

基于双模型的婴儿异常睡姿识别方法

黄小杰, 黄明, 臧福星, 章顺, 巢梓涵

江苏理工学院机械工程学院, 江苏 常州

收稿日期: 2024年1月9日; 录用日期: 2024年3月6日; 发布日期: 2024年3月13日

摘要

在父母对婴儿的监护中, 婴儿的异常睡姿会引起婴儿睡眠猝死, 尤其在一两岁经常发生, 研究表明, 婴儿的俯卧极容易引发这种现象, 及时、准确地对婴儿进行目标检测和睡姿识别, 对其健康安全有着至关重要的作用。为了综合满足婴儿监护的准确度、误报率和检测速度的要求, 提出了一种基于双模型结构婴儿异常睡姿识别方法。采用改进后YOLOv5与OpenPose相结合的方法, 结合YOLOv5的目标检测实时性和OpenPose的人体姿态估计的准确性, 通过引入感知机与门作为二分类器, 将两个模型的输出结果融合, 对婴儿异常睡姿进行准确的二元分类判定, 实验表明, 本文模型在婴儿睡姿检测中准确率达到94.7%, 误报率为2.0%, 能够有效处理婴儿异常睡姿检测任务, 在提高准确率的同时, 还具有较低的误报率, 可提高系统的可靠性和监护人的舒畅度。

关键词

婴儿监护, 睡姿识别, 双模型, 误报率

Recognition Method for Abnormal Sleeping Posture of Infants Based on Dual Models

Xiaojie Huang, Ming Huang, Fuxing Zang, Shun Zhang, Zihan Chao

School of Mechanical Engineering, Jiangsu University of Technology, Changzhou Jiangsu

Received: Jan. 9th, 2024; accepted: Mar. 6th, 2024; published: Mar. 13th, 2024

Abstract

In the parents' monitoring of the baby, the baby's abnormal sleeping position can cause sudden infant sleep death, especially in one or two years old, research shows that the baby's prone lying is very easy to cause this phenomenon, timely and accurate detection of the baby's target and sleeping position identification, has a vital role in its health and safety. In order to meet the requirements of accuracy, false alarm rate and detection speed of infant monitoring comprehensively, a

method of infant abnormal sleeping position recognition based on dual model structure was proposed. By using the improved method of combining YOLOv5 and OpenPose, combining the real-time target detection of YOLOv5 and the accuracy of human pose estimation of OpenPose, the perceptron and door were introduced as binary classifier, and the output results of the two models were fused to make accurate binary classification judgment on the abnormal sleeping position of infants. The experiment showed that, the accuracy rate of this model in infant sleeping position detection reached 94.7%, and the false alarm rate was 2.0%, which can effectively handle the task of abnormal sleeping position detection of infants. While improving accuracy, the model also has a low false alarm rate, which improves the reliability of the system and the comfort of guardians.

Keywords

Infant Monitoring, Sleep Pose Recognition, Dual Model, False Alarm Rate

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自 2017 年来,我国新生儿数量从 1700 万减少到 900 多万,对于婴儿的健康和监护成为重中之重,其中婴儿猝死综合征(Sudden Infant Death Syndrome),通常在婴儿睡眠时突然发生,这对婴儿睡眠环境尤其是睡眠姿势有着一定的要求,俯卧位睡眠极易增加猝死率,侧卧睡姿也会有同样的危险,仰卧睡眠能够减少 SIDS [1]的风险,这需要对姿势识别有着准确的判断,由于机器学习和深度学习的流行以及相关研究的进展,基于 RGB 图像的姿势识别在实用性和通用性方面变得更加强大,任志斌[2]提出一种基于模糊粗糙算法框架处理图像身体的遮挡和部分缺失,引入模糊粗糙集,采用属性约简的方法提高睡姿分类的速度,具有高效的性能;刘今越[3]等人提出了一种支持向量机 SVM 与遗传算法 GA 相结合解决睡姿中人体束缚和识别精度不高等问题;赵海文[4]等人利用柔性压力传感器采集坐卧姿势数据,使用 YOLOv5 对数据进行训练,实现了精确的姿势识别;杨明健[5]利用 OpenPose 姿态估计对睡眠图像进行人体关键点定位来进行睡姿判断,解决了背景复杂的问题并具有较高的识别率;周志文[6]基于人工神经网络(ANN)设计对应网络结构,将网络隐藏层节点数测试优化后设置网络参数训练,提高了训练与识别的准确率与效率,石用伍[7]等人利用心冲击图信号空间和时域的特点在深度学习睡姿检测模型引入注意力机制,完成对睡姿有效、无扰和准确的监测。

