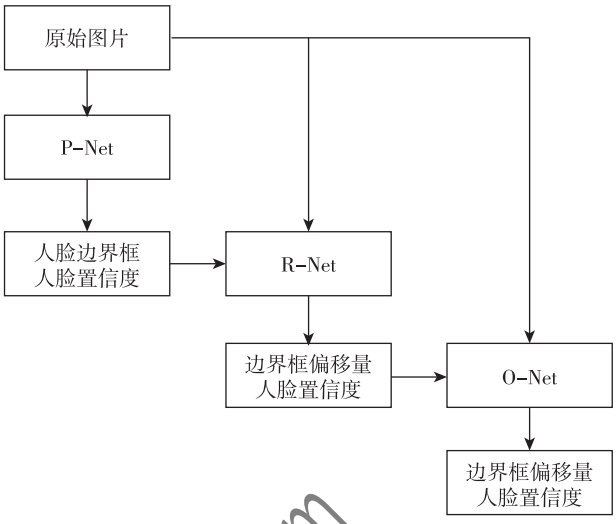


图 1　MTCNN 检测流程

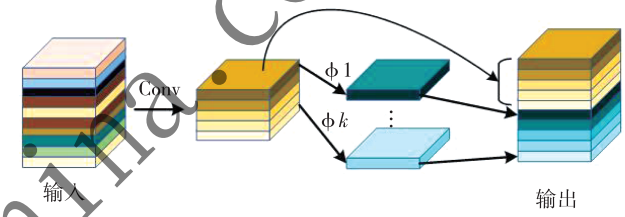


图 2　Ghost 模块

假设输入通道为 c，特征图高和宽为 h 和 w，输出数据的高和宽为 h' 和 w'，传统卷积核的数量为 n，传统卷积核大小为 k，线性变换卷积核大小为 d，变换数量为 s。理论上，使用 Ghost 卷积替换传统卷积的参数压缩比推算如式（1）所示：

$$\mathrm{rc}=\frac{n\times c\times k\times k}{\frac{n}{s}\times c\times k\times k+(s-1)\times\frac{n}{s}\times d\times d}\approx\frac{s\times c}{s+c-1}\approx s \tag{1}$$

加速比推算如式（2）所示：

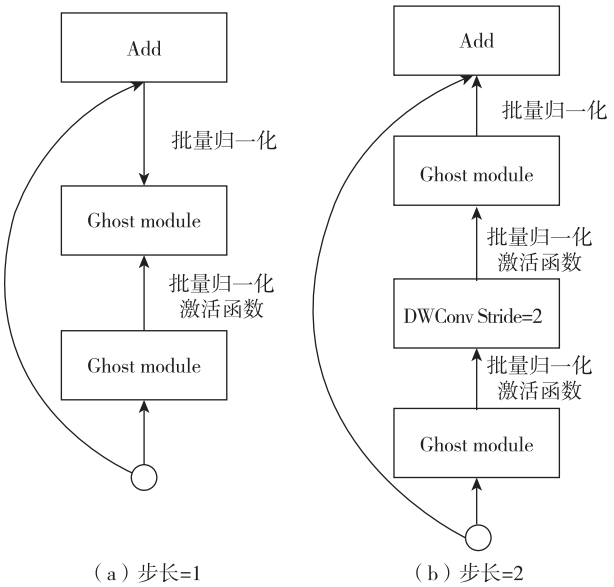
$$\mathrm{rs}=\frac{n\times h'\times w'\times c\times k\times k}{\frac{n}{s}\times h'\times w'\times c\times k\times k+(s-1)\times\frac{n}{s}\times h'\times w'\times d\times d}$$

$$\approx\frac{s\times c}{s+c-1}\approx s \tag{2}$$

从式（1）和式（2）可以看出，计算加速收益和参数压缩效果受变换数量影响，即生成“Ghost”特征图越多加速效果越好，检测精度也会随之下降。为平衡速度与精度一般将变换数量设置为 1/2。

GhostNet 网络由两种不同步长的 Bottleneck 块组成，如图 3 所示。步长为 1 时由两个 Ghost 模块和一个残差边组成，第一个 Ghost 模块用作扩展层，增加了通道数；第二个 Ghost 模块减少通道数。当步长为 2 时，在两个

