文章编号:1001-9014(XXXX)XX-0001-14

DOI:10. 11972/j. issn. 1001-9014. XXXX. XX. 001

YOLO-Fastest-IR:面向红外热像仪的超轻量级热红外人脸检测网络

李希才[1], 朱嘉禾[2], 董鹏翔[1], 王元庆[1*]

(1. 南京大学 电子科学与工程学院,江苏 南京 210023;
2. 南京大学 智能科学与技术学院,江苏 苏州 215163)

摘要:本文介绍了一种基于ARM CPU的高速鲁棒的双波段热成像测温相机,该测温仪由低分辨率长波红外探测器、数字温湿度的传感器和CMOS传感器组成。针对热红外图像中人脸与背景对比度大的现象,本文探索了一种平衡了人脸检测精度与速度的折衷方案,并提出了一个超型轻量级热红外人脸检测,将之命名为YOLO-Fastest-IR。基于YOLO-Fastest设计了四种不同尺度的热红外人脸检测器YOLO-Fastest-IR0至IR3。为了对4个超轻量级网络训练和测试,本文还设计了一套多用户低分辨率热人脸数据集(RGBT-MLTF),并对四个网络完成了训练。实验表明,轻量级卷积神经网络在热红外人脸检测任务中表现出色。该算法在定位精度和速度上均优于现有的人脸检测算法,且更适合部署在移动平台或嵌入式设备中。在红外图像(IR)中获取感兴趣区域后,根据热红外人脸检测结果对RGB相机进行引导,实现RGB人脸的精细定位。实验结果表明,YOLO-Fastest-IR在树莓派4B上的帧率高达92. 9 FPS,在RGBT-MLTF测试集中人脸定位成功率达97. 4%。最终实现了低成本、强鲁棒性和高实时性的测温系统集成,测温精度可达0. 3℃。

关 键 词:热成像测温相机; YOLO-Fastest-IR; 热红外人脸检测; 超轻量级目标检测

中图分类号:TP18

YOLO-Fastest-IR: Ultra-lightweight thermal infrared face detection method for infrared thermal camera

LI Xi-Cai[1], ZHU Jia-He[2], DONG Peng-Xiang[1], WANG Yuan-Qing[1*]

(1. School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China;
2. School of Intelligence Science and Technology, Nanjing University, Suzhou 215163, China)

Abstract: This paper presents a high-speed and robust dual-band infrared thermal camera based on an ARM CPU. It is composed of a low-resolution long-wavelength infrared detector, a digital temperature and humidity sensor, and a CMOS sensor. In view of the phenomenon of large contrast between face and background in thermal infrared image, this paper we search for a suitable accuracy-latency tradeoff for thermal face detection and propose a tiny-lightweight detector named YOLO-Fastest-IR. Four different scale YOLO-Fastest-IR0 to IR3 thermal infrared face detectors based on YOLO-Fastest are designed. To train and test four tiny-lightweight models, a multi-user low-resolution thermal face database (RGBT-MLTF) is collected, and the four networks are trained. Experiments reveal that the lightweight convolutional neural network can also perform well in the thermal infrared face detection task. And the algorithm is superior to the existing face detection algorithms in positioning accuracy and speed, which is more suitable for deployment in mobile platforms or embedded devices. After obtaining the region of interest in the infrared image (IR), the RGB camera is guided by the results of thermal infrared face detection, to realize the fine positioning of RGB face. The experimental results show that YOLO-Fastest-IR has a frame rate of 92. 9 FPS on a Raspberry Pi 4B and can successfully locate 97. 4% of the face in the RGBT-MLTF test set. The integration of infrared temperature measurement system with low cost, strong robustness and high real-time performance was ultimately achieved, the temperature measurement accuracy can reach 0. 3 degrees Celsius.

Key words: Infrared thermal camera, YOLO-Fastest-IR, Thermal face detection, Tiny-lightweight detector

PACS:

Foundation items: Supported by the Fundamental Research Funds for the Central Universities (2024300443) and the Natural Science Foundation of Jiangsu Province (Grants No BK20241224)

1 Introduction

Infrared thermal camera (ITC) has attracted widespread attention from various sectors of society due to their characteristics of large-scale rapid screening, automatic tracking, high-temperature area alarm, and fusion of visible light images, which can quickly track high-temperature individuals in the crowd [1-2]. During the epidemic period, it was widely used for inspection and quarantine in crowded public places such as airports, nucleic acid testing entrances, subway and train stations, and shopping centers. This approach not only reduces the chance of cross-infection between individuals but also prevents personnel congestion caused by large-scale temperature detection. In addition, it can also be used for monitoring chemical heat sources and real-time monitoring of animal body temperature on farms [3-4].

Face detection is the key technology for ITC, a set of high-speed, stable, low-cost and robust face detection algorithms can enable users to detect faces well under different conditions, and achieve accurate temperature measurement, which significantly affects the performance of ITC. In recent decades, despite substantial progress have been achieved in face detection, there have been many reports on infrared temperature measurement, numerous models have been proposed for thermometer [5-7], it remains a very challenging work for accurately and quickly locating faces in infrared images. However, most previous methods only used a single thermal infrared camera for rough facial detection based on morphology [8], or facial localization based on visible light images. In other words, the thermal camera first performed face detection through visible spectrum images, and then mapped the detected face position to infrared images for temperature measuring [9]. And they are hard to be directly detected from IR images, the disadvantage of using RGB camera is that it is easily affected by ambient light [10]. Moreover, some human-shaped objects (e. g. , tiny pillars and blurry traffic lights) have similar appearances to face [11] and they are easily mistaken for thermometer. As a result, many false alarms appear in the ITC, which affects the practical application of the temperature measuring camera. In general, RGB images cannot guarantee the high-quality for face detection, and more comprehensive information should be explored for thermometer.

Most ITC usually take high-resolution images as input to achieve high recall, which usually rely on costly graphic processing units (GPUs) for low latency [12]. For our knowledge, few works before have report lightweight ITC. Limited by infrared face detection technology and data set, Negishi and others use a mature face detection algorithm in the visible light image to locate the faces [13], and then map the detected faces coordinate to the corresponding infrared image for temperature measurement. In addition to the shortcomings of visible light face detection, this method also has the shortcomings of inaccurate coordinate mapping, high computing time and low frame rate.