随着深度学习和卷积神经网络(CNN)的兴起,对于姿态识别的检测效果有了显著的提高,但是仍然存在一些问题。现有的模型对于监测方面,往往在准确度和实时性上不能保持平衡,通过加深网络层数提高准确性,这样导致网络复杂不能满足监测速度;通过网络剪枝和量化去加快推理速度反而检测精度大幅下降;对于监测异常睡姿任务而言,还需要较低的误报率。

针对以上问题,本文提出一种基于 YOLOv5 与 OpenPose [8]的双模型婴儿睡姿识别方法,该方法结合了两者的优点能兼顾快速的目标定位和精确的识别能力,提供更全面的场景理解和信息,提高系统的灵活性和适应性,通过感知机与门将两者结合,形成并联系系统降低误报率。本文方法在具备较高准确性和实时性的同时能降低对婴儿监护的误报率。

2. YOLOv5 与 OpenPose 混合模型

为了减少监护人不必要的检查和干预,在准确度一定的同时,还具备相当的低误报率。本研究提出

了一种创新的方法，将两个模型结合，输入图像在分别经过改进后的 YOLOv5 和 OpenPose 得到婴儿姿态正常或异常判断，通过感知机与门，结合两者的输出，通过控制与门阈值(threshold)和权重(weight)，当两者结果判断都为真则为真，其余为假，从而完成婴儿睡姿检测的任务，总的流程图如图 1 所示。