Ghost 模块间添加 2 × 2 的深度可分离卷积完成宽高压缩操作，其他部分不变。GhostNet 由多个 Ghost Bottleneck 叠加块组成。



（a）步长=1　　（b）步长=2

图 3　Ghost Bottleneck 块

1.3　基于 GhostNet 的特征提取网络改进

本文所提出的 GhostNet-MTCNN 算法利用了 Ghost 模块的优势，以 MTCNN 网络架构为基础，使用 Ghost Bottleneck 块重新构建 P-Net、R-Net、O-Net 三个网络。为了更好地对图片数据进行特征提取，P-Net、R-Net、O-Net 三个网络的第一个卷积层使用普通卷积层，后续卷积层使用 Ghost Bottleneck 块对网络进行替换，其网络结构如图 4 所示。P-Net 输入图片像素大小为 12 × 12 × 3，经过 3 × 3 的普通卷积层和最大池化层的处理变为 5 × 5 × 16，通过步长为 1 和步长为 2 的 Ghost Bottleneck 块进行进一步的特征提取，最后使用 3 × 3 卷积将图片高宽变为 1 × 1 进而计算出人脸分类和边界框。因为 P-Net 网络不包含全连接层，所以网络可以处理不同像素大小的图片，将图片进行一系列缩放操作后送入网络从而生成不同尺寸的人脸候选框。R-Net 输入图片固定像素大小为 24 × 24 × 3，经过 3 × 3 的普通卷积层和最大池化层的特征提取后，图片像素大小变为 10 × 10 × 28，通过一个步长为 1 和两个步长为 2 的 Ghost Bottleneck 块进一步进行特征提取，经过全连接层生成 128 维向量计算出人脸和边界框偏移量，实现了将 P-Net 生成的大量人脸候选框进行精简。O-Net 输入图片固定像素大小为 48 × 48 × 3，经过 3 × 3 的普通卷积层和最大池化层图像像素大小变为 22 × 22 × 32，通过一个步长为 1 和三个步长为 2 的 Ghost Bottleneck 块进行进一步的特征提取，全连接层生成 256 维向量计算出人脸和边界框偏移量，生成模型的最终结果。改进后的模型检测流程如图 5 所示，图片经过 P-Net 处理后生成大量不同尺寸的人脸候选框，将候选框中的人脸剪切后送入 R-Net 网络，进行人脸筛选去除冗余候选框，将判断为存在人脸的候选框剪切后送入 O-Net 网络进行最终的判定。

