文章编号：1001-9014(2025)05-0790-11

DOI:10.11972/j.issn.1001-9014.2025.05.017

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. The system consists of a low-resolution long-wavelength infrared detector, a digital temperature and humidity sensor, and a CMOS sensor. In view of the significant contrast between face and background in thermal infrared images, this paper explores a suitable accuracy-latency tradeoff for thermal face detection and proposes a tiny, lightweight detector named YOLO-Fastest-IR. Four YOLO-Fastest-IR models (IR0 to IR3) with different scales are designed based on YOLO-Fastest. To train and evaluate these lightweight models, a multi-user low-resolution thermal face database (RGBT-MLTF) was collected, and the four networks were trained. Experiments demonstrate that the lightweight convolutional neural network performs well in thermal infrared face detection tasks. The proposed algorithm outperforms existing face detection methods in both positioning accuracy and speed, making it more suitable for deployment on mobile platforms or embedded devices. After obtaining the region of interest (ROI) in the infrared (IR) image, the RGB camera is guided by the thermal infrared face detection results to achieve fine positioning of the RGB face. Experimental results show that YOLO-Fastest-IR achieves a frame rate of 92.9 FPS on a Raspberry Pi 4B and successfully detects 97.4% of faces in the RGBT-MLTF test set. Ultimately, an infrared temperature measurement system with low cost, strong robustness, and high real-time performance was integrated, achieving a temperature measurement accuracy of 0.3 ℃.

Key words: artificial intelligence, infrared face detection, ultra-lightweight network, infrared thermal camera, YOLO-Fastest-IR

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,且更适宜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 ℃。

Received date: 2024-10-30, **revised date**: 2024-12-15　　**收稿日期**:2024-10-30,**修回日期**:2024-12-15

Foundation items: Supported by the Fundamental Research Funds for the Central Universities (2024300443); the Natural Science Foundation of Jiangsu Province (BK20241224).

Biography: LI Xi-Cai (1989-), male, Dehong, doctor degree. Research area involves stereo display technology, computer vision, and micro aerial vehicle control. E-mail: lixicai@nju.edu.cn.

[*] **Corresponding author**: E-mail: yqwang@nju.edu.cn

关 键 词:人工智能;热红外人脸检测;超轻量级网络;热成像测温相机;YOLO-Fastest-IR
中图分类号:TP18 **文献标识码**:A

Introduction

Infrared thermal cameras (ITCs) have attracted widespread attention across various sectors due to their capabilities in large-scale rapid screening, automatic tracking, high-temperature zone alarms, and visible-light image fusion, enabling efficient tracking of individuals with elevated temperatures in crowds[1-2]. During the pandemic, ITCs were widely deployed for inspection and quarantine in crowded public spaces such as airports, nucleic acid testing sites, subway/train stations, and shopping centers. This approach not only reduces the risk of cross-infection but also prevents congestion caused by large-scale temperature screening. Additionally, ITCs are applicable in chemical heat source monitoring and real-time livestock body temperature tracking on farms[3-4].

Face detection is a key technology for ITCs. A high-speed, stable, low-cost, and robust face detection algorithm enables effective face detection under varying conditions and ensures accurate temperature measurement, significantly impacting ITC performance. Despite substantial progress in face detection over recent decades, infrared temperature measurement remains challenging. Although numerous models have been proposed for thermometers[5-7], accurately and quickly locating faces in infrared images is still a difficult task. Most existing methods rely solely on a single thermal infrared camera for rudimentary facial detection based on morphological processing[8] or perform facial localization using visible light images. In such approaches, the thermal camera first detects faces in visible-spectrum images and then maps the positions to infrared images for temperature measurement[9]. However, faces are difficult to detect directly in IR images, and the use of RGB cameras is limited by their susceptibility to ambient light interference[10]. Additionally, human-shaped objects (e. g. , narrow pillars or blurry traffic lights) often resemble faces[11] and may be misidentified by thermometers, leading to false alarms in ITCs and compromising their practical application. In general, RGB images alone cannot guarantee high-quality face detection, and more comprehensive information should be explored to improve thermometer reliability.

Most ITCs typically utilize high-resolution images as input to achieve high recall rates, which usually rely on costly graphics processing units (GPUs) to maintain low latency[12]. To our knowledge, few studies have previously reported on lightweight ITCs. Limited by infrared face detection technology and dataset availability, Negishi et al. employed a mature visible-light face detection algorithm to locate faces[13], then mapped the detected face coordinates to corresponding infrared images for temperature measurement. However, in addition to inheriting the limitations of visible-light face detection, this method suffers from inaccurate coordinate mapping, high computational overhead, and low frame rates.

Chaitra Hegde et al. implemented PoseNet-based forehead positioning for temperature measurement and cyanosis detection on a Raspberry Pi edge computing platform [5]. A significant drawback of this approach is that both forehead and lip detection required computation on a Google Coral USB accelerator. This system not only exhibits slow face detection speeds and poor positioning accuracy but also incurs high costs, as near-real-time performance can only be achieved with the assistance of the Google Coral's Tensor Processing Unit (TPU) neural network accelerator.