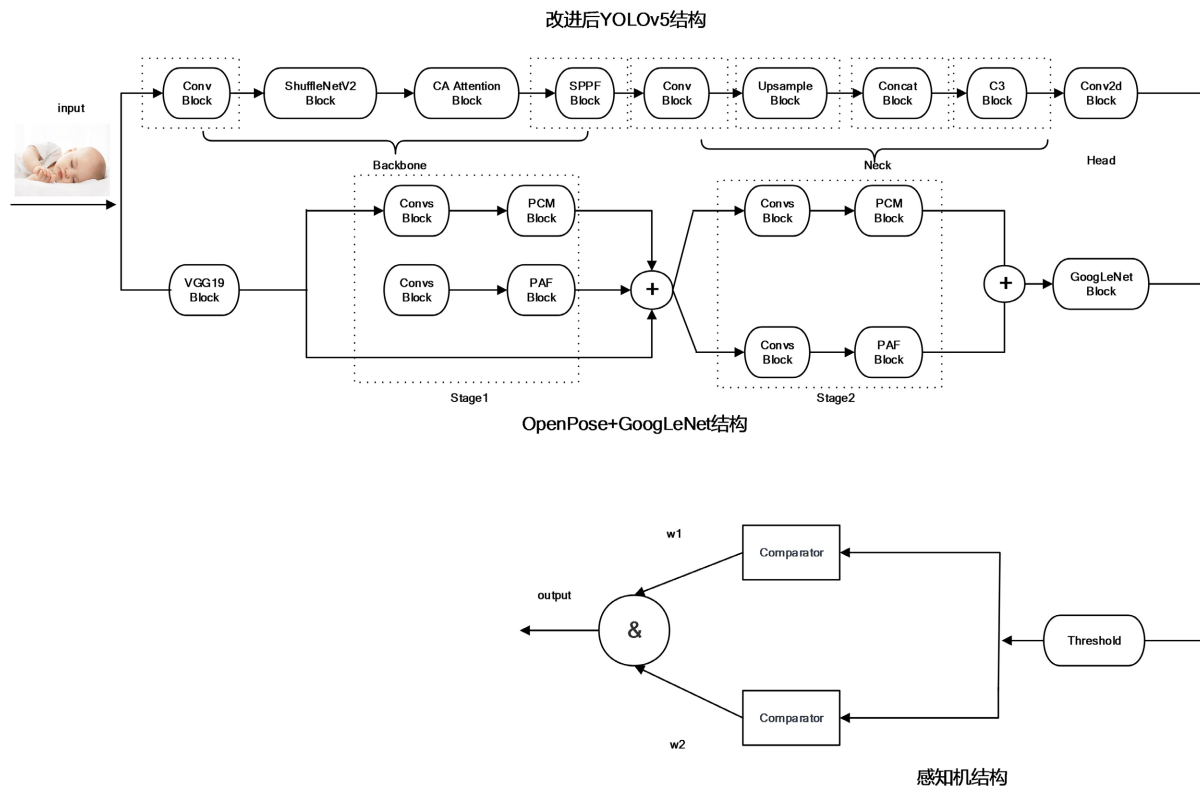


Figure 1. General working flow chart
图 1. 总工作流程图

2.1. YOLOv5 模型

YOLOv5 是一种快速而有效的目标检测模型，其主要由 Input、Backbone、Neck 和 Head 四部分组成，图像通过输入层进行处理并发送到主干进行特征提取，Backbone 获得不同大小的特征图，然后由特征融合网络(Neck)融合这些特征，生成三个特征图送入预测头(Head)，使用预设的先验框对特征图中的每个像素执行置信度计算和边界框回归，以获得包括对象类别、类别置信度、框坐标、宽度和高度信息的多维数组(BBoxes)，通过设置相应的阈值(confthreshold, objthreshold)来过滤阵列中的无用信息，并执行非最大值抑制(NMS)过程，可以输出最终的检测信息。

2.1.1. 主干网络

采用 ShuffleNetV2 [9]代替原本的 C3 网络，通过引入通道重排操作，将卷积操作的计算量大大降低，具有较高的准确性，采用分组卷积将模型参数量显著减少，降低模型复杂度，具有较高的部署和推理效率，其高速度的推理适用于对实时性要求较高的任务，如实时目标检测的视频分析，如图 2 所示。

在给定一个 1×1 的卷积层，其输入维度是 c_1 ，输出维度是 c_2 ，特征图(feature map)尺寸为 $h \times w$ ，其 MAC 值为

$$MAC = hwc_1c_2 + c_1c_2 \tag{1}$$

分组卷积(Group Convolution)可以减少模型的 FLOPs, 但是当 FLOPs 值固定时, 分组 g 越多, MAC 也会越大, 分组卷积的 FLOPs 计算公式如下:

$$B = hwc_1c_2/g \quad (2)$$

当 FLOPs 值固定, 输入的特征图尺寸固定 $c_1 \times h \times w$ 时, MAC 与分组数量 g 成正比,

$$MAC = hwc_1 + \frac{Bg}{c_1} + \frac{B}{hw} \quad (3)$$

在 ShuffleNetV2 使用一种新操作 channel split, 将原始通道拆成两部分, 模型碎片化降低并行度, 其中一个分支保持不变, 另一个分支包含三个卷积操作, 这些卷积操作使其输入输出的通道数保持一致, 减少 MAC, 在 1×1 卷积不再分组, 减少分组对 MAC 的影响, 同时 channel split 操作本身就已经将通道分成两组, 经过卷积之后, ShuffleNetV2 会将两个分支拼接在一起(concat), 而不是采用 Add 操作, 从而减少元素级操作, 最后 ShuffleNetV2 再用 channel shuffle 混合通道。