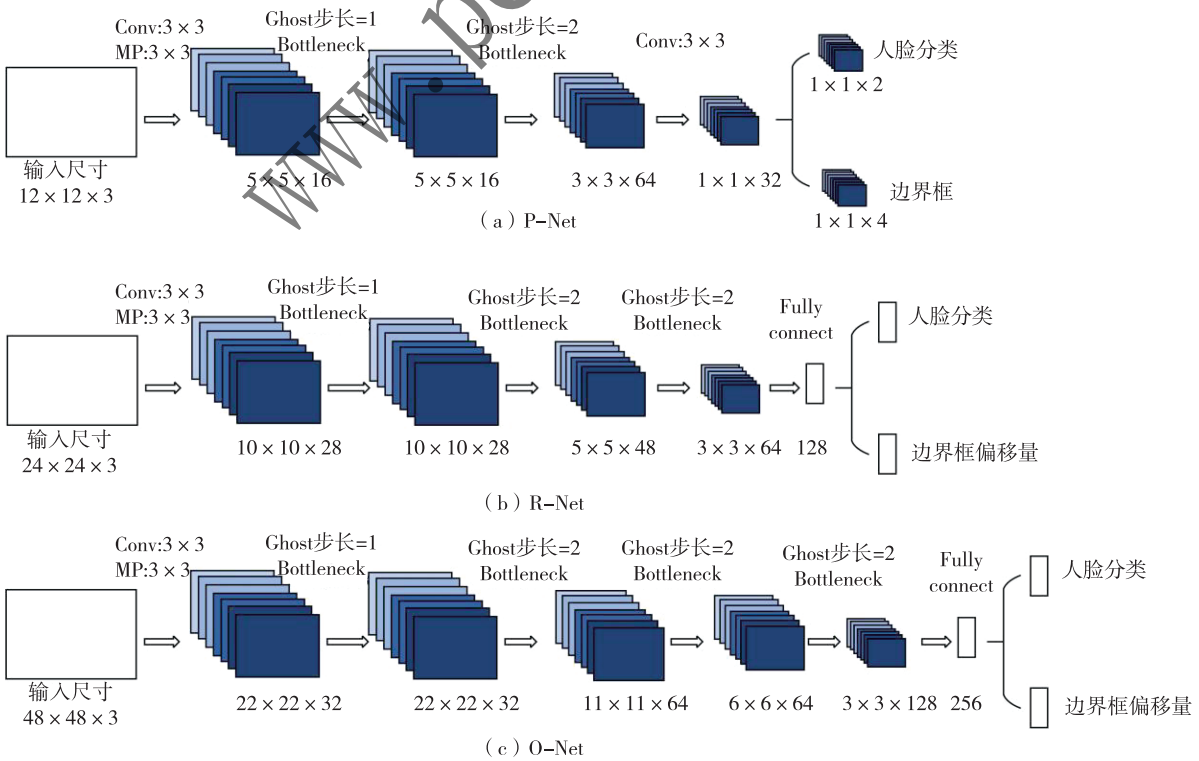


图 4　GhostNet-MTCNN 模型结构

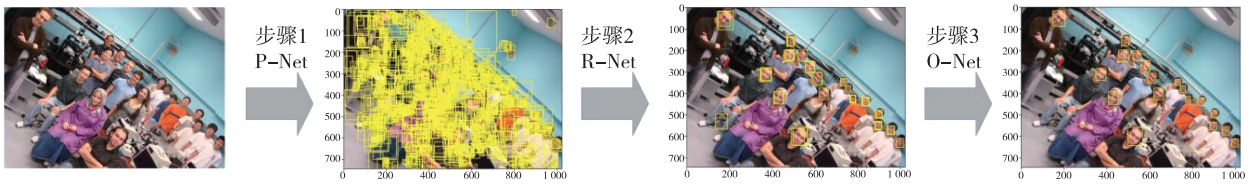


图 5　GhostNet-MTCNN 模型检测流程

2　实验结果分析

2.1　实验参数设置

本文实验测试使用的硬件配置为 Intel（R）Core（TM）i5-13500HX，NVIDIA RTX 4050，运行内存 16 GB，训练框架为 PyTorch 1.10.2，CUDA 12.0，编程语言为 Python 3.6，操作系统为 Windows11。选用 WIDER_ FACE 公开数据集进行实验，其中训练图片 12 793 张，包含不同场景下的人脸数据，实验中部分参数如表 1 所示。

表 1　部分实验参数

参数名	参数说明	参数值
Batchsize	单次训练样本数	[128，640，640]
Learningrate	网络初始学习率	[0.01，0.005，0.05]
Lr_ factor	学习率衰减因子	[0.6，0.1，0.1]
Thresh	每层网络置信度阈值	[0.6，0.7，0.7]
Min_ face	最小人脸尺寸	12

2.2　评估参数

衡量神经网络模型的优劣通常对比模型的 AP（Average Precision）值、精准率（Precision）、召回率（Recall）、模型参数等指标。AP 值是在不同召回率水平上精准率的平均值，它提供了一个分数来总结模型在分类阈值上的性能。精准率是模型正确检测到的人脸数除以模型检测到的总人脸数。高精准率意味着较少的假阳性（即错误地将非人脸识别为人脸）。召回率衡量的是模型识别出的人脸占实际人脸总数的比例，其计算为正确检测到的人脸数除以实际的人脸总数。高召回率意味着较少的假阴性（即漏掉的人脸）。模型参数指构成神经网络模型的可学习元素，如权重和偏置，模型参数的数量通常影响模型的复杂性和计算需求。

