arXiv:2302.04341v1 [eess.IV] 8 Feb 2023

基于视频记录的新生儿面部与面部标志点检测

Ethan Grooby[1,2,3], IEEE 学生会员、Chiranjibi Sitaula[1], Soodeh Ahani[2,3], Liisa Holsti[2], Atul Malhotra[4], Guy A. Dumont[2,3], IEEE 会士，以及 Faezeh Marzbanrad[1], IEEE 高级会员

摘要——本文探讨了新生儿的人脸及面部关键点自动检测，这是许多基于视频的新生儿健康应用中的重要第一步，例如生命体征估计、疼痛评估、睡眠-觉醒分类和黄疸检测。利用临床环境中的三个公开数据集，分别标注了366张图像（258名受试者）和89张图像（66名受试者），用于训练和测试。本文对两个基于 YOLO 的模型应用了迁移学习，并在每个训练轮次期间，通过随机水平翻转、光度颜色失真、平移和缩放对输入训练图像进行增强。此外，还探索了输入图像的重新定向以及已训练深度学习模型的融合。我们提出的基于 YOLOv7Face 的模型优于现有方法，在人脸检测方面达到84.8%的平均精度均值，在面部标志点检测方面达到0.072的归一化平均误差。总体而言，这将有助于开发完全自动化的新生儿健康评估算法。

临床相关性——准确的人脸及面部关键点检测可提供一种自动化的非接触式方案，以协助基于视频的新生儿健康应用。

I. 引言

视频监测为在临床、家庭和远程环境中监测新生儿提供了一种非接触式方法。通常，首先需要将患者的人脸区域检测为感兴趣区域（ROI），并确定相关的面部关键点。完成此步骤后，还需进行视频帧之间的感兴趣区域跟踪，并对跟踪到的感兴趣区域开展后续分析。

近年来，使用人脸 ROI 进行视频监测以开展婴儿心肺监测，越来越受到关注 [1]–[3]。心肺监测对于新生儿期重大并发症的早期检测、预后预测和持续监测至关重要 [4]。这有助于临床医生及时提供恰当的护理，从而尽可能降低发病率和死亡率 [4]。

*E. Grooby感谢MIME—莫纳什合作伙伴—CSIRO赞助的博士研究支持计划和研究培训计划（RTP）提供的支持。A. Malhotra的研究得到Kathleen Tinsley Trust和脑瘫联盟研究基金的支持。F. Marzbanrad获得Veski Victoria Fellowship 2021的支持。本研究得到莫纳什医学工程研究所（MIME）和莫纳什数据未来研究所（MDFI）的支持。[1]莫纳什大学电气与计算机系统工程系，墨尔本，维多利亚州，澳大利亚。[2]BC儿童医院研究所，温哥华，不列颠哥伦比亚省，加拿大。[3]英属哥伦比亚大学电气与计算机工程系，温哥华，不列颠哥伦比亚省，加拿大。[4]Monash Newborn、莫纳什儿童医院及莫纳什大学儿科学系，墨尔本，维多利亚州，澳大利亚。电子邮箱：ethan.grooby@monash.edu

同样，研究人员也探索了通过视频面部监测及时检测新生儿疼痛的可能性 [5]。这种监测有助于及时采取治疗措施，防止对神经发育和行为发育造成长期影响 [5]。此外，研究人员还探索了通过面部视频监测检测黄疸 [1] 以及对新生儿进行睡眠—觉醒分类 [6] 。

新生儿的人脸及面部关键点检测，尤其是在临床环境中，一直被证明颇具难度 [1], [7]–[9]。与成人相比，新生儿的面部特征不够明显 [7]，姿势更多变，而且奶瓶喂养和安抚奶嘴等常见遮挡也很普遍 [7]–[9]。此外，在临床环境中，背景场景复杂，可能涉及医疗设备和干预措施，且光照条件各异，这些都带来了额外挑战 [1], [9]。此外，许多人脸及面部关键点检测器主要使用成人数据进行训练 [1], [7]。总的来说，这导致新生儿面部感兴趣区域检测效果不佳，因此常常需要手动选择新生儿面部的感兴趣区域 [1],[10], [11]。

本文将在第二节概述若干现有的成人和新生儿人脸/关键点检测器。第三节提出采用迁移学习、图像重新定向和融合技术改进的深度学习模型。随后将在第四节和第五节评估并讨论现有方法和所提方法。