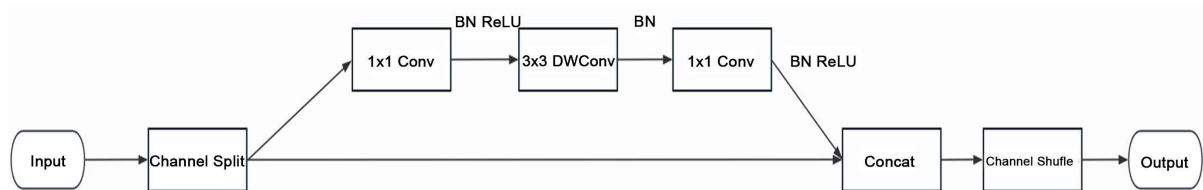


Figure 2. ShuffleNetV2 network structure

图 2. ShuffleNetV2 网络结构

2.1.2. 注意力机制

为了使网络更加关注重要的特征, 比如婴儿的头部、身躯、手脚等等, 在复杂的环境中也能使网络抑制不重要的特征, 使网络对噪声和干扰更加鲁棒, 加快模型的收敛速度。

CA [10]注意力机制(如图 3 所示)对输入的特征图进行平均池化, 得到每个通道的全局平均值, 获得宽度和高度两个方向的特征图:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < H} x_c(h, i) \quad (4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (5)$$

将上方的全局平均池化结果输入到一个全连接层中, 进行特征的映射和变换, 然后将特征图批量归一化通过 Sigmoid 激活函数得到形如 $1 \times (W + H) \times C/r$ 的特征图 f :

$$f = \delta(F_1 | z^h, z^w) \quad (6)$$

接着将特征图 f 进行 1×1 的卷积得到特征图 F_h 和 F_w , 再经 Sigmoid 激活函数得到特征图在高度和宽度上注意力权重:

$$g^h = \sigma(F_h(f^h)) \quad (7)$$

$$g^w = \sigma(F_w(f^w)) \quad (8)$$

最后 CA 注意力机制的输出公式为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (9)$$

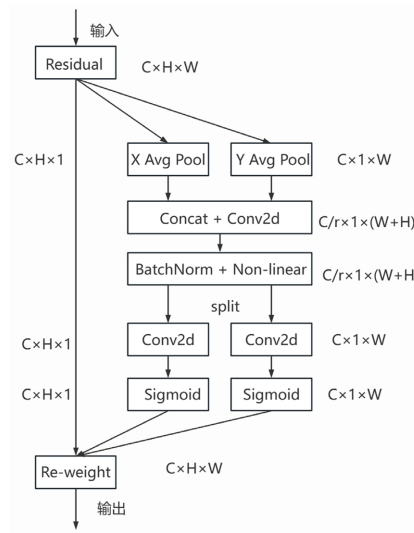


Figure 3. CA attention mechanism
图 3. CA 注意力机制

2.2. 基于 OpenPose 的骨架分类

OpenPose 通过输入图像或视频，经主干网络 VGG-19 前 10 层初始化并进行微调，生成一组特征映射 F 作为下一阶段的输入，在这一阶段会产生一组关节置信度映射 PCM (Part Confidence Maps) 和部位亲和场 PAF (Part Affinity Fields)，其中 PCM 表示像素在关节点的高斯响应，即 $S^1 = \rho^1(F)$ ，PAF 表示像素点在骨架中的走向，即 $L^1 = \phi^1(F)$ ，其中 ρ^1 和 ϕ^1 为第一阶段推理的 CNNs，在随后每一个阶段，都会将前一阶段的预测与特征 F 相结合，产生精细的预测。在前后传播的过程中，OpenPose 会输出一个特征图，每一个像素对应一个可能的关节点，通过对特征图进行解码和阈值处理，得出关节点的粗略位置，根据关节点之间的空间关系，OpenPose 输出表示人体姿态的关节点坐标，每一个关节点都由一个二维坐标 (x, y) 表示，