本文使用模型参数量、准确率（Precision，P）、召回率（Recall，R）、精度（AP）对模型进行评估，P、R、AP 的表达式如式（3）~（5）所示：

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{5}$$

其中，TP 代表被正确分类的正样本，FP 代表被错误分类的负样本，FN 代表被错误分类的正样本；AP 代表人脸检测的平均精度。

2.3　新方法与其他算法实验结果对比

本文使用相同的数据集和训练参数分别训练了三种模型 Original-MTCNN（原始的 MTCNN 网络）、MobileNetV3-MTCNN（使用 MobileNetV3 中 bneck 模块替换 MTCNN 的主干网络）以及本文所提出的 GhostNet-MTCNN 网络。在 WIDER_ FACE 数据集上使用验证集进行了测试和评估，其中包含 3 200 张图片。为了更好地评估模型的检测效果，将 P-Net、R-Net、O-Net 三个网络的置信度分别设置为 0.7、0.8、0.8。针对 Easy、Medium、Hard 三种不同类型的人脸图片，测试不同模型在人脸检测方面的精准率、召回率等指标。模型的各项评估指标如表 2 ~表 4 所示。

表 2　Easy 验证集人脸类型测试结果

模型	AP/%	召回率/%	精准率/%	参数量/M
Original-MTCNN	64	66.8	89.2	1.88
MobileNetV3-MTCNN	68	74.2	85.5	3.77
GhostNet-MTCNN	69.6	74.8	82.2	2.5

表 3　Medium 验证集人脸类型测试结果

模型	AP/%	召回率/%	精准率/%	参数量/M
Original-MTCNN	61.6	63.8	93.1	1.88
MobileNetV3-MTCNN	67.4	70.9	88.9	3.77
GhostNet-MTCNN	68.2	71.2	92	2.5

表 4　Hard 验证集人脸类型测试结果

模型	AP/%	召回率/%	精准率/%	参数量/M
Original-MTCNN	40.9	41.8	96.8	1.88
MobileNetV3-MTCNN	48.2	50	94.5	3.77
GhostNet-MTCNN	48.7	50.1	93.1	2.5

如表 2～表 4 所示，GhostNet-MTCNN 模型在 AP 值和召回率两项指标上相较于 MTCNN 模型有明显的提升。具体来说，在类型为 Easy 的人脸类型中 AP 值提升了 5.6%，召回率提升了 8%；在类型为 Medium 的人脸类型中 AP 值提升了 6.6%，召回率提升了 7.4%；在类型为 Hard 的人脸类型中 AP 值提升了 7.8%，召回率提升了 8.3%。本文所提出的模型与原始模型相比虽然模型参数增加了 0.62M，但在人脸检测的准确度和全面性上有非常大的提升。与 MobileNetV3-MTCNN 模型相比，本文提出的模型在模型参数上减少了 1.27M，并且对于 Easy、Medium、Hard 三种人脸类别的检测精度依次提高了 1.6%、0.8%、0.5%。与相同参数量的神经网络相比其检测效果更好，与相同检测效果的神经网络相比参数量更少。综上所示，GhostNet-MTCNN 更好地实现了模型的检测精度与模型参数之间的平衡，是边缘设备部署更优的一种选择。

为了更直观地看到模型的检测效果，选择小尺寸单一角度人脸以及小尺寸多角度人脸的图片进行模型测试。图 6、图 7 分别对比了三种模型在同一张图片下的检测结果。图 6 展现了在小尺寸单一角度人脸下的检测结果，图中最左侧的图片为原始 MTCNN 的检测结果，选择原图中较小一块区域放大进行对比，原始 MTCNN 检测到了 8 张人脸，右上角图片为 MobileNetV3-MTCNN 检测结果共计检测出 9 张人脸，右下角为本文 GhostNet-MTCNN 的检测结果共计检测出 14 张人脸，可以看出本文提出的模型可以检测出更多的小尺寸单一角度人脸。图 7 展现了三

