

# Deep Learning Based Infant and Child Monitoring System

Peng Wang

*College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China*

Jiale Jia

*College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China*

*E-mail: autowangpeng@tust.edu.cn, 15127223153@163.com*

## Abstract

This paper focuses on a baby monitoring system based on computer vision and multi-branch convolutional neural network, firstly, the collected photos are processed, and then the algorithm is implemented using openCV library to train to get the baby's facial target detection model, and secondly, based on the opencv algorithm and yolov8 algorithm technology to achieve the tracking and analysis of baby's behavioral trajectory and facial detection. Finally, we realized the functions of target tracking, night lightening, image segmentation and baby face detection, and the detection achieved good results.

*Keywords:* Deep learning, Neural network, Attribute recognition, Yolov8

## 1. Introduction

With the fast pace of modern life and changes in family structure, more and more parents need to take care of their babies as part of their daily routine. However, it is difficult to remain vigilant at all times while taking care of their babies, especially when they need to fulfill other tasks, which may lead to injuries or mood swings that can have a negative impact on the family. Therefore, the development of a baby safety monitoring system can help parents identify potential problems and take necessary measures in a timely manner, contributing to the safety and emotional stability of their babies.

This project aims to develop a baby monitoring system based on computer vision and multi-branch convolutional neural networks. The system can reduce the blind spot of people's monitoring of infants and young children, and detect the safety factor of infants and young children to help parents monitor the safety and emotional state of their babies.

The rest of this paper is organized as follows: The second section introduces the research on infant facial recognition both domestically and internationally; the third section discusses the technical solutions for detecting infant facial expressions; the fourth section presents the experimental results; and the fifth section summarizes the main content of this paper.

## 2. Domestic and International Research

Infant facial recognition detection system is a technology that combines computer vision, artificial intelligence and biometrics for recognizing, analyzing and verifying facial features of infants and children. Such technologies have applications in a variety of fields, including child health,

behavioral monitoring, and safety and security.

In China, with the rapid development of artificial intelligence technology, infant facial recognition technology has gradually gained attention and application. The research in China mainly focuses on the following areas: infant facial recognition accuracy, biometric identification applications, intelligent security and child protection. In foreign countries, the application of infant facial recognition technology is also growing, mainly focusing on the following areas: facial expression recognition and emotion calculation, medical and health monitoring, and intelligent parenting devices.

## 3. Technical Program

### 3.1. Establishment of a database

Crawler technique is used to collect enough baby expressions from multiple data sources to build a baby expression database [1]. Then used the sprite annotation assistant to label the inductive regions of the infant pictures in the database, which are mainly classified into three categories: POSITIVE, PEACE, and PASSIVE. Furthermore, I used rectangular boxes to crop out all the infant facial regions in the database, as a way to reduce the interference of other environmental factors on facial recognition. It is also necessary to categorize all the data images, due to the simplicity of the infant facial expressions, after all the images are labeled, they are exported to XML format. Then it is transferred to the format used by YOLO, through pre-processing, including data enhancement and data division. Data enhancement can be done by random cropping, rotating, scaling, flipping, etc. to expand the data volume and enhance the generalization ability of the model. Data partitioning can divide the dataset into training, validation and test sets for model training and evaluation.

### 3.2. Training models

Model training was performed using YOLOV8. Using the YOLO algorithm library, after collecting sufficient quantity and quality of data, it is fed into the model to be trained using a number of techniques such as deep neural networks, CNN (convolutional neural networks) [2], data augmentation, optimizers, and GPU acceleration, to build a face recognition model with good generalization performance. After the model is trained, it can be applied to real-time video streams to determine whether infants are in a positive, negative or calm emotional state by analyzing their facial expressions.

### 3.3. Model evaluation

In target detection, the important metrics for evaluating the performance of a model are Accuracy, Precision, Recall and mean average precision (mAP).

First, calculate the accuracy (Accuracy)

$$Acc = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Afterwards the accuracy is calculated (Precision ):

$$Precision = TP / (TP + FP) \quad (2)$$

Recall is then calculated:

$$Recall = TP / (TP + FN) \quad (3)$$

Where TP is positive class determined as positive, FP is negative class determined as positive, FN is positive class determined as negative, and TN is negative class determined as negative. The higher the recall, the more the model is able to detect the target correctly.

For each category, the Precision and Recall are calculated for different thresholds and plotted on the Precision-Recall curve. On the basis of the Precision-Recall curve, a formal evaluation is obtained by calculating the average of the Precision values corresponding to each Recall value:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_i + 1) \quad (4)$$

where  $r_1, r_2, \dots, r_n$  are the Recall values corresponding to the first interpolation at the first interpolation of the Precision interpolation segment in ascending order. The AP of all categories is the mAP:

$$mAP = \sum_{i=1}^k AP_i / k \quad (5)$$

From Fig.1. it can be seen that the higher the average accuracy, the better the model detects under different categories.

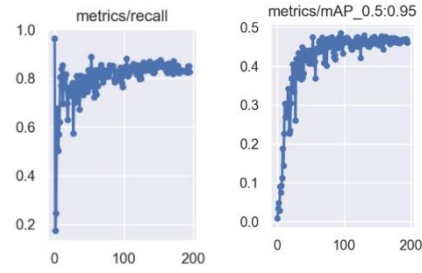


Fig.1 Precision-Recall Curve

### 3.4. Trace detection

By processing the video stream from the camera to obtain the relevant information about the person and locking it, using the Mean Shift algorithm, the user feature values can be extracted in the subsequent processing of the image to avoid following the target confusion.