Currently, visible-light face detection tasks predominantly utilize the MS COCO dataset, while facial analysis tasks can employ 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, introduced in 2003, followed by commonly used datasets like IRIS[18] and NVIE[19]. In 2021, Domenick Poster et al. [20] proposed the ARL-VTF dataset, the most recent thermal infrared facial dataset, which also comprehensively cataloged previous datasets in this domain. Existing thermal infrared facial datasets primarily focus on tasks such as facial recognition and emotion recognition. Consequently, these datasets typically feature images where a single face occupies most of the frame, with few or no background interference factors, resulting in generally high image resolution. Most existing thermal infrared face detection algorithms based on convolutional neural networks utilize custom datasets of single-user thermal infrared facial images collected by the researchers themselves, and these datasets are closed source. To the best of our knowledge, there currently exists no publicly available thermal infrared face dataset specifically designed for multi-user face detection tasks.

In 2021, Woongkyu Lee et al. proposed a temperature measurement method based on the SSD model[21]. They customized SSD to identify human face locations through transfer learning, achieving a speed of 160 FPS on NVIDIA Jetson AGX while directly detecting faces in infrared images. However, a limitation of this method is its difficulty in accurately locating suspicious high-temperature targets in visible images when multiple targets are present. Friedrich et al. [22] developed an eye corner detection algorithm for thermal infrared face detection, leveraging the characteristic that eye regions typically show the highest temperature while facial areas show the lowest. Reese et al. [23] introduced a gray projection analysis (Projection Profile Analysis, PPA) method, where they calculated the gray projection curve of thermal infrared images and determined face regions by analyzing both the curve and its first derivative. 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 Vio-

la 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].

Experimental results demonstrate that conventional machine learning algorithms primarily designed for visible-light face detection achieve higher accuracy and lower false positive rates compared to those specifically developed for thermal infrared images. While thermal infrared-specific detection algorithms exhibit shorter running times, the PICO algorithm stands out with the highest computational efficiency.

In recent years, deep learning approaches have been increasingly applied to thermal infrared face detection tasks. In 2017, Alicja et al. modified the InceptionV3 architecture by removing the global pooling operation[29], This adaptation enabled separate classification on a 64-grid 8×8 feature map, with face regions determined by grids exhibiting face probabilities exceeding 0.5. However, their study was confined to single-user thermal infrared face detection. In 2019, Silva et al. [30] adapted the YOLOv3 network[31], training it on thermal infrared facial datasets and truncating the final prediction feature map during detection to achieve both high accuracy and efficiency. Although YOLO-based detectors inherently support multi-target detection, their application focused solely on driver detection in autonomous systems, consequently limiting their dataset to single-user scenarios. Most existing methods directly employ infrared images to train generic models for infrared target detection, resulting in models containing substantial redundant information that limits detection speed improvements. Our work demonstrates that combining infrared image characteristics with model compound scaling can significantly enhance model efficiency, paving the way for future advancements in this field.

This paper presents a high-speed, robust dual-band face detection system implemented on an ARM CPU for ITCs. Our ITC system integrates three key components: (1) a low-resolution infrared detector, (2) a CMOS sensor, and (3) an environmental temperature monitoring sensor. The system employs the infrared camera for initial face localization, while the visible-light camera provides supplementary verification of corresponding facial features in infrared images. We additionally incorporate thermal radiation attenuation compensation based on distance and implement stereo ranging through dataset fitting. To address these challenges, we propose an ultra-lightweight thermal infrared face detection network and investigate the impact of various network architectures on detection performance. For model training, we developed a comprehensive dual-band face detection dataset comprising 2,030 RGB-thermal image pairs with 138,389 annotated faces. We validated our approach through experimental prototypes deployed on Raspberry Pi systems with ARM CPUs. Through systematic compound scaling of network depth, input resolution, and channel width, we identified an optimal ultra-lightweight architecture specifically tailored for thermal infrared face detection applications.

1 Principle of the dual-band infrared thermal 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. The raw infrared data captured by the IR camera undergoes parallel processing through two distinct pathways. In the first processing branch, the original 16-bit data is dynamically normalized to 8-bit grayscale values based on the detected maximum and minimum temperatures, following the conversion rule specified in Eq. (1):

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

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. In the second processing branch, the raw data is preserved as backup for 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.

After obtaining the face region in the thermal infrared image, the region of interest is synchronously mapped to the corresponding visible light 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.