Chaitra Hegde et al. conducted PoseNet based forehead positioning temperature measurement and cyanosis inspection at edge computing platform raspberry pi [5]. The disadvantages of this method are the forehead and lip detection were computed on the Google Coral USB accelerator. This system not only has slow facial detection speed and poor positioning accuracy, but also high cost, only with the help of the Accelerator TPU of Google Coral′s Coral USB accelerator neural network accelerator can the near real-time effect be barely achieved.

At present, visible face detection tasks are usually based on the MS COCO dataset, while visual tasks faces can be trained and tested using datasets such as Helen [14], IBUG [15], and 300-W [16]. For thermal infrared visual tasks, UND [17] was the earliest thermal infrared facial dataset proposed in 2003, followed by commonly used datasets such as IRIS [18] and NVIE [19]. In 2021, Domenick Poster et al. [20] proposed the latest thermal infrared facial dataset ARL-VTF and listed most of the previous thermal infrared facial datasets. The existing thermal infrared facial datasets are mainly aimed at tasks such as facial recognition and emotion recognition. Therefore, in the dataset, a single face occupies most of the image, and there are few or no interference factors in the background, resulting

in usually high image resolution. The existing thermal infrared facial detection algorithms based on convolutional neural networks most are using trained single user thermal infrared facial images obtained by the author themselves, and these datasets are not publicly available. Within the scope of the author's understanding, there is no thermal infrared face dataset suitable for multi user face detection tasks.

Woongkyu lee et al. based on SSD model in 2021 proposed a temperature measurement method [21], they customize SSD to identify the location of human faces via transfer learning, which can run to 160 FPS on NVIDA Jetson AGX, it directly detects faces in infrared images. The disadvantage of this method is that when there are multiple targets, it is difficult to accurately locate suspicious high-temperature targets in visible images. Friedrich et al. [22] proposed the eye corner detection algorithm for thermal infrared face detection based on the characteristics of the highest eye temperature and the lowest face temperature. Reese et al. [23] calculated the gray projection curve of the thermal infrared image, and determined the face range by analyzing the gray projection curve and its first derivative, which is called gray projection analysis (Projection Profile Analysis, PPA). Marcin Kopaczka et al. [24] analyzed and compared two detection algorithms for thermal infrared images, along with five algorithms predominantly used for visible light face detection. These include the Viola Jones algorithm [25], a variant Viola Jones algorithm replacing Harr features with local binary pattern features, and a face detection approach combining directional gradient histograms with support vector machines [26-27]. Deformable component model [27] and pixel intensity comparisons organized (PICO) in decision trees [28].

Experiments have shown that these machine learning algorithms mainly used for visible light face detection have higher accuracy and lower false positives compared to detection algorithms proposed for thermal infrared images. The detection algorithm proposed for thermal infrared images has a shorter running time, but the PICO algorithm has the highest computational efficiency.

In recent years, researchers have also applied deep learning to thermal infrared facial detection tasks. In 2017, Alicja et al. adjusted the InceptionV3 network and removed the global pooling operation [29], allowing the last 8×8 feature map to be classified on 64 grids separately, using a set of grids with a face probability greater than 0.5 as the face range. However, their research is limited to single user thermal infrared face detection. In 2019, Silva et al. [30] adjusted the YOLOv3 [31] network, trained YOLOv3 using a thermal infrared facial dataset, and cut off the last predicted feature map during detection, achieving high accuracy and efficiency. The target detector based on YOLO has the ability of multi-target detection, but their goal is to detect the driver in the autopilot system, so their data set also contains only a single user. However, most previous methods directly used infrared images to train general models for detecting targets in infrared images, but the models had a lot of redundant information and could not improve detection speed. In this work, we found that combining the characteristics of infrared images with composite scaling of the model can greatly improve the efficiency of the model and promote future research in this field.

In this paper, we propose a high-speed and robust dual band face detection system on ARM CPU for ITC. Our ITC mainly consists of a low-resolution infrared detector, an CMOS sensor and an environmental temperature monitoring sensor. The infrared camera is used to locate the face in the scene, and the visible camera is used to further confirm the face information corresponding to the infrared image. At the same time, the attenuation compensation of thermal radiation with distance is considered, and the stereo ranging is carried out by data set fitting. In addition, focusing on this problem, we propose an ultra-lightweight thermal infrared face detection network in terms of algorithm, explore the impact of several different network models on thermal infrared face detection. And at the same time, to train the proposed model, we designed a dual band face detection dataset, we introduce a large-scale RGBT dual light dataset, which contains 2,030 pairs of RGB-thermal images with

138,389 annotated faces. To verify the proposed model, we designed a set of experimental prototypes, and deployed the algorithm to the raspberry pi system with ARM CPU. An ultra-lightweight infrared face detection network suitable for thermal infrared face detection is found by composite scaling of network depth, resolution and width.

2 Principle of the temperature measurement system

As shown in Fig. 1, a binocular stereo vision system composed of an infrared camera and an RGB camera. The raw infrared data obtained by the IR camera is separated into two branches for further processing. On the one hand, the raw sixteen-bit data is dynamically normalized to an eight-bit gray value according to the maximum and minimum temperature, the conversion rule as shown in Eq. (1). Where I_{gray} is the gray value of infare image, $RawData[i]$ is 16 bits infrared raw data, $MaxValue$ and $MinValue$ are the maximum and minimum values in the current frame's infrared data. On the other hand, it is used as the backup data of temperature measurement. The original resolution of infrared image is 80×60 pixels, and the resolution is resized to 160×120 after interpolation. The gray image is directly used as the input of the face detector to obtain the region of interest for infrared temperature measurement.

$$\left[I_{gray}\right] = \left(RawData[i] - MinValue\right) * \frac{255}{MaxValue - MinValue} \tag{1}$$

After obtaining the face region in the thermal infrared image, the region of interest is synchronously mapped to the visible image, thereby achieving the task of facial detection or identity recognition in the visible light image, and the amount of visible data is greatly compressed. In addition, based on the geometric relationship of binocular stereo vision composed of infrared and RGB cameras, the distance between the measured individual and the camera can also be obtained. Finally, the temperature is corrected and compensated according to the distance and environmental information to improve the temperature measurement accuracy.

One of the advantages of our RGB face localization is splitting a complicated real-world computer vision task into two easier ones that can be well solved by current deep learning methods. If we stick to a single visible RGB camera for cascaded or simultaneous face detection and eye localization, the input resolution of the CNN will inevitably be large, resulting in a computationally heavy network. In this paper, we make the most of guiding mode by using two tiny-lightweight CNNs. The dual-band infrared guidance system not only largely reduces the computational cost but also maintains high accuracy and robustness. It effectively addresses the trade-off between high tracking speed, high tracking accuracy, and strong robustness in conventional visual tracking systems.