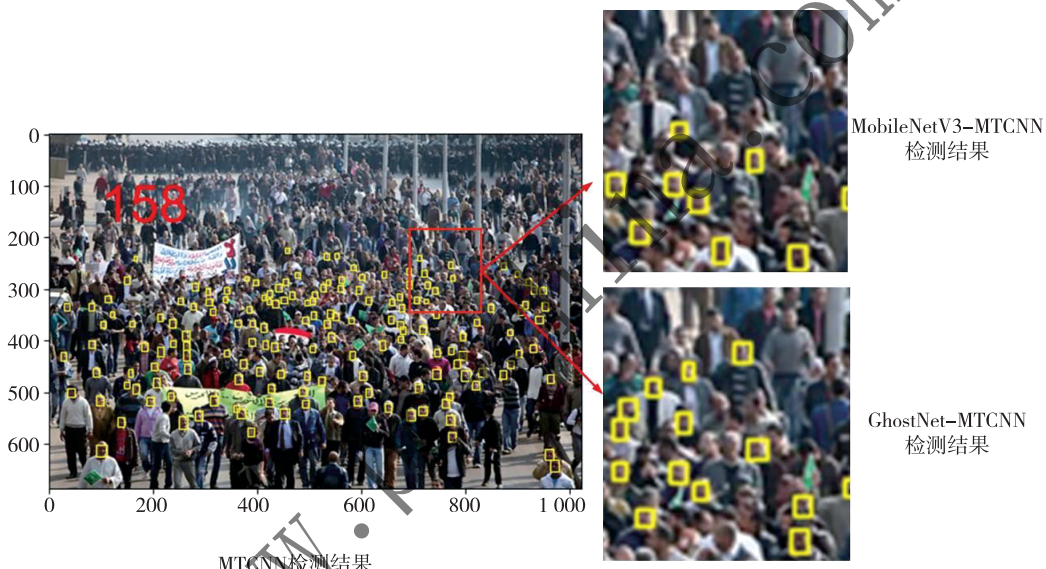


图 6　小尺寸单一角度人脸模型检测结果

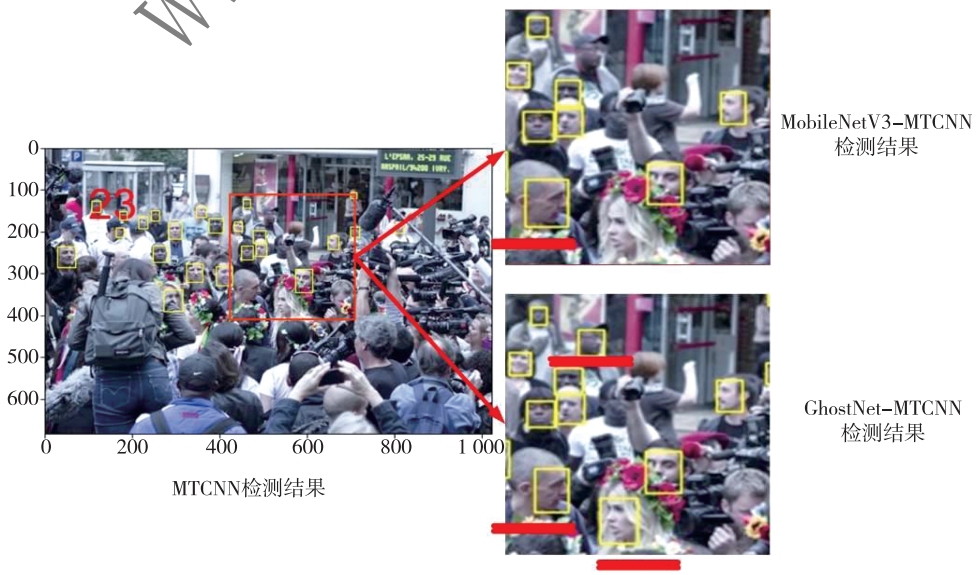


图 7　小尺寸多角度模型检测结果

种模型在小尺寸多角度人脸的检测结果，图中最左侧原始 MTCNN 模型总共检测出 23 张人脸。选择原图中较小一块区域放大进行对比，原始 MTCNN 检测到了 7 张人脸，右上角 MobileNetV3-MTCNN 模型检测出 8 张人脸，右下角 GhostNet-MTCNN 检测出 10 张人脸，其中画横线的人脸为相比原始模型多检测出的人脸。综上所述，在复杂环境下本文提出的模型相比于其他模型可以检测出更多的小尺寸单一角度和小尺寸多角度人脸，体现了本文改进模型有更好的检测结果。