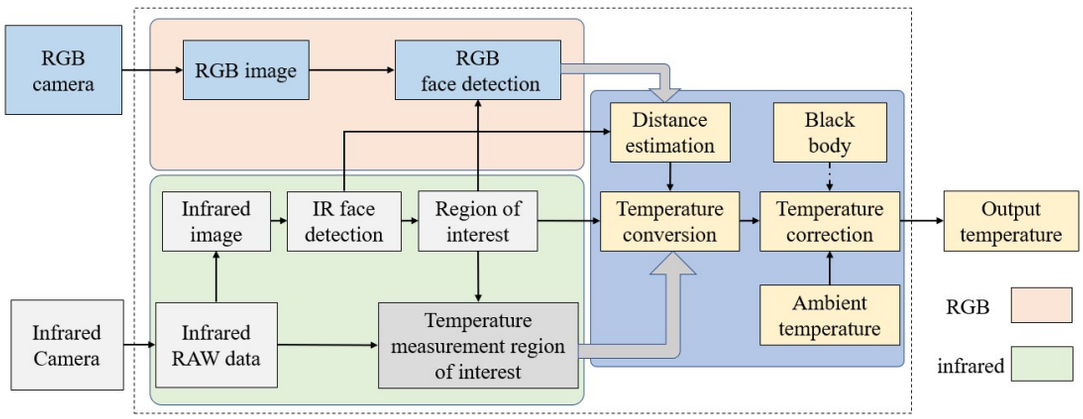


Fig. 1 The procedure of the working process of the dual band ITC system
图1 双波段红外测温系统组成和工作流程

2 Ultra-lightweight object detection network design and training

2.1 The ultra-lightweight object detection network YOLO-Fastest-IR

Since most of the body's thermal radiation is typically attenuated by clothing, the facial region consistently exhibits the highest grayscale intensity in thermal images, creating strong contrast against the background. In infrared (IR) imaging, faces manifest as bright oval patterns with indistinguishable facial features. This phenomenon presents two major challenges: it not only complicates high-precision computer vision tasks such as facial recognition, emotion analysis, and landmark detection, but also significantly restricts the extractable feature space for convolutional neural networks (CNNs), leading to limited discriminative deep features for thermal facial detection. Notably, facial patterns demonstrate distinct shape and aspect ratio characteristics compared to common thermal interference sources (e. g. , electronic displays, fluorescent lamps, and heated containers). Therefore, theoretically, lightweight convolutional neural networks can stably detect faces in IR images.

As illustrated in Fig. 2, we validated the aforementioned hypothesis by developing four compact convolutional neural networks of varying complexity levels, adopting architectural principles from YOLO-Fastest[32]. The red and blue blocks in the schematic diagram denote the lite convolution module and lite residual module, respectively. This study systematically investigates the impact of network scale through three fundamental dimensions: input resolution, network depth, and channel width.

Regarding resolution optimization, contemporary object detection networks typically employ high input resolutions (416-800 pixels) to accommodate datasets like

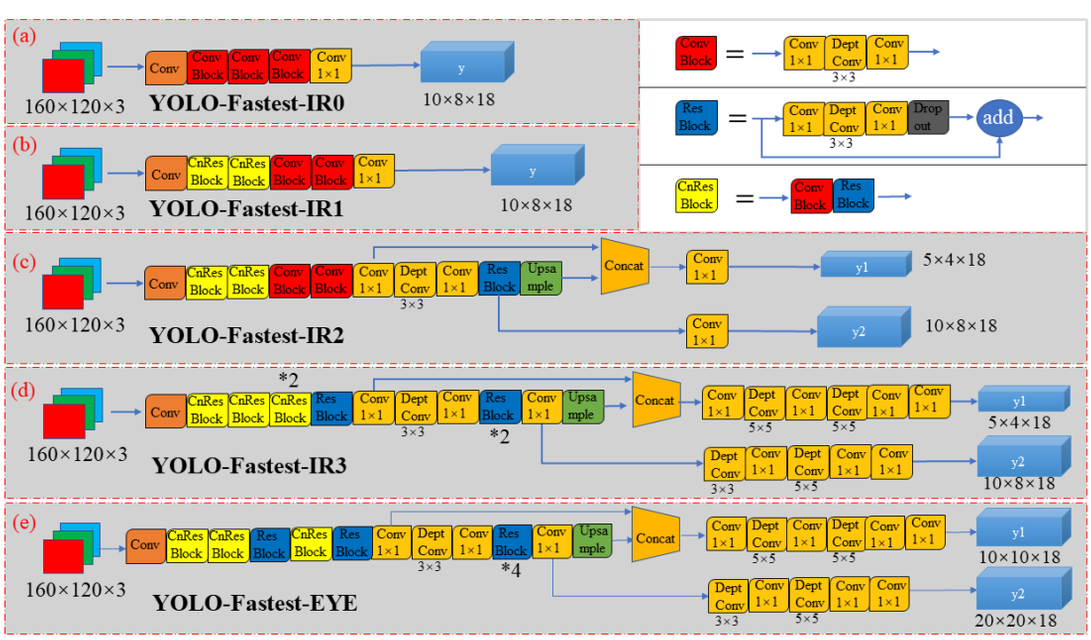


Fig. 2 The YOLO-Fastest-IR network structures with four different levels of complexity
图2 四种不同复杂度的热红外人脸检测网络 YOLO-Fastest-IR 结构