通过关节点之间的连接，可以形成人体的姿态骨架，本文模型中用 24 个关节点，如图 4 所示，0-鼻子，1-脖子，2-右肩，3-右肘，4-右手腕，5-左肩，6-左肘，7-左手腕，8-胯中心，9-右臀，10-右膝，11-右踝，12-左胯，13-左膝，14-左踝，15-右眼，16-左眼，17-右耳，18-左耳，19-左脚内，20-左脚外，21-左脚跟，22-右脚内，23-右脚外，24-右脚跟。

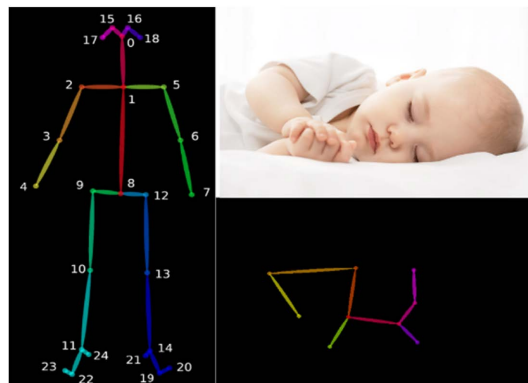


Figure 4. Skeleton extraction
图 4. 骨架提取

2.3. 感知机与门

感知机(perceptron)可接受多个输入信号, 本文将两个模型输出的结果作为输入, 输出一个信号, 如图 5 所示, x_1 、 x_2 是输入信号, y 是输出信号, w_1 、 w_2 是权重, 输入信号被送往神经元分别乘以固定的权重, 计算传送过来的信号总和, 只有当总和超过阈值 θ (threshold)才会输出 1, 确定一分类超平面, 其感知机数学过程亦可定义为:

$$y = \begin{cases} 0 & \text{if } (w_1x_1 + w_2x_2 + \theta) \leq 0 \\ 1 & \text{if } (w_1x_1 + w_2x_2 + \theta) > 0 \end{cases} \quad (10)$$

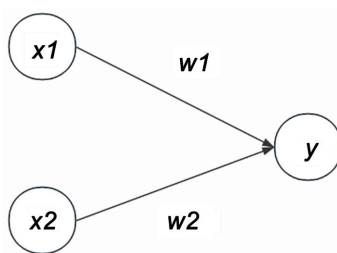


Figure 5. Perceptron

图 5. 感知机

与门表现两个输入均为 1 时输出为 1, 其他情况输出为 0, 根据式(10)训练一个感知机求解超平面方程来表示与门, 根据输出 y 值将结果分为“0”和“1”两类, 来代表婴儿睡姿正常和异常的状态, 如图 6 所示。

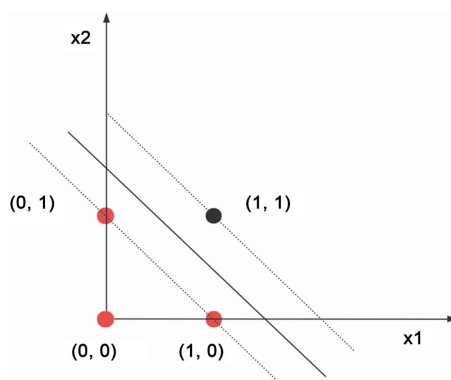


Figure 6. Hyperplane with gate

图 6. 与门超平面

3. 实验

本文使用的计算机硬件设备有: Intel®Core™i5—8300CPU, NVIDIA GeForce GTX 1060 GPU, 8GB 内存。在 64 位 Windows 11 的实验环境下进行实验。

3.1. 数据集和评价标准

由于没有公开的婴儿睡姿数据集, 本文使用网络爬虫技术爬取婴儿睡姿图片, 经过筛选、过滤与整理, 通过数据增强将自制数据集扩充到数千张, 以提高模型的泛化性, 将收集的图片分成三类, 分别为

仰卧、侧卧和俯卧，如图 7 所示，共有 3075 多张，以 7:3 的比例划分为训练集和测试集，其中趴卧侧卧共有 2100 张，俯卧共有 975 张，将其 2153 张作为训练集，922 张作为测试集。