II. 现有研究

A. 成人的人脸及面部关键点检测

当前最先进的人脸及面部关键点检测方法采用了机器学习，尤其是深度学习 [12]–[15]。一种传统的成人 face detection 方法是 OpenCV 的 Haar级联方法 [16]。该方法会提取 Haar特征，并使用 AdaBoost 算法检测人脸 [16]。

对于深度学习方法，Dlib 人脸检测采用方向梯度直方图和基于最大间隔的卷积神经网络（CNN）进行人脸检测 [17]。多任务级联卷积神经网络（MTCNN） [18], 利用人脸检测与对齐以及相关标志点之间的关联来提升性能。MTCNN 分为三个阶段，逐步预测人脸和五个标志点（即右眼、左眼、鼻子、右嘴角和左嘴角），并使用 WIDER FACE数据集进行训练 [19]。

近年来，单次多框检测器（SSD）已广泛用于人脸检测 [13]–[15]。SSD [13] 使用一组

默认边界框，这些边界框在每个特征图位置具有不同的纵横比和尺度。在人脸检测过程中，每个默认框都会输出是否存在人脸的信息，并对边界框进行适当调整，使其与人脸匹配 [13]。

RetinaFace [14] 是一种新型 SSD，统一实现了人脸边界框预测和 2D 面部关键点检测。该模型使用 WIDER FACE数据集 [19],进行训练，并输出五个面部关键点（即右眼和左眼、鼻子、右嘴角和左嘴角）。

BlazeFace [15]是一种轻量级人脸检测器，专为移动 GPU 推理设计，适用于前置和后置摄像头。其特征提取网络的设计受 MobileNetV1/V2 启发，锚框方案则由 SSD [15]修改而来。MediaPipe 提供了 BlazeFace 架构的更新实现，可输出人脸检测结果和 6 个面部关键点（即右眼和左眼、鼻子、嘴中心以及左耳和右耳） [15]。

B. 新生儿人脸及面部关键点检测

与针对成人的现有研究类似，过去针对新生儿的研究大多聚焦于基于深度学习的方法 [1]–[3], [9]。其中，YOLO 框架尤其受到欢迎 [1], [2], [9]。

对于人脸检测，Awais 等人 [6] 提出了一种半自动化的基于强度的方法。RGB 色彩空间被转换为 CIELAB 色彩空间。随后，通过检查每个 CIELAB 通道中的强度值 [6]，确定合适的阈值。最后，将连通区域数量最多的区域确定为面部区域 [6]。

Olmi 等人 [20] 基于聚合通道特征算法，开发了一种用于新生儿重症监护室（NICU）中新生儿的全自动人脸检测方法。该模型使用在意大利佛罗伦萨 AOU Careggi 的神经生理病理学与新生儿临床科室收集的新生儿数据进行训练 [20]。

对于基于深度学习的人脸检测器，Khanam 等人 [2]使用了在 MS COCO 数据集 [22] 上预训练的 YOLOv3 模型 [21]。随后，利用从互联网获取的 473 张新生儿图像进行迁移学习 [22]。

Kyrollos 等人 [3] 构建了一个采用 ResNet 主干网络和 RetinaNet 检测头的深度学习模型 [23], ，该模型在 ImageNet 数据集 [24]上进行了预训练。随后，他们使用入住安大略东部儿童医院 NICU 的患者数据对该模型进行了进一步训练 [1]。

在 [3], Dosso 等人 [1] 的研究基础上，NICUface 应运而生。NICUface 专为在复杂的新生儿重症监护室环境中检测新生儿人脸而设计。开发 NICUface 时，研究人员评估了两种基础模型：YOLOv5Face [25] 和 RetinaFace [14] ，因为这两种模型在初步测试中表现最佳。随后，研究人员使用公开数据集对这些模型进行了训练，包括新生儿心率估计数据库（NBHR） [8] 、iCOPEvid 数据集 [5], ，以及此前从安大略东部儿童医院收集的数据 [1]。总体而言，与现有基础模型和基于 RetinaFace 的 NICUface 相比，基于 YOLOv5Face 的 NICUface 表现最佳。