3 Thermal infrared face detect

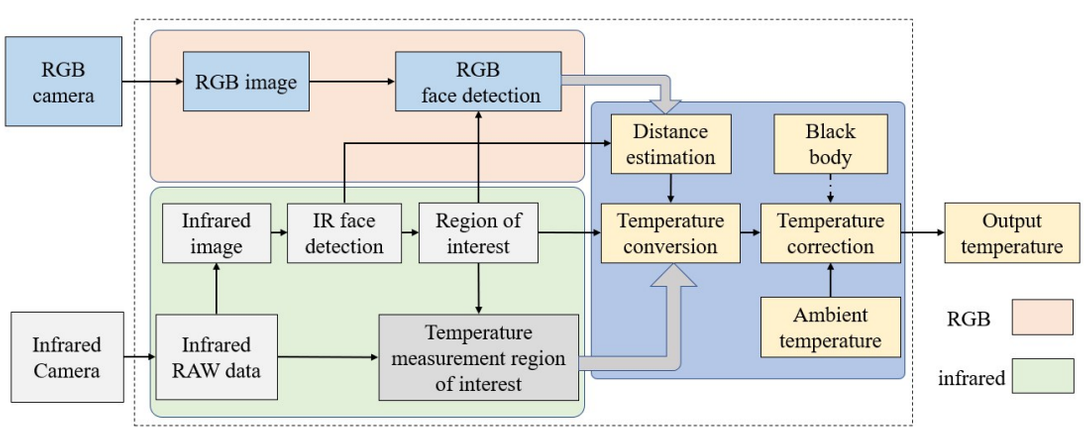


Fig. 1 The procedure of the working process of the dual band ITC system
图1 双波段红外测温系统组成和工作流程

method

3.1 The Introduction of lightweight network YOLO-Fastest-IR

As the most of the body's thermal radiation is usually blocked by clothing. Therefore, the facial part is often the place with the highest gray value in the image, and the contrast with the background is very high. The face appears as a white oval shape in the IR image, and the facial features are basically indistinguishable. On the one hand, this creates difficulties for highly refined computer vision tasks such as facial recognition, emotion recognition, and face landmark detection. On the other hand, it limits the features that convolutional neural networks can extract. That is, there are basically no effective deep features for thermal infrared facial detection to be extracted by neural networks. In addition, there are significant differences in shape and aspect ratio between the face and other possible heat sources, such as screens, fluorescent lamps, and cups filled with hot water. Therefore, theoretically, lightweight convolutional neural networks can stably detect faces in IR images.

As shown in Fig. 2, to verify the hypothesis proposed above, we designed four tiny lightweight convolutional networks with different complexity levels, inspired by YOLO-Fastest [32]. The red block and blue block in the figure represent the lite convolution module and lite residual module, respectively. This article explores the relationship between network scale in three aspects: resolution, depth, and channel number.

In terms of resolution, current object detection networks are mostly designed for datasets such as COCO, so the input resolution is relatively high, ranging from 416 to 800. To enhance the computational efficiency of lightweight object detection networks, reducing the input image resolution has become a common strategy for speed improvement. If directly down sampling to a size of 320×240, image details will inevitably be lost. However, the infrared camera used in this article has a physical resolution of only 160×120, so using bilinear interpolation to enlarge the image to 320×240 may not have practical significance. To be compatible with the physical resolution of infrared cameras, the input resolution of YOLO-Fastest was set to 160×120, and variables were controlled in depth and width. Four different lengths and widths of infrared face detection networks were designed differentially, and these networks were named YOLO-Fastest-IR.

The first 3×3 convolution step in all four networks has a size of 2 and uses zero padding, thus reducing

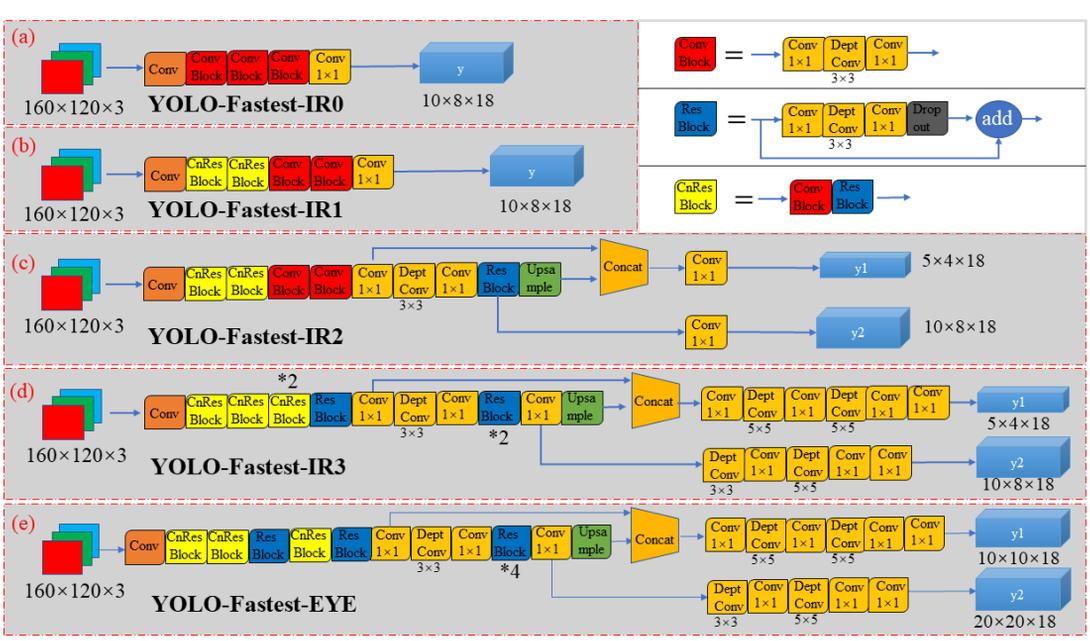


Fig. 2　The YOLO-Fastest-IR network structures with four different levels of complexity
图2　四种不同复杂度的热红外人脸检测网络 YOLO-Fastest-IR 结构

the length and width of the input image to half of its original size. The input and output feature maps of all lite residual modules maintain the same size and are not scaled. The first lite convolution module in IR1 to IR3 does not perform feature map size scaling. In IR0, the first lite convolution module, as well as all other lite convolution modules in the four YOLO-Fastest-IR models, employs depth-separable convolutions with a stride of 2, using zero padding to reduce the length and width of the input image by half. Both IR0 and IR1 undergo four feature map size reductions, resulting in a feature map with a grid size of 10×8. In the four types of YOLO-Fastest-IR, the number of convolutional kernels gradually increases with the increase of network depth. That is, the number of feature map channels increases with the increase of network depth, the specific details are as follows.