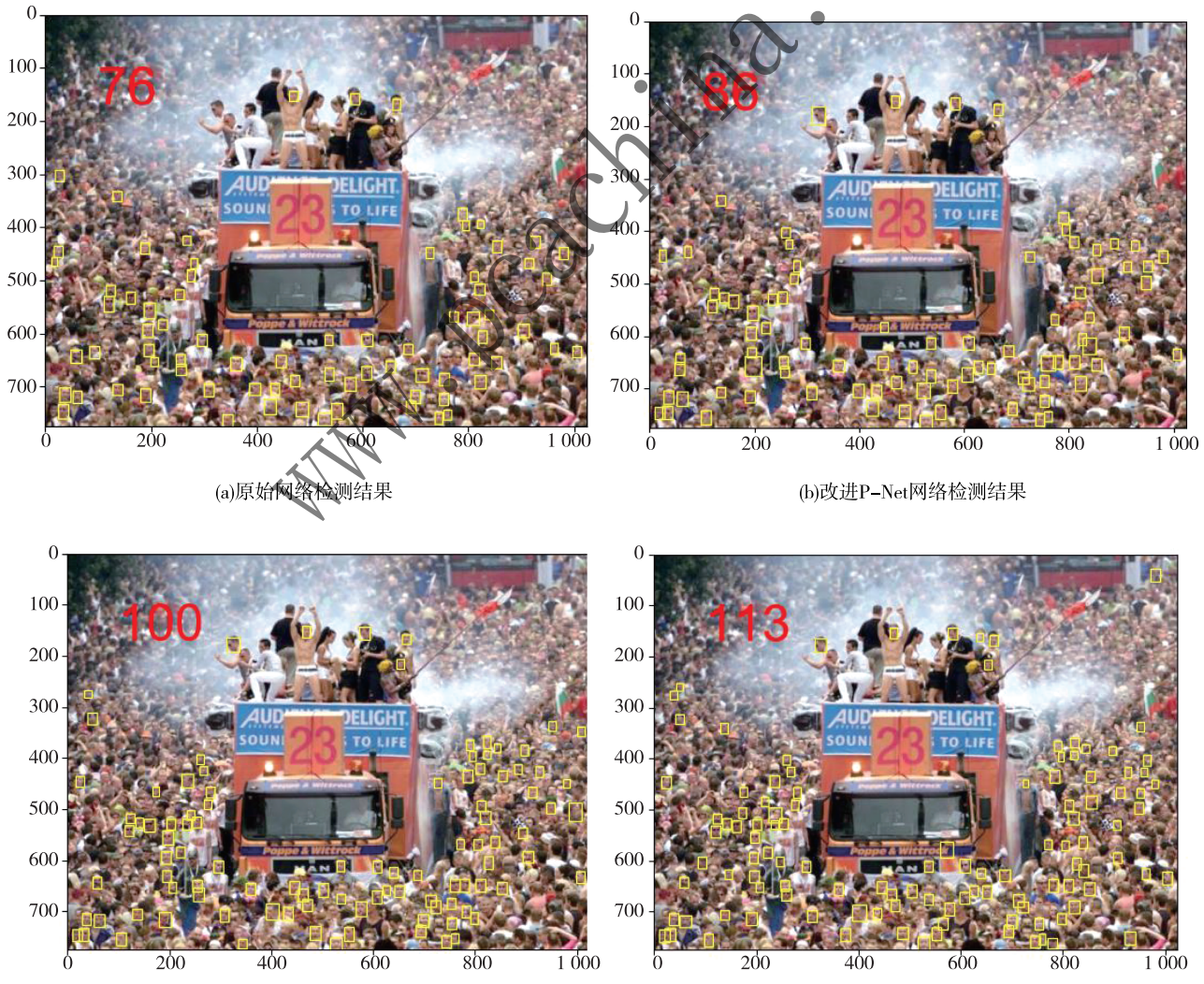
2.4 模型消融

为了验证本文方法对 MTCNN 性能提升的有效性，在训练集和验证集都相同的情况下对验证集为 Hard 的人脸类别设计消融实验，由于 MTCNN 是多级级联结构，对 P-Net、R-Net、O-Net 三个网络依次进行改进。结果如表 5 所示。