COCO. For lightweight detection networks, resolution reduction has become a prevalent strategy for computational efficiency enhancement. However, direct downsampling to 320×240 pixels inevitably compromises image details. Notably, our infrared camera's native resolution is limited to 160×120 pixels, making bilinear interpolation upscaling to 320×240 potentially counterproductive. To maintain compatibility with the infrared sensor's physical constraints, we configured YOLO-Fastest with a native 160×120 input resolution while controlling variables in depth and width dimensions. This approach yielded four distinct infrared face detection architectures, collectively designated as YOLO-Fastest-IR variants, each featuring optimized depth-width combinations

All four network variants employ an initial 3×3 convolutional layer with stride 2 and zero-padding, reducing the input spatial dimensions by half. Throughout the networks, lite residual modules maintain identical input-output dimensions without spatial reduction. The first lite convolution module in IR1-IR3 preserves feature map dimensions, while in IR0, all lite convolution modules utilize depthwise separable convolutions (stride=2 with zero-padding) for progressive halving of spatial resolution. Both IR0 and IR1 undergo four spatial reduction stages, yielding final feature maps of 10×8 grid resolution. The architectures demonstrate progressive channel expansion with increasing network depth, as detailed below:

(1) YOLO-Fastest-IR0: As shown in Fig. 2(a), YOLO-Fastest employs only three lite convolutional modules, excluding its first and last two convolutional layers. The final output grid size is 10×8, making it the network with the fewest layers and the simplest structure. Due to its shallow architecture, gradient vanishing and network degradation are virtually absent, eliminating the need for residual modules.

(2) YOLO-Fastest-IR1: As shown in Fig. 2(b), a residual module was incorporated into the backbone network, which consists of four lite convolution modules and two lite residual modules. The final output grid size is 10×8. (3) YOLO-Fastest-IR2: As shown in Fig. 2(c), a multi-scale prediction strategy was introduced in the neck network, which employs five lite convolutional modules and three lite residual modules. The final output consists of two feature maps with grid sizes of 5×4 and 10×8, responsible for detecting large and small targets, respectively. The head network uses only a single convolutional layer.

(4) YOLO-Fastest-IR3: As shown in Fig. 2(d), the network layers were further deepened, with six convolutional layers employed in the head network. The architecture ultimately outputs two feature maps with grid sizes of 5×4 and 10×8, making it the network configuration with the greatest depth and most complex structure among the compared versions.

(5) YOLO-Fastest-EYE: The overall network structure is shown in Fig. 2(e). The facial bounding box maintains an approximately 1:1 aspect ratio. For RGB images with 640×480 resolution, when the user's face is approximately 1 meter from the camera, the bounding box size measures about 160×160 pixels. Accordingly, YOLO-Fastest-EYE's input image size is set to 160×160. Following four downsampling operations and one upsampling of the feature map, the network produces a final output tensor of size 20×20×18. Each grid cell corresponds to 5% of the feature map's length and width. As the network only predicts specific facial features (such as eyes), the output feature map is limited to 18 channels.

2.2 Training of the ultra-lightweight object detection network YOLO-Fastest-IR

To facilitate comprehensive training of the proposed networks, we developed the RGBT Multi-user Low-resolution Thermal Face (RGBT-MLTF) dataset. The dataset comprises 26, 800 thermal images captured using a LeptON3.0 infrared camera (160×120 native resolution), with each image containing 1-4 facial instances. Notably, 76% of the images contain ≥2 faces, ensuring adequate multi-face representation. 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 2 680 images, 1 627 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 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.

3 Experiment and discussion

3.1 Coordinate mapping relationship between infrared and RGB binocular cameras

The infrared sensor employed in this study is FLIR's Lepton 3.5 module with a resolution of 160×120 pixels, while the visible-light module utilizes a Raspberry Pi camera equipped with an OV5647 sensor (OmniVision Technologies) offering 1280×720 resolution. Given the significant disparity in both spatial resolution and physical alignment between the infrared and RGB imaging systems, direct mapping of thermal facial bounding box coordinates to visible-light images is infeasible. To address this, we established a spatial correspondence model between the two imaging modalities through regression analysis of the annotated RGBT-MLTF dataset. Specifically, the coordinate transformation between infrared-detected facial regions and their RGB counterparts is computed using Eq. (2).