Figure 7. Data set preparation
图 7. 数据集准备

本文使用准确率 P (Precision)、帧率 FPS(frame per second)以及误报率 FPR (False Positive Rate)指标进行性能评估，表达式为 $P = TP/(TP + FP)$ ， $FPR = FP/(FP + TN)$ ，其中 TP (True Positive)模型正确检测异常睡姿，FP (False Positive)模型错误将正常睡姿误分类为异常睡姿，TN (True Negative)模型正确检测正常睡姿。

3.2. 基于 YOLOv5 的婴儿睡姿检测

将 YOLOv5 与其他主流目标检测算法如 YOLOv4、SSD 和 Faster RCNN 进行了对比，通过表 1 的综合性能评估可发现 YOLOv5 在多个关键指标上表现出色。相较于 Faster RCNN 和 SSD，虽然它们在准确率方面较优，但无法满足实时监测的需求，而 YOLOv5 在保持高准确率的同时实现了更为出色的检测速率。虽然 YOLOv4 在准确率、召回率、平均精度和实时性方面表现稳定，但相较于 YOLOv5 仍存在一些性能上的差距。因此，YOLOv5 在综合性能和实时性方面都取得了显著的优势，其设计理念和算法架构在目标检测领域展现了独特的特色。

Table 1. Comparison between YOLOv5 and different detection algorithms
表 1. YOLOv5 与不同检测算法对比

Model	Precision	Recall	mAP@0.5	FPS
Faster-RCNN	87.7%	90.8%	92.4%	28
SSD	88.4%	91.2%	91.7%	39
YOLOv4	84.1%	88%	88.2%	46
YOLOv5	87.1%	90.5%	90.49%	62

3.3. 基于 OpenPose 的婴儿检测

在基于 OpenPose 的婴儿检测中，将婴儿提取骨骼特征输入 GoogLeNet [11]网络模型进行训练，经过

300 次迭代训练，模型的准确性达到 92.8%，与其他 AlexNet [12]、VggNet [13]与 ResNet [14]相比，具有高效的准确性，且参数量不大，计算复杂度较低，模型总体性能比其他网络更好，如下表 2。

Table 2. Comparison of four different models based on OpenPose

表 2. 基于 OpenPose 四个不同模型对比

Model	Parameter	FLOPs	Precision
OpenPose + GoogLeNet	5 M	1.5 G	92.8%
OpenPose + AlexNet	6 M	7.3 G	90.2%
OpenPose + VggNet	13 M	15.7 G	89.7%
OpenPose + ResNet	11 M	10.8 G	88.3%

3.4. 消融实验

为了进一步提高 YOLOv5 的准确性和检测速率，在自建的数据集上做了消融实验，如下表 3，在 YOLOv5 中将主干网络 C3 替换成轻量型 ShuffleNetV2 网络，或者在网络添加 CA 注意力机制模块。实验中设置相同的学习率、批量大小和迭代次数等参数，当更换轻量型 ShuffleNetV2 网络后可以看出模型在监测速度上提升了 16 个点，而添加注意力机制使得模型提高了准确率 3.1 个百分点，同时更改这两方面，改进后的 YOLOv5 相比原模型提高准确率 3.5% 和帧率 5 fps。

Table 3. Ablation results of YOLOv5

表 3. YOLOv5 消融实验结果

ShuffleNetV2	CA 注意力机制	Precision	FPS
×	×	87.1%	62
×	√	90.2%	59
√	×	88.4%	78
√	√	90.6%	67

在同一条件下，将本文的算法模型分别和原型 YOLOv5、改进后的 YOLOv5 和 OpenPose 算法进行对比，如下表 4 所示，本文算法通过感知机与门结合改进 YOLOv5 和 OpenPose，相比较其他三个模型，提高一定的准确性的同时降低监测的误报率，也能满足实时监测的要求，图 8 为部分婴儿睡姿检测结果和混淆矩阵的可视化，可见本文算法在婴儿睡姿识别方面有着出色的表现，由混淆矩阵可得在误报率上仅为 2%，意味着在监测过程中，仅有极少数误报事件，几乎不会发生虚警。这一结果反映了本文监测系统的综合性能得到显著提高。