(1) YOLO-Fastest-IR0: As shown in Fig. 2 (a), excluding the first and last two convolutional layers of YOLO-Fastest, only three lite convolutional modules were used. The final output grid size is 10×8, which is the network with the least number of layers and the simplest network structure. Due to its small number of layers, there is basically no phenomenon of gradient vanishing and network degradation, so residual modules are not used.

(2) YOLO-Fastest-IR1: As shown in Fig. 2 (b), a residual module was introduced into the backbone network, using 4 lite convolution modules and 2 lite residual modules, the final output grid size was 10×8.

(3) YOLO-Fastest-IR2: As shown in Fig. 2 (c), a multi-scale prediction strategy was introduced in the neck network, and 5 lite convolutional modules and 3 lite residual modules were used. The final output was two feature maps with grids sizes of 5×4 and 10×8 respectively, responsible for predicting large and small targets. Only one convolutional layer was used in the head network.

(4) YOLO-Fastest-IR3: As shown in Fig. 2 (d), further deepen the network layers and use 6 convolutional layers in the head network, ultimately outputting two feature maps with grids of 5×4 and 10×8, making it the network with the highest number of layers and the most complex structure.

(5) YOLO-Fastest-EYE: The overall network structure is shown in Fig. 2(e). The aspect ratio of the facial bounding box is roughly close to 1:1. The RGB image resolution is 640×480, when the user's face is about 1 meter away from the camera, the bounding box size is about 160×160. Therefore, it is specified that the image size input by YOLO-Fastest-EYE is set to 160×160. After four down sampling and one up sampling of the feature map, the final output is a tensor of size 20×20×18, the ratio of each grid to the length and width of the entire feature map is 5%. Since only targets such as the eyes are predicted, the number of output feature map channels is 18.

3.2 Infrared face detection dataset and model training

To complete the training of the proposed network, we also designed an RGBT multi user low resolution thermal face database (RGBT-MLTF) in this paper. This dataset contains 26800 images captured by an infrared camera lepto3.0 with a resolution of 160×120. Each image contains 1 to 4 faces, with no less than 2 faces accounting for 76% of the dataset. As shown in Fig. 3, to improve the generalization ability of the model and prevent overfitting, we conducted long-term experiments under different environmental lighting and temperature conditions, including weak light conditions, high exposure scene, high temperature environment, and low temperature environment. The dataset covers almost all conventional application scenarios. The dataset is annotated using labeling, which approximates the head as an ellipse and labels the outer tangent rectangle of the ellipse as a real face rectangle. The dataset annotates distant faces, incomplete faces, and lateral faces, but the back of the head is not marked. The final dataset contains a total of 5102 faces from 22 people.

Among 2680 images, 1627 images were randomly selected as the training set, 520 were used as the cross validation set, and 533 were used as the testing set. The training set is used to train the four proposed thermal infrared face detectors, the cross-validation set is

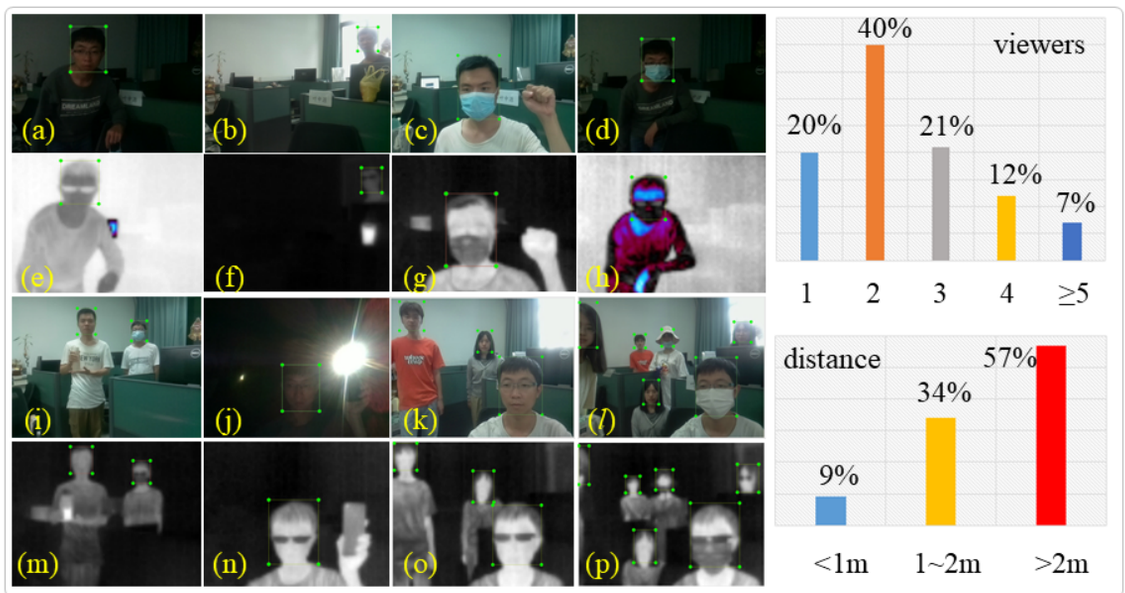


Fig. 3　Examples and statistics of the RGBT-MLTF dataset
图3　部分RGBT-MLTF数据集示例和统计信息

used to select the optimal network weights from several training sessions, and the test set is used to test and compare the performance of different neural network models.