$$X_{\mathrm{RGB}} = 0.881\,2 \cdot X_{\mathrm{IR}} + 0.084\,4 . \tag{2}$$

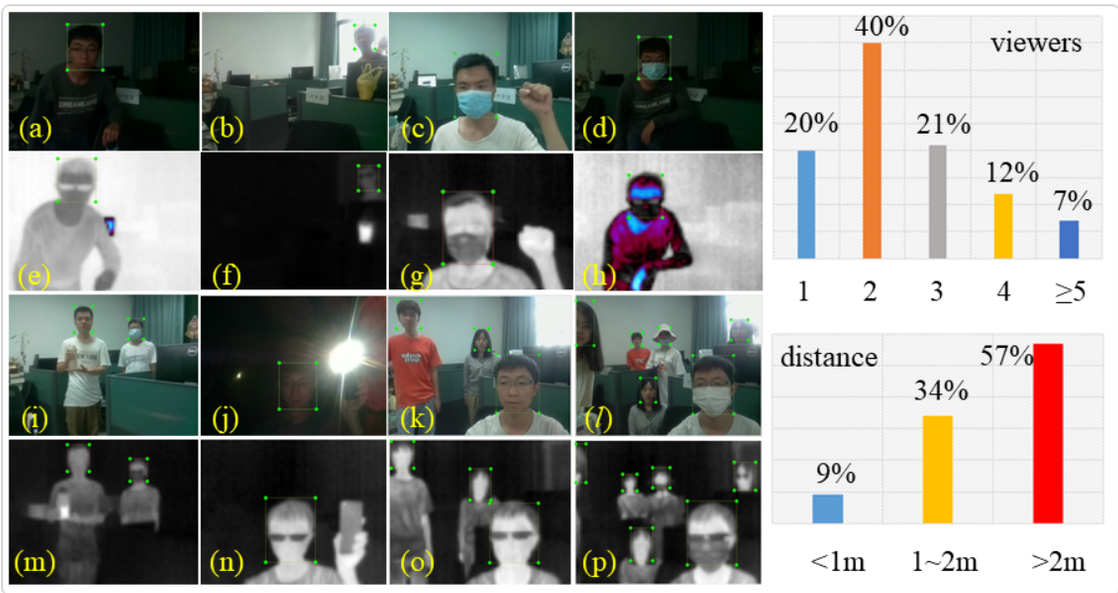


Fig. 3 Examples and statistics of the RGBT-MLTF dataset
图3 部分RGBT-MLTF数据集示例和统计信息

where X_{IR} denotes the normalized x-coordinate (relative to image width) of the original infrared bounding box, and X_{RGB} represents the corresponding normalized x-coordinate in the transformed RGB image. The horizontal and vertical axes correspond to the IR and RGB coordinate systems, respectively. Each sampling point corresponds to a matched pair of face bounding boxes in the dual-spectrum images. Through linear regression analysis, the transformation relationship as expressed in Eq. (2).

3.2 Experiment of the thermal infrared face detection

In this study, we employ the RGBT-MLTF dataset to train four proposed thermal face detection networks (YOLO-Fastest-IR series) along with three benchmark models: YOLO-V4 and YOLO-V8s [33], YOLO-Fastest. We then compare the detection performance of our proposed networks with these state-of-the-art object detection algorithms. As demonstrated in the first column of Fig. 4(a1-g1), our four thermal infrared face detection networks maintain stable face detection capability even in challenging scenes with interference. However, YOLO-Fastest-IR0 and YOLO-Fastest-IR1 exhibit some false positives, occasionally misidentifying monitors or raised fists in the background as faces. The second column in Fig. 4(a2-g2) reveals that in multi-person scenarios, our four scaled networks successfully detect thermal faces across various sizes, including partially occluded faces. As shown in the third column in Fig. 4(a3-g3), all networks except the shallowest YOLO-Fastest-IR1 and YOLO-Fastest-IR0 achieve reliable face detection in thermal images. These two smallest networks still show occasional false positives, while the other five models demonstrate robust performance.

To further evaluate the generalization capability of our four YOLO-Fastest-IR models, we performed validation tests using a high-resolution thermal infrared facial dataset from Marcin et al. [34]. The results in Fig. 4(a4-g4) demonstrate varying generalization performance across detectors: while YOLO-Fastest-IR0 and YOLO-Fastest-IR1 exhibit bounding box deviations from ground truth, YOLO-Fastest-IR2, YOLO-Fastest-IR3, and YOLOv8s achieve accurate face localization (near 100% confidence) on thermal images significantly different from the training set. Notably, YOLO-V4 fails to produce valid predictions at standard confidence thresholds (0.5), only generating detections when the threshold is lowered to 0.1 as shown in Fig. 4(f4), suggesting potential overfitting to the training data. Additionally, as shown in Fig. 4(a5-g5), although the dataset contains only a limited number of pseudo-color samples, the IR face detection networks YOLO-Fastest-IR0 and YOLO-Fastest-IR1 proposed in this study can still accurately localize faces in pseudo-color images. In contrast, YOLO-V8s fails to reliably detect faces in such pseudo-color images.