NICUface 还能够输出 5 个面部关键点（即右眼、左眼、鼻子、右嘴角和左嘴角）；但是，更新后的模型并未专门针对关键点检测重新训练，因此仅进行了定性评估。

同样，Hausmann 等人 [9] 开发了一种基于 YOLOv5 模型的人脸检测器 [26]。该模型使用坦帕总医院采集的图像进行训练，样本的性别构成为男性 47.3%、女性 52.7%，族裔构成为白人 72.7%、非裔美国人 20%、亚裔 7.3%。总体而言，[9] 该模型的表现显著优于 YOLO 模型基线 [26]。

关于面部关键点，Wan 等人 [7] 基于 HRNet 的面部关键点检测方法开发了一种模型 [27]。该模型以人脸检测边界框作为输入，随后输出 68 个面部关键点，涉及眼睛、眉毛、鼻子、嘴巴和人脸轮廓 [7]。他们还使用从互联网获取的婴儿图像对该模型进行了进一步训练 [7] 。

III. 方法

A. 数据采集

1) 数据集：本研究使用了三个公开数据集 [5], [7],和[8] 。

NBHR [8] 包含1,130段视频，记录了257名0至6日龄的新生儿，总时长为9.6小时，并同步采集了光电容积描记图和心率生理信号。这些新生儿招募自中国忻州市人民医院妇产科。生物学性别比例大致相当（男婴48.6%，女婴51.4%）。相对于每名婴儿所在位置，摄像机角度各不相同；部分视频中存在面部遮挡，原因包括奶瓶喂养和睡姿；此外，还采集了多种自然病房照明条件下的视频 [8]。

婴儿标注面部数据集（InfAnFace） [7]，包含从谷歌、YouTube和广告中获取的410张婴儿图像。这些图像标注了68个面部关键点和姿态属性 [7]。

iCOPEvid 数据集 [5]包含 234 段视频（每段视频 20 秒），记录了 49 名新生儿经历一系列伤害性刺激、一段休息时间以及一次急性疼痛刺激的情况。这些新生儿视频拍摄于美国密苏里州斯普林菲尔德仁慈医院的新生儿病房，拍摄对象为出生后 34 至 70 小时的婴儿 [5]。生理性别比例大致均衡（男性 53.1%，女性 46.9%）；16 名新生儿被襁褓包裹，另有 33 名新生儿可自由活动 [5]。族裔分布为：41 名白人、1 名非裔美国人、1 名韩国人、2 名西班牙裔和 4 名跨种族 [5]。

2) 数据提取：对于 NBHR [8] 和 iCOPEvid [5]数据集，提取每段视频的第一帧进行标注。此外，删除了同一受试者在不同视频中相似的新生儿图像，以构建更多样化的数据集。对于 InfAnFace 数据集 [7], ，若图像无法代表医院环境或新生儿年龄阶段，则将其排除。

总体而言，共获得来自324名新生儿的455张图像用于标注。这些图像包括来自257名新生儿的352张图像、来自18名婴儿的18张图像，以及85张图像

分别来自NBHR [8], InfAnFace [7], 和 iCOPEvid [5]数据集中的49名新生儿。

3) 标注：我们对提供的数据集进行了人工标注。人脸边界框覆盖从额头到下巴、两耳之间的区域。对于部分遮挡的情况，如果遮挡区域内仍能看到人脸的部分区域或其部分轮廓，则将其标注为人脸。否则，如果人脸的某个完整区域被遮挡，则不将其纳入边界框，因为难以推断被遮挡人脸的大小。此外，我们还标注了 6 个面部关键点，分别是：右眼、左眼、鼻子、右嘴角、嘴中心和左嘴角。

B. 迁移学习

1) 基础模型：根据表 I 中报告的人脸检测性能，选择 YOLOv5 和 YOLOv7Face [25] 作为基础模型。YOLOv5 的初始权重取自 [9]。YOLOv7Face 的初始权重采用了 GitHub 上的“yolov7”模型：https://github.com/derronqi/yolov7-face，该模型是在 WIDER FACE数据集 [19]上预训练的。

2）数据增强：在每个训练轮次期间，执行了随机水平翻转（概率为50%）、光度颜色失真（即 ±1%、 ±10%、 ±10%的HSV色相、饱和度和值）、平移（即 ±10%）和缩放（即 ±10%）增强。

3）超参数：对于YOLOv5，超参数的设置与 [1]中类似。训练期间采用了初始学习率0.0032、最终学习率0.12、预热动量0.5、最终动量0.843、2个训练轮次的预热周期以及20%的交并比（IoU）阈值，并使用随机梯度下降进行优化。