4　Experiment and discussion

The infrared sensor used in this article is the lepton 3.5 from FLIR, with a resolution of 160×120. The visible light module uses a Raspberry Pi camera, and its sensor is OV5647 produced by OV company, with a resolution of 1280×720. Due to the differences in resolution and position between IR cameras and RGB cameras, the horizontal and vertical coordinates of the thermal infrared facial bounding box cannot be directly used in visible images. Therefore, we fitted the positional relationship between infrared and RGB images using the annotated RGBT-MLTF dataset. In short, the relationship between the annotated IR face boxes and the corresponding RGB face boxes is directly calculated by Eq. (2).

$$X_{RGB} = 0.8812 * X_{IR} + 0.0844 \tag{2}$$

where X_{IR} is the x coordinate of the original IR box, X_{RGB} is the x coordinate of the shifted RGB box. Both X_{IR} and X_{RGB} are relative values to the width of images. The x coordinates of IR and RGB face bounding boxes are on the horizontal and vertical axis respectively. Each of the sampling point represents a pair of face bounding boxes in the two spectrum images. with linear regression, the relationship is obtained as shown in Eq. (2).

In this article, we use the RGBT-MLTF dataset to train YOLO-V4 and YOLO-V8s [33], YOLO-Fastest, and the four proposed thermal face detection networks YOLO-Fastest-IR, respectively. And compared the prediction performance of the four proposed network models with the current state of the art object detection algorithms. As shown in the first column of Fig. 4(a1~g1), even in scenes with interference, the four designed thermal infrared face detection networks can stably detect faces in images. YOLO-Fastest-IR0 and YOLO-Fastest-IR1 have certain false positives, which can detect the monitors or raised fist in the background as a face. As shown in Fig. 4(a2~g2), in a multi-user scenario, the four networks proposed in this paper with different scales can not only effectively detect thermal infrared faces of different sizes, but also detect half faces with occlusion. As shown Fig. 4(a3~g3), except for the shortest YOLO-Fastest-IR1 and YOLO-Fastest-IR0, which occasionally have some false positives, the other five networks can effectively detect faces in thermal infrared images.

To further verify the generalization ability of the

four YOLO-Fastest-IR, we conducted generalization tests on the high-resolution thermal infrared facial dataset provided by Marcin et al [34], and the test results are shown in Fig. 4(a4~g4). The detection results show that several IR face detectors have certain generalization ability. Among which, YOLO-Fastest-IR0, YOLO-Fastest-IR1 predicted the boundary box has a certain deviation from the real face, YOLO-Fastest-IR2, YOLO-Fastest-IR3 and YOLO-V8s can accurately extract the thermal infrared face region from the thermal infrared face image that is significantly different from the training set, and the confidence is close to 100%. However, YOLO-V4 is no longer able to predict the image under commonly used confidence thresholds. When the confidence threshold is reduced to 0.1, the results in Fig. 4(f4) are outputted, this indicates that YOLO-V4 may have some over fitting to the training set. As shown in the Fig. 4(a5~g5), although the data set only contains a fewer number of pseudo color samples, the IR face detection network YOLO-Fastest-IR0 and YOLO-Fastest-IR1 designed in this paper can still predict the position of the face in the pseudo color image. However, YOLO-V8s cannot accurately predict faces from pseudo color images.

Fig. 5(a) shows the experimental results of RGB face detection and eye location guided by thermal infrared proposed in this paper. First, the position of the face in the infrared image is determined by YOLO-Fastest-IR, and then the YOLO-Fast-EYE is used to predict the position of the eyes based on the region of interest determined by thermal infrared. The result show that YOLO-Fastest-IR can better solve the localization problem of occluded faces. It proves that the infrared guidance effectively avoids the abnormal temperature measurement of the ITC, thus reducing the false alarm rate. Fig. 5(e~h) shows the results of RGB face and eye detection using YOLO-Fastest-EYE under poor lighting conditions. Fig. 5(i~l) shows the results of multi-target face detection. In summary, the algorithm presented in this paper demonstrates strong robustness, capable of handling diverse lighting conditions and occlusion interferences, while also supporting the detection of multiple targets.

The comparison results of AP values and FPS for several different networks are shown in Fig. 6. It can be seen that except for YOLO-V4, YOLO-V8s and YOLO-Fastest, four YOLO-Fast-IR can all meet the real-time detection standards, and their AP50 values exceed 90%. The prediction performance of the lightweight network is close to that of YOLO-V4. The AP values was tested based on the RGBT-MLTF dataset, and the frame rate was measured on the Raspberry Pi 4B CPU. By compressing the network, the inference speed of the network can be greatly improved, but the accuracy has not significantly decreased. This also fully demonstrates that for infrared facial detection tasks, a lightweight convolutional neural network with a simple structure is enough to extract infrared facial features and complete facial localization. Based on the experimental results, it can be observed that the overall trend of YOLO-Fastest-IR is: the deeper the network, the slower the inference speed, but the higher the average precision. YOLO-Fastest-IR2 in Fig. 6 has achieved a better precision mean than the deep network, which may be caused by the accidental factors during training that lead to the better convergence result of the network.

In the visual task of object detection, 30 frames per second is usually used as the standard to divide real-time and non-real-time. YOLO-V4 and YOLO-V8s can achieve the real-time operation standard on the GPU, but it takes more than 1 second to complete the reasoning of a single picture on the Raspberry Pi 4B, which cannot meet the real-time requirements. YOLO-V4 can reach 98.95% AP50 on RGBT-MLTF dataset after training, which is far more than the highest 81.3% AP50 on MS COCO dataset. As mentioned above, the RGBT-MLTF dataset covers as many factors as possible that may interfere with thermal infrared face detection. Compared with the images captured in the actual application scenario of the ITC system in this paper, the images in the RGBT-MLTF dataset are more difficult for the detector. Through the performance of the above several networks on the dataset, it is proved that the thermal infrared face detection task in the application scenario of ITC can be better com-