The Mean Shift algorithm is a non-parametric clustering algorithm based on kernel density. The algorithm assumes that the datasets of different clusters follow different probability density distributions, finds the direction of the fastest density growth at any local point, and finds the region with high sample density corresponding to the maximum value of the distribution.

Therefore, the Mean Shift algorithm flow is:

- (1) Calculate the mean drift vector for each sample:

$$m_h(x)$$

- (2) For each sample point with  $m_h(x)$  Perform a translation, :

$$x_i = x_i + m_h(x_i) \quad (6)$$

- (3) Repeat (1)(2) until the sample points converge:

$$m_h(x) = 0 \quad (7)$$

- (4) Samples that converge to the same point are considered members of the same cluster class.

## 4. Presentation of Results

### 4.1. Training results

First we simply constructed a confusion matrix as in Fig.2 that facilitates better analysis of important indicators of model performance later.



Fig.2 Confusion matrix

After completing the training of the model, we then analyzed the accuracy, precision, and recall. Fig.3 is the F1

curve, positive, peace, and passive three curves are basically the same, with the growth of confidence firstly increased, then stable and finally decreased. Fig.4 and Fig.5 are the curves of precision and recall, respectively.

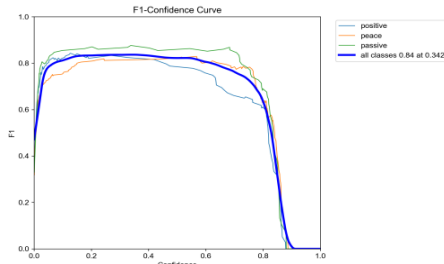


Fig.3 F1 curve

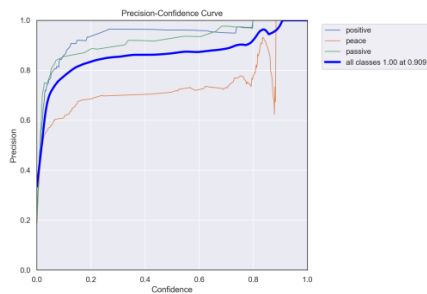


Fig.4 Accuracy curve

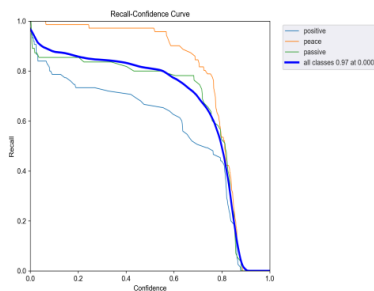


Fig.5 Recall curve

After that the trained model can analyze the expression of each baby photo as in Fig.6 in which each photo indicates the state as well as the degree of the moment.



Fig.6 Partial identification

#### 4.2. Functional realization

The full range of core functions of a baby monitoring system, including nighttime light boosting, image segmentation, and target tracking.

##### (1) Nighttime Light Enhancement

On the premise of the original clarity, the product traverses the pixels to brighten their gray values, adding a

nighttime brightening effect to make the picture brighter and more stable in low-light conditions. As Fig.7 shown.



Fig.7 Lightening treatment

##### (2) Image segmentation

Gradient flow field based image segmentation is accomplished with watershed function to improve the accuracy of face recognition. Using image segmentation to extract the local features of the face can reduce the interference of background information on the face recognition algorithm and improve the accuracy of recognition. As Fig.8 shown.

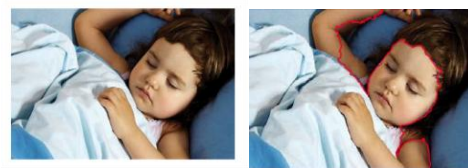


Fig.8 Segmentation Processing

##### (3) Goal tracking

The Mean-shift algorithm is mainly used to mean shift the image to track the position of some regions in the video frame.

The facial recognition system continuously tracks and recognizes the baby's face in a video stream or sequence of images and can more accurately locate facial features, thereby improving the accuracy of facial recognition. As shown in Fig.9 .

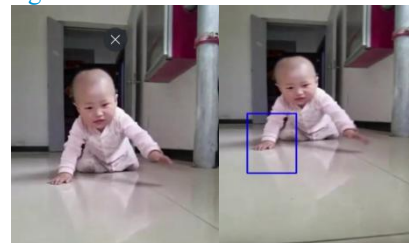


Fig.9 Tracking Processing

## 5. Conclusion

In this paper, a baby monitoring system is successfully built by establishing a database, training the model and applying the yolo algorithm program, and finally realizing the functions of nighttime light enhancement, image segmentation, target tracking, and baby face detection.

## References

1. Simeng Yan, Wenming Zheng, Chuangao Tang, et al, "ARL-IL CNN for Automatic Facial Expression Recognition of Infants under 24 Months of Age." *Journal of Physics: Conference Series*, vol. 1518, 2020, p. 012027
2. Yue Sun, Caifeng Shan, Tao Tan, et al, "Detecting Discomfort in Infants through Facial Expressions." *Physiological Measurement*, vol. 40, no. 11, 2019, p . 115006

---

---

### Authors Introduction

Ms. Peng Wang



She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

Mr. Jiale Jia



He was admitted to the School of Electronic Information and Automation at Tianjin University of Science and Technology in 2022. He is currently pursuing his undergraduate studies at Tianjin University of Science and Technology.