对于YOLOv7Face，超参数的设置与YOLOv7Face [25]中的示例设置类似。训练期间采用了学习率0.01、预热动量0.8、最终动量0.937、3个训练轮次的预热周期以及20%的IoU阈值，并使用随机梯度下降进行优化。

4) 训练：数据被拆分为约 80% 的训练集和 20% 的测试集，其中训练集包含来自 258 名受试者的 366 张图像，测试评估集包含来自 66 名受试者的 89 张图像。确保同一受试者的所有图像仅出现在训练集或测试集中。

在训练集中，数据进一步拆分以进行五折交叉验证，同时确保每名受试者的图像均位于同一折中。随后，探索了模型的训练轮次以及模型中冻结层的数量。

根据初步交叉验证结果，最终的 YOLOv5 模型在未冻结任何层的情况下训练了 100 轮，最终的 YOLOv7Face 模型在未冻结任何层的情况下训练了 50 轮。两个模型的批量大小均为 16。

C. 方向

现有检测器通常是在图像中人脸保持正立的情况下训练的 [1]。由于新生儿姿势多变，将图像旋转至四种不同的方向，可能会增加模型获得更理想姿势的概率。

因此，使用同一个人脸及面部关键点检测器，将四张图像输入检测器。这四张图像分别是原始图像旋转 0、90、180 和 270 度后的版本。随后，选取置信度检测结果最高的图像，用于人脸及关键点估计。

D. 融合

基于表 I 和表 II 所呈现的结果，考虑了两种融合方法。

首先，由于经过重新定向的 RetinaNet 和所提出的 YOLOv7Face 对面部关键点产生了最低的平均归一化误差（MNE），因此考虑采用基于多数投票的融合模型。具体而言，选择置信度最高的方法（即将四个 RetinaFace 输出中的最大置信度与所提出的 YOLOv7Face 进行比较）来估计面部关键点。采用该融合模型时，平均执行时间约为每张图像 2.09 秒。

为减少计算时间，使用 MediaPipe 前置摄像头模型来估计输入图像相对于 RetinaFace 的最佳方向。在该融合模型中，平均执行时间缩短至每张图像约 1.58 秒。

E. 评估

对于人脸及面部关键点检测，由于每张图像中仅有一名新生儿，因此将置信度阈值设为 5%，以最大限度地增加检测数量，并选择置信度最高的检测结果进行评估。如果未检测到人脸，则将其定义为假阴性。

1) 人脸检测：平均精度（AP）由式（1）定义，并使用特定的交并比阈值。交并比定义为参考边界框与估计边界框之间的重叠面积除以并集面积。交并比阈值用于确定将人脸检测结果分类为真阳性（TP）还是假阳性（FP）。本文考虑两种 AP 指标：交并比阈值为 50% 时的平均精度（AP50），以及交并比阈值从 50% 到 95%、以 5% 为增量时的平均精度均值（mAP）。

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

2）面部标志点检测：使用 MNE 评估所有测试样本 (N) 中参考（ ref ）与估计（ est ）标志点之间的关键点检测情况。归一化距离为参考人脸边界框面积（ref_bbox）（2）、（3）。

$$norm_error(i) = \frac{\sqrt{(ref_x^i - est_x^i)^2 + (ref_y^i - est_y^i)^2}}{\sqrt{ref_bbox_{width}^i \times ref_bbox_{height}^i}} \tag{2}$$

$$MNE = \frac{\sum_{i=1}^{N} norm_error(i)}{N} \tag{3}$$

表I：人脸检测器比较研究

方法	AP50 (%)	mAP (%)	空 (%)	Time (ms)
成人现有方法（第II-A节）				
MediaPipe [15] 正面	86.0	30.5	3.4	**9**
MediaPipe [15] Back	41.2	16.5	80.9	11
OpenCV [16]	10.0	7.0	88.8	115
SSD [13]	75.0	55.0	95.5	123
Dlib [17]	**100.0**	50.0	97.8	126
MTCNN [18]	85.0	51.5	77.5	686
RetinaFace [14]	97.1	55.9	21.3	182
YOLOv7Face [25]	**100**.0	64.6	0.0	1364
新生儿现有方法（第 II-B 节）				
NICUface [1]	83.1	61.5	33.7	1096
YOLOv5 [9]	87.7	46.4	9.0	211
所提方法（第 III-B 节）				
YOLOv5*	98.8	81.8	6.7	211
YOLOv7Face*	**100.0**	**84.7**	**0.0**	1364
方向（第III-C节）				
MediaPipe Front	88.8	30.8	**0.0**	35
RetinaFace	**100.0**	67.5	2.2	728
NICUface	**100.0**	79.0	**0.0**	4383
YOLOv5	87.6	47.5	**0.0**	845
YOLOv7Face	98.9	72.6	**0.0**	5457
YOLOv5*	98.9	80.0	2.2	845
YOLOv7Face*	**100.0**	83.4	**0.0**	5457