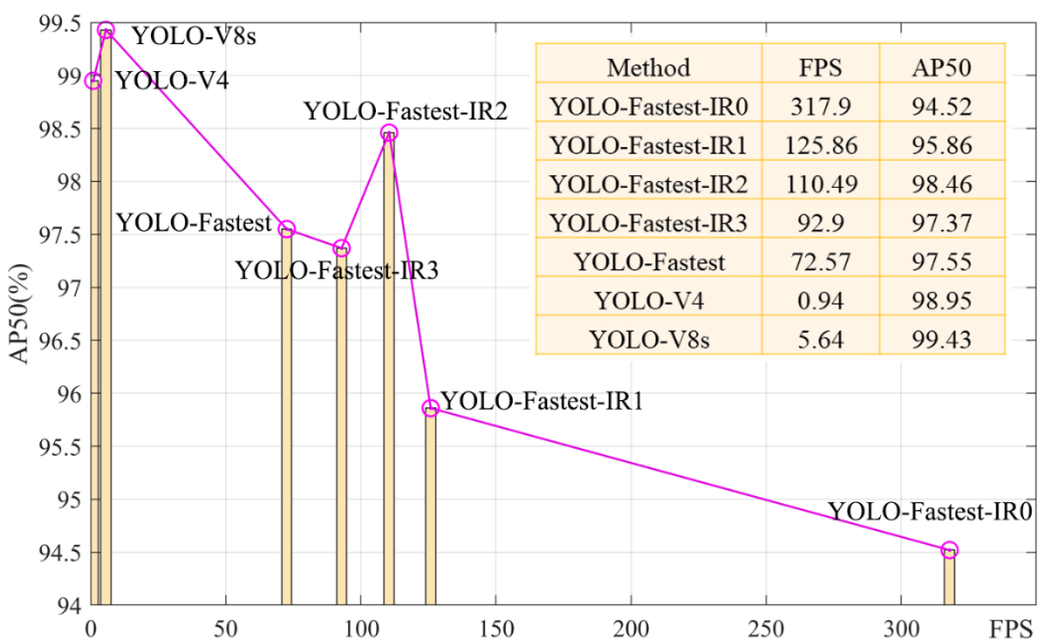
Fig. 4　The comparative and generalization experiments of different models: (a1~a5) the detection effect of YOLO-Fast-IR0; (b1~b5) the detection effect of YOLO-Fast-IR1; (c1~c5) the detection effect of YOLO-Fast-IR2; (d1~d5) the detection effect of YOLO-Fast-IR3; (e1~e5) the detection effect of YOLO-Fast; (f1~f5) the detection effect of YOLO-V4; (g1~g5) the detection effect of YOLO-V8s.
图4　不同模型的对比测试和泛化实验结果分析

pleted by using deep learning.

4.3　Temperature measurement experiments

In this section, we will introduce the details of temperature measurement. We set the black body temperature as the vertical axis and the raw data captured by IR camera as the horizontal axis. The black body temperature value is measured several times (20 times) and take the average value. The relationship between the temperature values and the raw IR data is the blue curve in Fig. 7 and its linear fitting curve is the magenta dashed line. To verify the relationship between thermal radiation and distance, we conduct tests at intervals of 25 cm, starting from 25 cm to 225 cm. At each interval, we perform 20 tests and take the average value. The relationship curve is shown as the red curve in Fig. 7, and the linear fitting result is a green dotted line. The fitting result is shown in Eq. (3). When i=1 or i=2, it represents the relationship between temperature and grayscale or distance respectively. After fitting $a_1 = 19.645$, $b_1 = 0.1163$, $a_2 =$

Fig. 5 IR face detection samples of YOLO-Fastest-IR and eye localization in the RGB images results: (a~d) YOLO-Fastest-IR and YOLO-Fastest-Eye is robust against angle of face inter-viewer occlusion, environmental occlusion; (e~h) extreme lighting condition; (i~*l*) multi target and distant viewers.
图5 YOLO-Fastest-IR热红外人脸检测结果及在RGB图像中人眼定位效果



Method	FPS	AP50
YOLO-Fastest-IR0	317.9	94.52
YOLO-Fastest-IR1	125.86	95.86
YOLO-Fastest-IR2	110.49	98.46
YOLO-Fastest-IR3	92.9	97.37
YOLO-Fastest	72.57	97.55
YOLO-V4	0.94	98.95
YOLO-V8s	5.64	99.43

Fig. 6 Comparison of the proposed YOLO-Fastest-IR and other object detectors on the RGBT-MLEL face subset
图6 YOLO-Fastest-IR与主流目标检测器在RGBT-MLEL面部数据集上的测试结果对比

37.514, $b_2 = -0.00794$.

$$y = a_i + b_i x \tag{3}$$

According to Planck's radiation law, the gray value of each pixel in the infrared image is proportional to the thermal radiation energy of the corresponding point on the surface of the measured object. But the temperature captured by the thermal imager is the radiation temperature T_r of the object surface, not the real temperature T_0 of the object. The real temperature of the object is equal to the temperature of the blackbody radiating the same energy. Therefore, in actual temperature measurement, it is necessary to calibrate the thermal imager with high-precision blackbody to find the mapping relationship between blackbody temperature and sensor output voltage, the temperature of the black body is preset, the relationship between T_r and

T_0 as shown in Eq. (4).

$$T_0 = \{\frac{1}{\varepsilon}[\frac{1}{t_a}T_r^{\lambda} - (1-\varepsilon)T_u^{\lambda} - (\frac{1}{t_a}-1)T_a^{\lambda}]\}^{\frac{1}{\lambda}} \quad (4)$$

Where λ is a wavelength dependent parameter, which varies depending on the material of the IR camera used. For InSb (3-5 μ m) detectors, the value of λ= 8. 68; For HgCdTe (6-9 μ m) detectors, the value of λ=5. 33; The detector we use is HgCdTe (8-14 μm), so the λ=4. 09; Due to the fact that the impact of atmospheric transmittance can be ignored in close range temperature measurement, so t_a=1, and Eq. (5) can be obtained.

This is the temperature measurement calculation formula for the surface of the gray body. In the Eq. (5), where ε is the skin emissivity, usually taken as 0. 98, T_r is the thermal radiation detected by the infrared detector, T_u is the ambient temperature, which can be measured by temperature and humidity sensors. Therefore, the temperature at the forehead can be estimated.

$$T_M = \frac{1}{\varepsilon} * (T_r^{\lambda} - (1-\varepsilon) * T_u^{\lambda})^{\frac{1}{\lambda}} \quad (5)$$

To further verify the proposed method, we design and carry out the temperature measurement experiment based on our binocular vision system. It can be seen in Fig. 8, whether in close or long distance, single or multi person scenarios, the algorithm can stably detect the position of the face and accurately measure the temperature of the forehead. Meanwhile, as shown in Fig. 5 and Fig. 8, the proposed IR face detection algorithm