Figure 5(a) presents the experimental results of our proposed thermal infrared-guided RGB face detection and eye localization method. The system first employs YOLO-Fastest-IR to identify facial regions in infrared images, then utilizes YOLO-Fastest-EYE to precisely locate eyes within these thermally determined regions of interest. Our results demonstrate that this infrared-guided approach effectively addresses the challenge of occluded face detection while preventing abnormal temperature measurements by the ITC, thereby significantly reducing false alarm rates. As shown in Fig. 5(e-h), YOLO-Fastest-EYE maintains reliable face and eye detection performance even under poor lighting conditions. Furthermore, Fig. 5(i-l) illustrates the system's capability in multi-target face detection scenarios. Collectively, these results confirm that our algorithm exhibits strong robust-

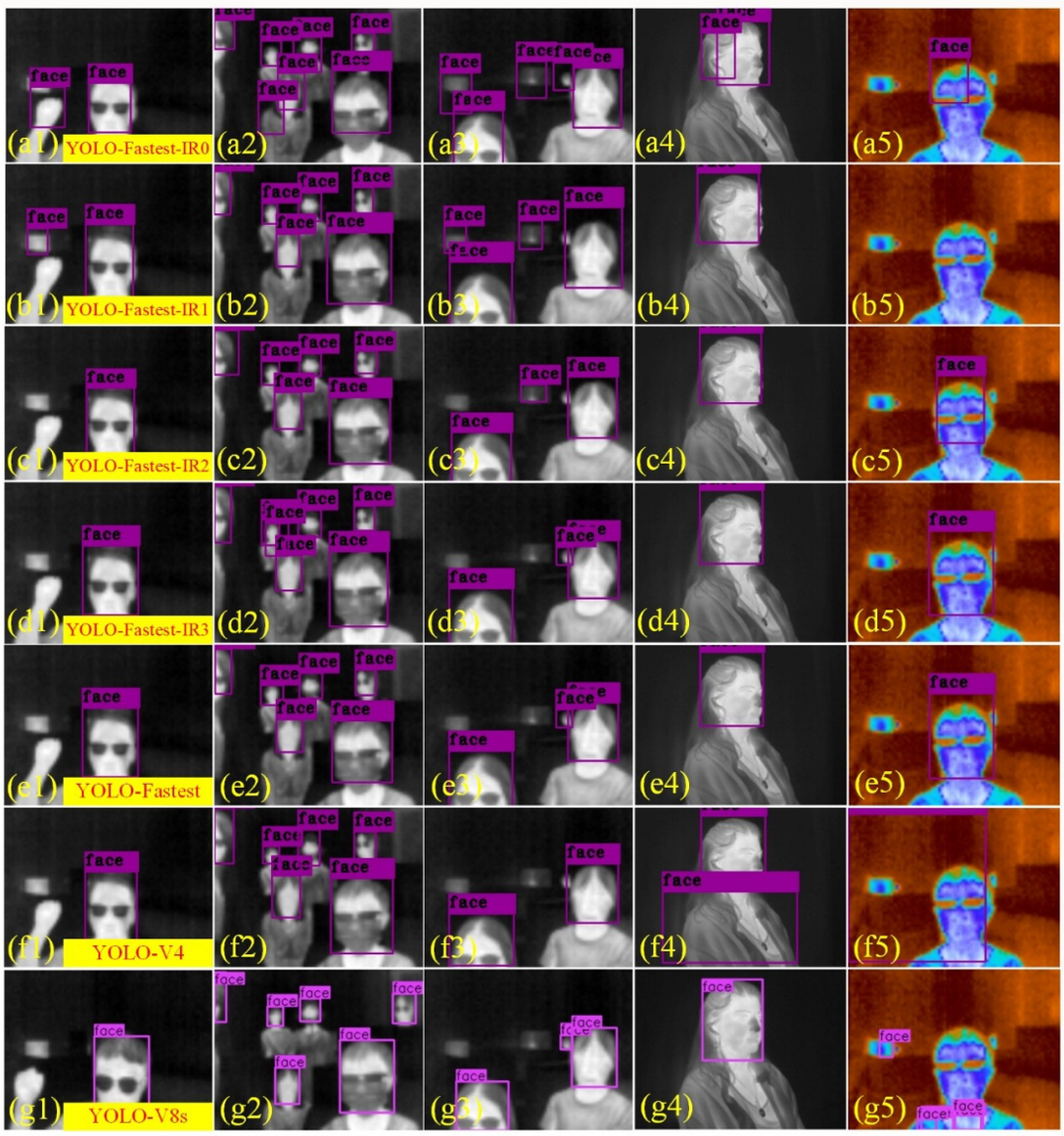


Fig. 4 The comparative and generalization experiments of different models: (a1~a5) the detection effect of YOLO-Fast-IR0; (b1~b5) the detection effect of YOLO-Fast-IR1; (c1~c5) the detection effect of YOLO-Fast-IR2; (d1~d5) the detection effect of YOLO-Fast-IR3; (e1~e5) the detection effect of YOLO-Fast; (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:(a1~a5) YOLO-Fast-IR0的检测效果;(b1~b5) YOLO-Fast-IR1的检测效果;(c1~c5) YOLO-Fast-IR2的检测效果;(d1~d5) YOLO-Fast-IR3的检测效果;(e1~e5) YOLO-Fast的检测效果;(f1~f5) YOLO-V4的检测效果;(g1~g5) YOLO-V8s的检测效果

ness across varying lighting conditions and occlusion challenges while supporting efficient multi-target detection.