*拟议的重新训练模型。AP50 和 mAP 的定义见第III-E.1节，时间指第III-E.3节定义的平均计算时间。Empty 表示未检测到人脸的图像百分比。

表 II：面部关键点检测器的比较研究

方法	All	Eyes	Nose	嘴巴	空 (%)
成人现有方法（第II-A节）					
MediaPipe [15] 正面	0.192	0.204	0.184	0.175	3.4
RetinaFace [14]	0.095	0.102	0.096	0.088	21.3
YOLOv7Face [25]	0.228	0.242	0.227	0.215	**0.0**
新生儿现有方法（第II-B节）					
NICUface [1]	0.320	0.323	0.267	0.343	33.7
所提方法（第III-B节）					
YOLOv7Face*	0.072	0.077	0.062	0.072	**0.0**
方向（第III-C节）					
MediaPipe Front	0.109	0.119	0.102	0.095	**0.0**
RetinaFace	0.075	**0.073**	0.073	0.078	2.2
YOLOv7Face	0.190	0.198	0.180	0.186	**0.0**
NICUface	0.083	0.074	**0.047**	0.109	**0.0**
YOLOv7Face*	0.111	0.122	0.087	0.112	**0.0**
融合方法（第III-D节）					
RetinaFace + YOLOv7Face*	**0.070**	0.075	0.059	0.070	**0.0**
MediaPipe Front + RetinaFace + YOLOv7Face*	0.071	0.078	0.062	**0.069**	**0.0**

*提出的重新训练模型。面部标志点检测采用第III-E.2节中定义的 MNE 进行评估。“All”指所有面部关键点合并后的 MNE，而“Eyes”、“Nose”和“Mouth”指特定标志点的 MNE。Empty 指未检测到面部标志点的图像所占百分比；这些图像未纳入 MNE 计算。

3）执行时间：我们将推理/测试阶段每张图像的平均执行时间作为另一项评估指标。速度更快的方法更适合实时处理和手机实现。使用 Python 在配备 2.3GHz 8 核 Intel i9 中央处理器的 MacBook Pro 上计算每张图像的平均执行时间。

4) 显著性检验：采用配对威尔科克森符号秩检验，以确定交并比和归一化误差是否显著更优，使用的 p 值阈值为 0.05。

IV. 结果

表 I 展示了人脸检测结果。在现有方法中，尤其是 OpenCV、SSD、Dlib 和 MTCNN 难以检测新生儿的人脸，超过四分之三的人脸未能检出。在现有方法中，YOLOv7Face 的表现明显最佳，AP50 为 100%，mAP 为 64.6%。两种所提方法均较现有方法有显著提升，其中重新训练的 YOLOv7Face 模型表现最佳，AP50 为 100%，mAP 为 84.7%。

将图像重新定向至置信度最高的方向后，除所提出的 YOLOv7Face 模型外，所有方法的结果均有所改善。特别是，现有方法 RetinaFace、YOLOv5、YOLOv7Face 和 NICUface 均取得了显著改善。即使经过重新定向，所提出的模型表现仍显著更好；但所提出的 YOLOv5 模型以及重新定向后的 YOLOv5 模型与 NICUface 相比时例外，其 p值分别为 0.3 和 0.1。

表 II 展示了面部标志点检测结果。与人脸检测的情况类似，重新定向显著改善了现有方法的结果，而所提出的 YOLOv7Face 模型表现则有所下降。与所有现有方法及重新定向后的现有方法相比，所提出的 YOLOv7Face 模型在整体 MNE 方面均有显著改善。所提出的 YOLOv7Face 模型在所有标志点、眼睛、鼻子和嘴巴标志点上的 MNE 值分别为 0.072、0.077、0.062 和 0.072。