根据表 5 结果显示，依次修改 P-Net、R-Net、O-Net 三个网络结构模型，检测精度逐步提升，其检测结果如图 8 所示。其中，图 8（a）所示为原始网络的检测结果，共计检测出人脸数量为 76 个；图 8（b）为修改 P-Net 网络后的检测结果，共计检测出人脸数量为 86 个；图 8（c）为修改 P-Net、R-Net 网络后的检测结果，共计检测出人脸数量为 100 个；图 8（d）为修改 P-Net、R-Net、O-Net 网络后的检测结果，共计检测出人脸数量为 113 个。随着对网络结构的修改模型检测能力逐步上升。综上所述，本文所提出的 GhostNet-MTCNN 模型对复杂环境下小尺寸人脸的检测有更好的检测效果。

表 5 消融实验

基准网络	P-Net	R-Net	O-Net	AP/%	召回率/%	精准率/%
GhostNet-MTCNN				40.9	41.8	96.8
			√	41.8	42.4	98.1
		√	√	46.8	47.4	97.3
	√	√	√	48.7	50.1	93.1



(a)原始网络检测结果　(b)改进P-Net网络检测结果

(c)改进P-Net、R-Net网络检测结果　(d)改进P-Net、R-Net、O-Net网络检测结果

图 8 模型消融实验检测结果

3 结论

为了在复杂背景下对小尺寸多角度人脸进行更好的检测，同时能够更容易地实现在边缘设备上部署模型，本文提出了一种新的神经网络模型（GhostNet-MTCNN），该模型以 MTCNN 为基础，利用 Ghost Bottleneck 模组重新构建特征提取网络。通过与原始 MTCNN 和 MobileNetV3-MTCNN 的对比实验以及模型消融实验证明了所提模型可以在复杂场景下较好地实现小尺寸人脸的检测，同时也能够有效平衡参数量和精度使其成为边缘设备部署的更优选择。经过以上改进可以在人流量较多的场景下取得不错的检测效果。如何将模型部署到边缘设备是接下来可以研究的方向。

参考文献

[1] ROWLEY H A, BALUJA S, KANADE T. Neural network-based face detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (1): 23 – 38.

[2] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57: 137 – 154.

[3] ZHAO Z. Application of improved CNN-based face detection technology in public administration [J]. Journal of Computational Methods in Sciences and Engineering, 2023, 23 (4): 1985 – 1997.

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 – 587.

[5] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137 – 1149.

[6] 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.

[7] REDMOM J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 – 7271.

[8] REDMOM J, FARHADI A. Yolov3: an incremental improvement [J]. arXiv preprint arXiv: 1804. 02767, 2018.

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

[10] ZHANG K P, ZHANG Z P, LI Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499 – 1503.

[11] KU H, DONG W. Face recognition based on MTCNN and convolutional neural network [J]. Frontiers in Signal Processing, 2020, 4 (1): 37 – 42.

[12] 陈政生. 基于深度学习的口罩佩戴检测方法研究 [D]. 武汉：华中师范大学，2021.

[13] HAN K, WANG Y H, TIAN Q, et al. Ghostnet: more features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 – 1589.

（收稿日期：2024 – 01 – 08）

作者简介：

黄杰（1999 – ），男，硕士研究生，主要研究方向：目标检测、深度学习、边缘部署。

刘芬（1979 – ），通信作者，女，硕士，副教授，主要研究方向：微波电路设计、数字信号处理、人工智能。E-mail：liuftute@ 126. com。

（上接第 45 页）

[12] ADAM S, MATTHEW B. One-shot learning with memory-augmented neural networks [J]. IEEE Transactions on Software Engineering, 2022, 49 (4): 1661 – 1682.

[13] MUNKHDALAI T, YU H. Meta networks [C]//International Conference on Machine Learning, 2017: 2554 – 2563.

[14] XU L, CHOY C S, LI Y W. Deep sparse rectifier neural networks for speech denoising [C]//2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), 2016: 26 – 33.

（收稿日期：2023 – 09 – 21）

作者简介：

孙悦（1994 – ），通信作者，女，硕士，工程师，主要研究方向：模式识别。E-mail：1286857718@ qq. com。

彭圆（1972 – ），女，博士，研究员，主要研究方向：模式识别。

贾连徽（1985 – ），男，博士，高级工程师，主要研究方向：目标特性。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部