Fig. 6 presents the comparative performance metrics (AP values and FPS) of various network architectures. Our experimental results demonstrate that all four YOLO-Fast-IR variants satisfy real-time detection requirements while maintaining AP50 values above 90%, outperforming YOLO-V4, YOLO-V8s, and YOLO-Fastest in terms of efficiency. Notably, these lightweight networks achieve detection accuracy comparable to YOLO-V4. All AP values were evaluated on the RGBT-MLTF dataset, with frame rates measured on Raspberry Pi 4B CPU hardware. The study reveals that network compression significantly improves inference speed without substantial accuracy degradation, confirming that structurally simple lightweight convolutional neural networks are sufficiently capable of extracting infrared facial features and achieving reliable face localization. As shown in Fig. 6, YOLO-Fastest-IR networks exhibit a clear performance trade-off: deeper architectures yield higher precision at the cost of reduced inference speed. An interesting observation is that YOLO-Fastest-IR2 achieves superior mean precision compared to deeper networks, potentially attributable to favorable convergence conditions during training.

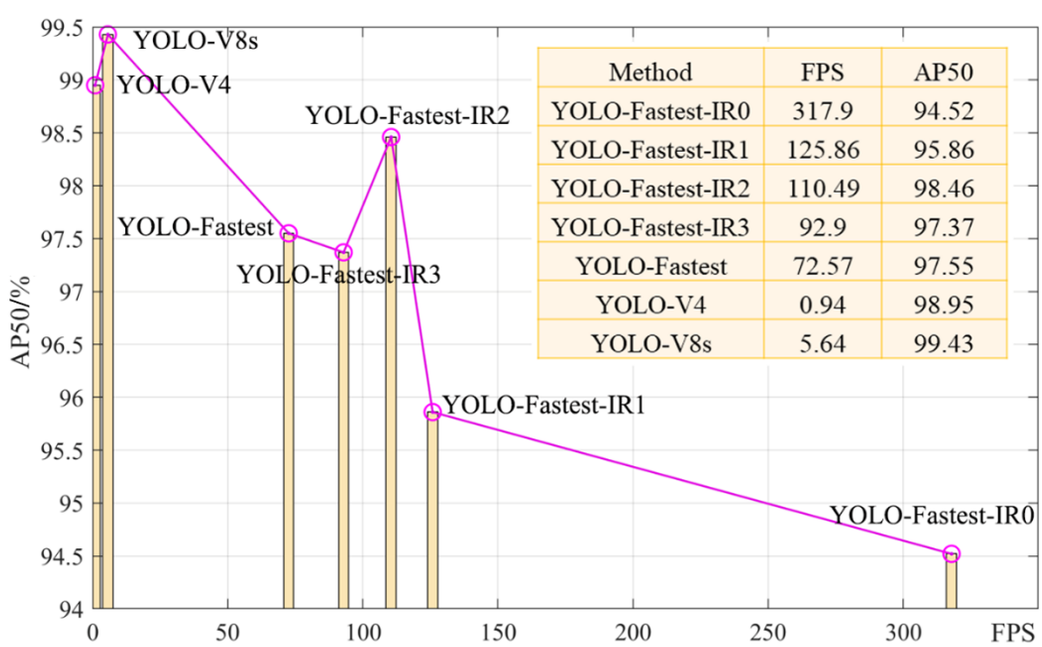
In object detection tasks, 30 frames per second (FPS) is conventionally established as the threshold dis-

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 are robust against variations in face angle, inter-viewer occlusion, and environmental occlusion; (e~h) extreme lighting conditions; (i~*l*) multi-target and distant viewer detection
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:(a-d) YOLO-Fastest-IR 与 YOLO-Fastest-Eye 对脸部角度变化、观察者间遮挡及环境遮挡具有鲁棒性;(e-h) 极端光照条件;(i-l) 多目标及远距离观察者检测

tinguishing real-time from non-real-time performance. While YOLO-V4 and YOLO-V8s achieve real-time operation on GPU platforms, their inference speed drops significantly on Raspberry Pi 4B hardware, requiring over 1 second per frame - far below real-time requirements. Notably, YOLO-V4 attains 98.95% AP50 on our RGBT-MLTF dataset, substantially outperforming its 81.3% AP50 performance on the MS COCO benchmark. This performance gap highlights the unique challenges of our dataset, which intentionally incorporates diverse interference factors affecting thermal infrared face detection. In fact, the RGBT-MLTF dataset presents greater detection difficulty than images captured in actual ITC system applications. The comparative performance analysis of these networks demonstrates that deep learning approaches are particularly well-suited for addressing thermal infrared face detection challenges in practical ITC deployment scenarios.