与仅使用提出的 YOLOv7Face 模型相比，融合模型1（RetinaFace + YOLOv7Face*）带来的变化并不显著。尽管该融合模型平均误差较低，但有25个标志点估计结果变差，改善的有20个。融合模型2（MediaPipe Front + RetinaFace + YOLOv7Face*）则整体上带来了显著但幅度较小的改善；所有、眼睛、鼻子和嘴巴标志点的 MNE 分别为0.071、0.078、0.062和0.069。

V. 讨论

总体而言，所提出的模型在人脸及面部关键点检测方面均有所改进。需要注意的是，NICUface 模型未经修改，直接采用了 [1] 中的模型。然而，该模型是在相同的 iCOPEvid 和 NBHR 数据集上训练的。因此，NICUface 的结果可能存在夸大。

这些模型的一个局限是，YOLOv5 和 YOLOv7Face 的平均执行时间分别为 211 毫秒和 1,364 毫秒。由于执行时间相较于其

他方法较长，这可能会妨碍实时处理。不过，人脸检测通常只需在感兴趣区域跟踪之前的第一步中进行，而这一步的计算开销较低。因此，在采用这些模型的新生儿视频监测系统整体实现中，仍有可能实现实时处理。未来可以研究利用 BlazeFace [15]实现手机实时处理，这或许会有所帮助。

事实证明，重新定向方法对于现有方法而言是一种非常有效的处理流程，既能提高人脸和面部关键点检测的成功率，也能提升其准确性。这凸显了一种应对新生儿可能处于的各种姿势/体位的简单流程。然而，执行该流程所需的时间将是原来的四倍。

对于所提出的模型，YOLOv5 的改进并不显著，而 YOLOv7Face 使用重新定向方法后结果有所下降。一种可能的解释是，这些模型是在未经重新定向、包含各种新生儿姿势和体位的原始输入图像上训练的。相反，重新定向改变了新生儿相对于临床环境的关系。

VI. 结论

本文介绍了两种重新训练的基于 YOLO 的模型，在人脸和面部关键点检测方面表现出显著改进。此外，图像的重新定向显著提升了现有方法的性能，而所提出的方法融合仅带来了轻微改进。未来应进一步研究新生儿人脸分割，以便更准确地定位人脸。

参考文献

[1] Y. S. Dosso, D. Kyrollos, K. J. Greenwood, J. Harrold, and J. R. Green, 《NICUface：复杂新生儿重症监护室场景中的鲁棒新生儿人脸检测》，IEEE Access，第10卷，第62 893–62 909页，2022年。[2] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, K. Gibson, J. Chahl 等，《基于神经网络的新生儿重症监护室婴儿生命体征非接触式自动监测》，Journal of Imaging，第7卷，第8期，第122页，2021年。[3] D. G. Kyrollos, J. B. Tanner, K. Greenwood, J. Harrold, and J. R. Green, 《利用机器视觉进行非接触式新生儿呼吸频率估计》，载于2021年IEEE传感器应用研讨会（SAS）。IEEE，2021年，第1–6页。[4] E. Grooby, C. Sitaula, T. Chang Kwok, D. Sharkey, F. Marzbanrad, and A. Malhotra, 《用于新生儿心肺监测的人工智能驱动可穿戴技术：第1部分——可穿戴技术》，Pediatric Research，第1–13页，2023年。[5] S. Brahnam, L. Nanni, S. McMurtrey, A. Lumini, R. Brattin, M. Slack, and T. Barrier, 《使用iCOPEvid数据集以及从局部描述符高斯模型中提取的描述符集成进行视频新生儿疼痛检测》，Applied Computing and Informatics，2020年。[6] M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink 等，《基于视频面部表情的混合DCNN-SVM模型，用于新生儿睡眠和清醒状态分类》，IEEE 生物医学与健康信息学期刊，第25卷，第5期，第1441–1449页，2021年。[7] M. Wan, S. Zhu, L. Luan, G. Prateek, X. Huang, R. Schwartz-Mette, M. Hayes, E. Zimmerman, and S. Ostadabbas, 《InfAnFace：弥合真实场景下面部关键点估计中的婴儿—成人领域差异》，载于2022年第26届国际模式识别会议（ICPR）。IEEE，2022年，第4486–4492页。