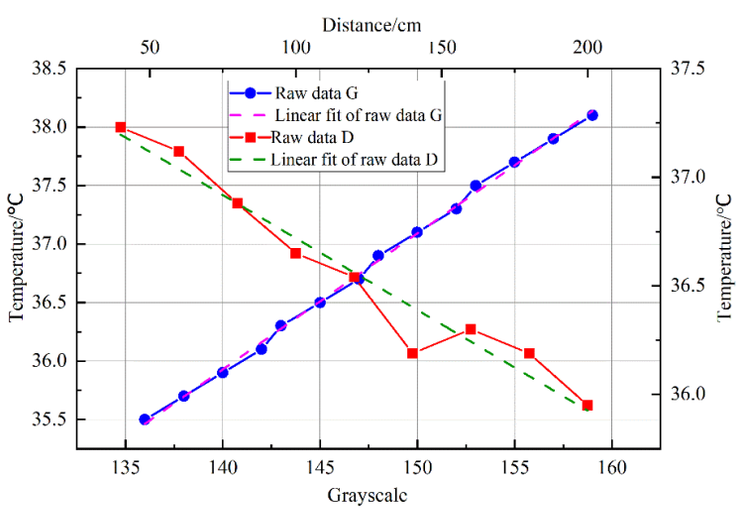


Fig. 7　The variation of grayscale values with temperature at different distances
图7　不同距离下灰度值随温度的变化关系

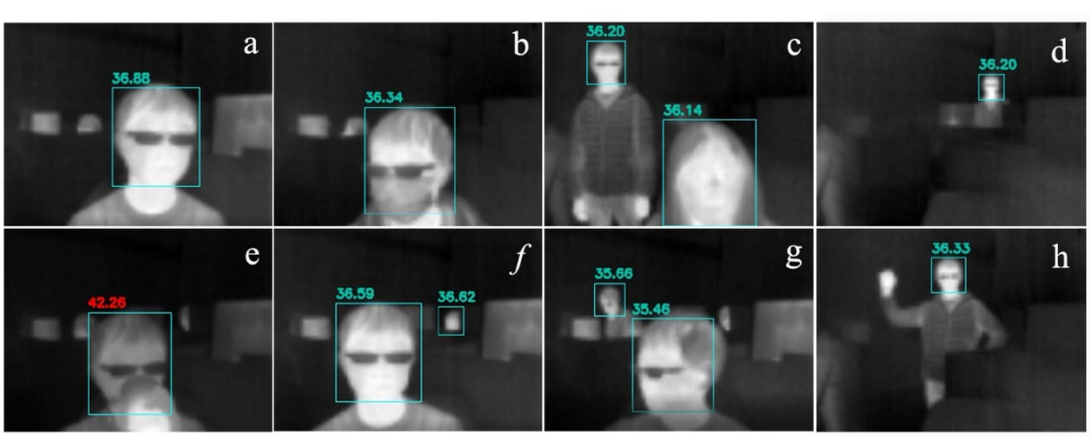


Fig 8　The real time temperature measurement experiment of temperature measurement system: (a) normal sitting and standing; (b) wear a mask; (c) interference testing at different distances; (d) remote temperature measurement experiment; (e) high temperature warning test; (*f*) remote multi-target temperature measurement experiment; (g) side face test; (h) fist interference experiment.
图8　红外测温系统的实时温度测量实验

also has strong robustness and can effectively detect faces wearing masks, occlusion, and different poses. Fig. 8(e) shows the high temperature warning examples caused by a part of the cup body entering the facial area during the process of drinking hot water. To avoid this phenomenon in practical applications, we will accurately locate the temperature measurement area and position combine with the position of the eyes.

In order to verify the temperature measurement accuracy of ITC, it was tested using a blackbody at different ambient temperatures and distances, each data point was measured five times, and took the average. As shown in Fig. 9(a), regardless of the ambient temperature, the measured temperature decreases linearly with distance increasing, and the higher the ambient temperature, the higher the measurement result. Then, we gather several volunteers for the experiments and used a forehead gun as a controlled experiment. The test results are shown in the Fig. 9(b), the experiment covered 11 sets of temperature measurements at different distances. With the distance increases, the original temperature measured by the infrared camera decreased, as shown the blue curve in Fig. 9(b). After distance correction, the measurement results are basically consistent at different distances, as shown the green curve in Fig. 9(b). At close range, the measurement accuracy of the ITC in this article is basically consistent with the measurement accuracy of the forehead gun. With the increase of distance, the measurement accuracy of the infrared temperature measurement camera is better than that of the forehead gun, which not only has high temperature measurement accuracy but also achieved good repeatability stability, the temperature measurement accuracy can reach 0.3 degrees Celsius.

5 Conclusion

In this paper, we develop a dual band ITC, it can be used to measure the temperature of forehead which is composed of an infrared detector and a RGB sensor. In addition, the thermometer also integrates a temperature and humidity sensor for sensing environmental temperature and humidity. Accordingly, this paper also proposes four tiny-lightweight thermal infrared face detectors of different scales, namely YOLO-Fastest-IR0 to YOLO-Fastest-IR3. Through training and testing the above models in the RGBT-MLTF dataset proposed in this paper, it is proved that YOLO-Fast-IR is more suitable for deployment in mobile device and edge computing embedded platform than existing algorithms such as YOLO-V4 and YOLO-Fastest, and its tiny version runs fastest. Although the accuracy of face location is slightly decreased, it also meets the application accuracy requirements of infrared temperature measurement equipment. Among them, YOLO-Fastest-IR0, which has the smallest network scale, cannot complete thermal infrared face detection well due to the lack of deep networks. The average accuracy of other thermal infrared face detectors can reach more than 95%, and the frame rate can reach more than 90 FPS on Raspberry Pi 4B. Compared with YOLO-Fast-

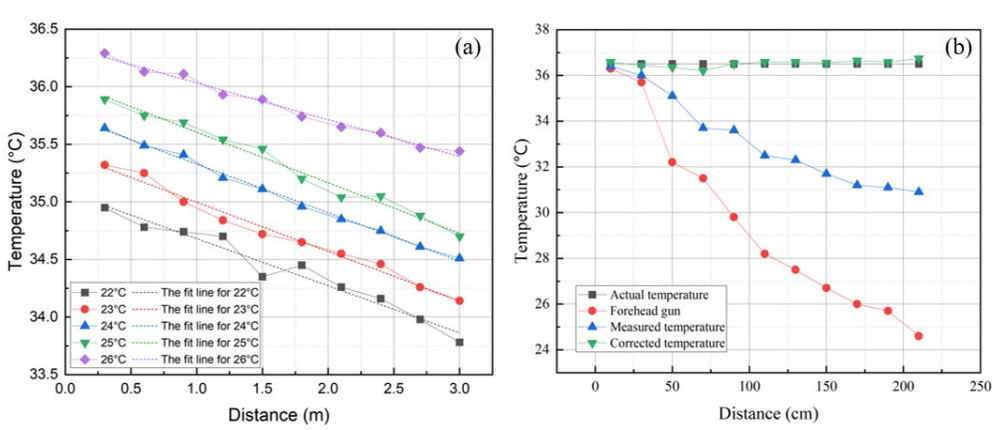


Fig. 9 Analysis of temperature measurement accuracy of ITC: (a) the variation relationship of different temperatures of blackbody under different environmental temperatures and distances; (b) temperature correction experiment.
图9 红外测温系统的测温精度分析

est and YOLO-V4 and YOLO-V8s, it shows that these tiny-lightweight convolutional neural networks have greatly improved the operating efficiency on the premise of less precision loss. According to the different visual tasks, the convolutional neural network can be adjusted in structure to achieve the optimal combination of precision and speed. The experimental results show that our ITC is effective, and the proposed face detection methods have excellent performances.