Table 4. Results of different mixed models on infant sleeping position recognition

表 4. 不同混合模型对婴儿睡姿识别结果

方法	Precision	FPR	FPS
YOLOv5	87.1%	7.2%	62
YOLOv5 + ShuffleNetV2 + CA	90.6%	5.4%	67
OpenPose + GoogLeNet	92.8%	9.9%	50
本文	94.7%	2.0%	53

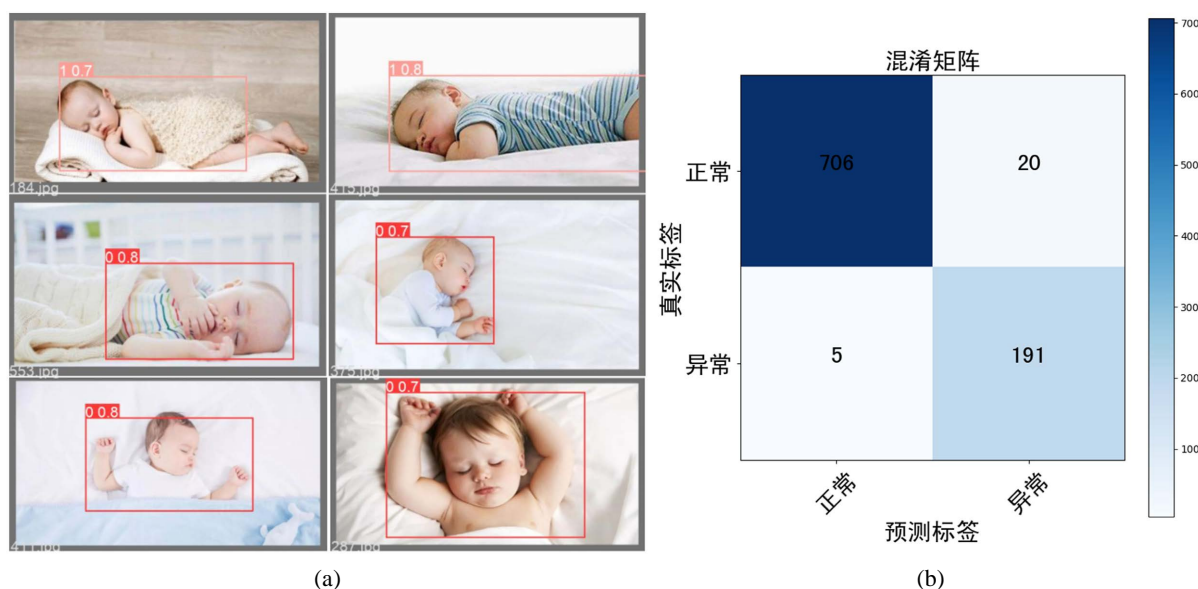


Figure 8. Visualized results
图 8. 可视化结果

4. 结束语

本文基于 YOLOv5 和 OpenPose 双模型结构成功构建了婴儿睡姿识别模型，并在自制婴儿睡姿数据集实验中取得了显著的成果。采用了轻量级 ShuffleNetV2 作为 YOLOv5 主干网络，同时引入了 CA 注意力模块，以及对 OpenPose 人体姿态骨骼图的 GoogleNet 网络训练，为了进一步提高系统的准确性和减少误报率，引入了感知机与门，将两个模型的信息有效融合，为系统决策提供了更加综合和可靠的依据。

实验结果表明，本文模型在婴儿睡姿检测中达到了 94.7% 的准确率，仅有 2.0% 的误报率。而且，监测速度足以满足实时检测的需求，婴儿睡姿监测系统的性能和可靠性也得到了有效的提升。未来，我们将进一步优化和拓展这一模型，以更好地满足婴儿护理和监测的需求，为社会健康事业做出更多贡献。