[8] B. Huang, W. Chen, C.-L. Lin, C.-F. Juang, Y. Xing, Y. Wang, and J. Wang, 《基于时空神经网络的非接触式新生儿心率监测数据集与基准》，《人工智能工程应用》，第106卷，104447，2021年，第p.页。[9] J. Hausmann, M. S. Salekin, G. Zamzmi, D. Goldgof, and Y. Sun, 《真实临床环境中的鲁棒新生儿人脸检测》，arXiv预印本 arXiv:2204.00655，2022年。[10] M. Paul, S. Karthik, J. Joseph, M. Sivaprakasam, J. Kumutha, S. Leonhardt, and C. H. Antink, 《利用基于摄像头的成像进行非接触式新生儿脉率感测：一项临床可行性研究》，Physiological Measurement，第41卷，第2期，第024001页，2020年。[11] K. Gibson, A. Al-Naji, J. Fleet, M. Steen, A. Esterman, J. Chahl, J. Huynh, and S. Morris, 《基于计算机视觉系统的早产儿非接触式心率和呼吸频率监测：一项方法比较研究》，Pediatric Research，第86卷，第6期，第738–741页，2019年。[12] A. Kumar, A. Kaur, and M. Kumar, 《人脸检测技术：综述》，《人工智能评论》，第52卷，第948，2019年第2期，第927页。[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, 《SSD：单次多框检测器》，载于《欧洲计算机视觉会议》。Springer，2016年，第21–37页。[14] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, 《RetinaFace：真实场景中的单次多级人脸定位》，载于CVPR，2020年。[15] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, 《BlazeFace：移动GPU上的亚毫秒级神经网络人脸检测》，arXiv预印本 arXiv:1907.05047，2019年。[16] G. Bradski, 《OpenCV库》，Dr. Dobb's Journal of Software Tools，2000年。[17] D. E. King, 《Dlib-ml：机器学习工具包》，Journal of Machine Learning Research，第10卷，第1755–1758页，2009年。[18] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, 《利用多任务级联卷积网络联合进行人脸检测与对齐》，《IEEE信号处理快报》，第23卷，第10期，第1499–1503页，2016年。[19] S. Yang, P. Luo, C.-C. Loy, and X. Tang, 《WIDER FACE：人脸检测基准》，载于IEEE计算机视觉与模式识别会议论文集，2016年，第5525–5533页。[20] B. Olmi, C. Manfredi, L. Frassineti, C. Dani, S. Lori, G. Bertini, S. Gabbanini, and A. Lanatà, 《用于新生儿重症监护病房新生儿人脸检测的聚合通道特征》，载于2022年第44届IEEE医学与生物学工程学会年度国际会议（EMBC）。IEEE，2022年，第455–458页。[21] J. Redmon and A. Farhadi, 《YOLOv3：渐进式改进》，arXiv预印本 arXiv:1804.02767，2018年。[22] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, 《Microsoft COCO：上下文中的常见物体》，载于《计算机视觉——ECCV 2014：第13届欧洲会议，瑞士苏黎世，2014年9月6日至12日，论文集，第V部分13》。Springer，2014年，第740–755页。[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, 《用于密集目标检测的焦点损失》，载于IEEE国际计算机视觉会议论文集，2017年，第2980–2988页。[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 《ImageNet：大型层次化图像数据库》，载于2009年IEEE计算机视觉与模式识别会议。IEEE，2009年，第248–255页。[25] D. Qi, W. Tan, Q. Yao, and J. Liu, 《YOLO5Face：为何重新发明人脸检测器》，arXiv预印本 arXiv:2105.12931，2021年。[26] G. Jocher, A. Stoken, J. Borovec, NanoCode012, ChristopherSTAN, L. Changyu, Laughing, tkianai, A. Hogan, lorenzomammana, yxNONG, AlexWang1900, L. Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, F. Ingham, Frederik, Guilhen, Hatovix, J. Poznanski, J. Fang, L. Yu, changyu98, M. Wang, N. Gupta, O. Akhtar, PetrDvoracek, and P. Rai, 《ultralytics/yolov5：v3.1——错误修复与性能改进》，2020年10月。 [在线]。可访问：https://doi.org/10 5281 4154370 。/zenodo。J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang 等，《用于视觉识别的深度高分辨率表征学习》，《IEEE模式分析与机器智能汇刊》，第43卷，第10期，第3349–3364页，2020年。[27]