References

[1] Haghmohammadi H F, Necsulescu D S, Vahidi M, Remote measurement of body temperature for an indoor moving crowd[C]. Proceedings of IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2018,(pp.1-6).

[2] Ring E F J, Jung A, Zuber J, et al. Detecting fever in polish children by infrared thermography[C]. Proceedings of 9th International Conference on Quantitative InfraRed Thermograph (QIRT), 2008.

[3] Ramelan A, Ajie G S, Ibrahim M H, et al. Design low cost and contactless temperature measurement gate based on the internet of things (IoT)[C]. IOP Conference Series Materials Science and Engineering (ICIMECE), 2021, (pp. 1096).

[4] Ye X, Gao S, Li F. ACE-STDN: An infrared small target detection network with adaptive contrast enhancement[J]. J. Infrared Millim. Waves, 2023, 42(5):701-710.

[5] Hegde C, Jiang Z, Suresha P B, et al. AutoTriage - An open source edge computing raspberry Pi-based clinical screening system[J]. medRxiv, 2020, 1-13.

[6] Švantner M, Vacíková P, Honner M. Non-contact charge temperature measurement on industrial continuous furnaces and steel charge emissivity analysis[J]. Infrared Physics & Technology, 2013, 61:20-26.

[7] Ng E Y, Kaw G J, Chang W M. Analysis of IR thermal imager for mass blind fever screening[J]. Microvascular Research, 2004, 68(2):104-109.

[8] Jiři M, Virginia E, Marcos F. Face segmentation: A comparison between visible and thermal images[C]. IEEE International Carnahan Conference on Security Technology, 2010.

[9] Somboonkaew A, Prempree P, Vuttivong S, et al. Mobile-platform for automatic fever screening system based on infrared forehead temperature[C]. 2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC), 2017.

[10] Li X, W Q, Xiao B, et al. High speed and robust infrared-guiding multiuser eye localization system for autostereoscopic display[J]. Applied Optics, 2020, 59(14): 4199-4208.

[11] Mucha W, Kampel M. Depth and thermal images in face detection - A detailed comparison between image modalities[C], Proceedings of the 5th International Conference on Machine Vision and Applications (ICMVA), 2022, (pp.16-21).

[12] Miao Z, Zhang Y, Li W. Real-time infrared target detection based on center points[J], J. Infrared Millim. Waves, 2021, 40(6):858-864.

[13] Negishi T, Sun G, Liu H, et al. Stable contactless sensing of vital signs using RGB-thermal image fusion system with facial tracking for infection screening[C]. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, (pp. 4371-4374).

[14] Helen dataset. http://www.ifp.illinois.edu/~vuongle2/helen/

[15] IBUG dataset. https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/

[16] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: the first facial landmark localization challenge[C]. Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops, 2013.

[17] Chen X, Flynn P J, Bowyer K W. IR and visible light face recognition[J]. Computer Vision and Image Understanding, 2005, 99(3):332-358.

[18] IRIS dataset. https://archive.ics.uci.edu/ml/datasets/Iris/

[19] NVIE dataset. http://nvie.ustc.edu.cn/

[20] Poster D, Thielke M, Nguyen R, et al. A large-scale, time-synchronized visible and thermal face dataset[J]. 2021.

[21] Lee W, Kwon H. Choi J. Thermal face detection for high-speed AI thermometer[C], Proceedings of 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC), 2021,(pp.163-167).

[22] Friedrich G, Yeshurun Y. Seeing People in the Dark: Face Recognition in Infrared Images[J]. Springer Berlin Heidelberg, 2002.

[23] Reese K W, Zheng Y, Elmaghraby A. A comparison of face detection algorithms in visible and thermal spectrums [C]. Proceedings of Int'l Conf. on Advances in Computer Science and Application (CSA), 2012,(pp.49-53).

[24] Kopaczka M, Nestler J, Merhof D. Face detection in thermal infrared images: A comparison of algorithm- and machine-learning-based approaches[C]. Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems, 2017,(pp.518-529).

[25] Viola P A, Jones M J. Rapid object detection using a boosted cascade of simple features[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001,(pp.511-518).

[26] Xiaoyu W, Jihong C, Pingjiang W, et al. Infrared human face auto locating based on SVM and a smart thermal biometrics system[C]. Proceedings of Sixth International Conference on Intelligent Systems Design and Applications, 2006.

[27] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Proceedings of IEEE computer society conference on computer vision and pattern recognition (CVPR), 2005.

[28] Nenad M, Miroslav F, Igor S P, Jörgen A, Robert F. Object detection with pixel intensity comparisons organized in decision trees[J]. arXiv preprint arXiv: 1305.4537, 2014.

[29] Kwasniewska A, Ruminski J, Rad P. Deep features class activation map for thermal face detection and tracking[C]. Proceedings of 10th International Conference on Human System Interactions (HIS), 2017,(pp.41-47).
[30] Gustavo S, Rui M, André F, Pedro C, Luís C. Face detection in thermal images with YOLOv3[C]. International Symposium on Visual Computing (ISVC), 2019, (pp. 89-99).
[31] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[32] Yolo-Fastest. [Online]. Available: https://github. com/dog-qiuqiu/Yolo-Fastest
[33] YOLOv8. [Online]. Available: https://github. com/ultralytics/ultralytics
[34] Kopaczka M, Kolk R, Schock J, Burkhard F, Merhof D. A thermal infrared face database with facial landmarks and emotion labels[J], IEEE Trans. Instrum. Instrum. Meas., 2018, 68:1-6.