参考文献

- [1] 邢唯杰, 周菲菲, 王靖, 等. 预防婴儿猝死综合征的安全睡眠环境证据总结[J]. 中国护理管理, 2020, 20(12): 1831-1837. <https://doi.org/10.3969/j.issn.1672-1756.2020.12.016>
- [2] 任志斌, 李洋, 郭士杰, 郭志红, 刘秀丽. 基于模糊粗糙集的睡姿压力图像识别[J]. 计算机工程与应用, 2018, 54(3): 172-177.
- [3] 刘今越, 周志文, 贾晓辉, 郭士杰. 基于柔性压力传感器的睡姿识别方法研究[J]. 传感器与微系统, 2021, 40(9): 10-13. [https://doi.org/10.13873/J.1000-9787\(2021\)09-0010-04](https://doi.org/10.13873/J.1000-9787(2021)09-0010-04)
- [4] 赵海文, 王佳阔, 齐德瑄, 郭士杰. 基于压力感知的坐卧姿势识别方法研究[J]. 传感器与微系统, 2023, 42(1): 30-33. [https://doi.org/10.13873/J.1000-9787\(2023\)01-0030-04](https://doi.org/10.13873/J.1000-9787(2023)01-0030-04)
- [5] 杨明健, 黎镜林, 郭锐坤, 唐小煜. 基于 OpenPose 的人体睡姿识别实现与研究[J]. 物理实验, 2019, 39(8): 45-49. <https://doi.org/10.19655/j.cnki.1005-4642.2019.08.008>
- [6] 周志文. 压力传感器性能优化及睡姿识别方法的研究[D]: [硕士学位论文]. 天津: 河北工业大学, 2020. <https://doi.org/10.27105/d.cnki.ghbgu.2020.000252>
- [7] 石用伍, 李小勇, 石用德, 等. 基于注意力机制的空时融合深度学习睡姿监测算法研究[J]. 中国医疗设备, 2022, 37(7): 39-44.
- [8] Yang, N. and Zhao, J. (2022) Dangerous Driving Behavior Recognition Based on Improved YoloV5 and Openpose. *IAENG International Journal of Computer Science*, 49, No. 4.
- [9] Cengil, E., Çınar, A. and Yıldırım, M. (2022) An Efficient and Fast Lightweight-Model with ShuffleNetv2 Based on

-
- YOLOv5 for Detection of Hardhat-Wearing. *Review of Computer Engineering Studies*, **9**, 116-123. <https://doi.org/10.18280/rces.090305>
- [10] Guangbo, L., Guolong, S. and Jun, J. (2023) YOLOv5-KCB: A New Method for Individual Pig Detection Using Optimized K-Means, CA Attention Mechanism and a Bi-Directional Feature Pyramid Network. *Sensors*, **23**, 5242. <https://doi.org/10.3390/s23115242>
- [11] Harish Babu, G. and Venkatram, N. (2022) An Efficient Image Dahazing Using Googlenet Based Convolution Neural Networks. *Multimedia Tools and Applications*, **81**, 43897-43917. <https://doi.org/10.1007/s11042-022-13222-2>
- [12] Lu, J. and Wan, X.T. (2022) Image Recognition Algorithm Based on Improved AlexNet and Shared Parameter Transfer Learning. *Academic Journal of Computing & Information Science*, **5**, No. 12. <https://doi.org/10.25236/AJCIS.2022.051202>
- [13] Prasetyo, E., Suciati, N. and Fatichah, C. (2022) Multi-Level Residual Network VGGNet for Fish Species Classification. *Journal of King Saud University-Computer and Information Sciences*, **34**, 5286-5295. <https://doi.org/10.1016/j.jksuci.2021.05.015>
- [14] Qian, H., Zhou, X. and Zheng, M. (2020) Abnormal Behavior Detection and Recognition Method Based on Improved ResNet Model. *Computers, Materials & Continua*, **65**, 2153-2167. <https://doi.org/10.32604/cmc.2020.011843>