3.3 Temperature measurement experiments

This section details our temperature measurement



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

methodology. We established calibration curves by plotting blackbody reference temperatures (vertical axis) against corresponding infrared camera raw data (horizontal axis), with each temperature point averaged over 20 measurements. As shown in Fig. 7, the experimental data (blue curve) and its linear fit (magenta dashed line) demonstrate the fundamental temperature-radiation relationship. To investigate thermal radiation's distance dependence, we conducted systematic measurements at 25 cm intervals from 25-225 cm, with 20 trials per distance averaged for reliability. The resulting distance-dependent characteristics are plotted as the red curve in Fig. 7, with its linear approximation shown as the green dotted line. These relationships are mathematically expressed in Eq. (3), where i=1 corresponds to the temperature-gray-level correlation (a_1=19.645, b_1=0.1163) and i=2 represents the distance dependence (a_2 = 37.514, b_2=-0.00794).

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

Based on Planck's radiation law, the grayscale value of each infrared image pixel exhibits a direct proportionality to the thermal radiation energy at the corresponding point on the target surface. However, thermal imagers measure the radiation temperature (T_r) rather than the true object temperature (T_0), where the latter represents the equivalent blackbody temperature emitting identical radiative energy. Consequently, accurate temperature measurement requires calibration using high-precision blackbody sources to establish the precise mapping relationship between preset blackbody temperatures and sensor output voltages. The fundamental relationship between radiation temperature (T_r) and true temperature (T_0) is mathematically expressed in Eq. (4).

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

The wavelength-dependent parameter λ varies according to the infrared detector material characteristics: λ = 8.68 for InSb detectors (3-5 μm spectral range), λ = 5.33 for HgCdTe detectors (6-9 μm range), and λ = 4.09 for our implemented HgCdTe detector (8-14 μm range). Since atmospheric transmittance (t_a) effects are negligible in close-range thermometry applications, we assume t_a=1, yielding the gray-body surface temperature calculation formula in Eq. (5). This governing equation incorporates three key parameters: ε represents skin emissivity (typically 0.98 for human tissue), T denotes the radiation temperature detected by the infrared sensor, and T_u signifies ambient temperature measured using auxiliary temperature/humidity sensors. Through this formulation, we can accurately estimate forehead temperature in clinical measurement scenarios.

$$T_M = \frac{1}{\varepsilon}\cdot(T_r^{\lambda} - (1-\varepsilon)\cdot T_u^{\lambda})^{\frac{1}{\lambda}}. \tag{5}$$

To validate the proposed method's effectiveness, we conducted comprehensive temperature measurement experiments using our binocular vision system. Experimental results in Fig. 8 demonstrate that our algorithm maintains stable face detection and accurate forehead temperature measurement across varying distances (both close and long range) and different scenarios (single or multiple subjects). Furthermore, as evidenced in Figs. 5 and 8, the infrared face detection algorithm exhibits strong robustness, successfully handling challenging cases including masked faces, partial occlusions, and diverse facial poses. Notably, Fig. 8(e) illustrates a false high-temperature warning scenario where a cup's surface intrudes into the facial region during drinking. To mitigate such occurrences in practical implementations, we employ precise eye-position-based localization to strictly define the valid temperature measurement area.

To validate the temperature measurement accuracy of our ITC system, we conducted rigorous testing under varying ambient temperatures and distances using a blackbody reference source. Each measurement point was averaged over five trials to ensure reliability. As shown in Fig. 9(a), the measured temperature exhibits a linear decrease with increasing distance, and higher ambient temperatures consistently yield elevated mea-

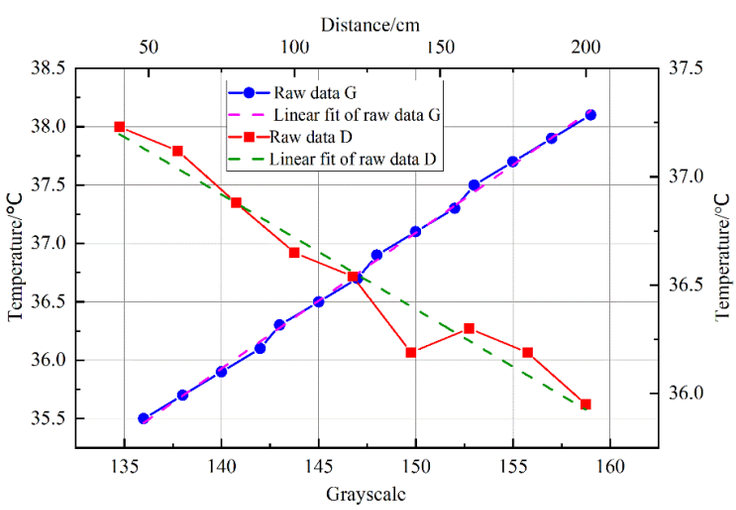


Fig. 7 The variation of grayscale values with temperature at different distances
图7 不同距离下灰度值随温度的变化关系

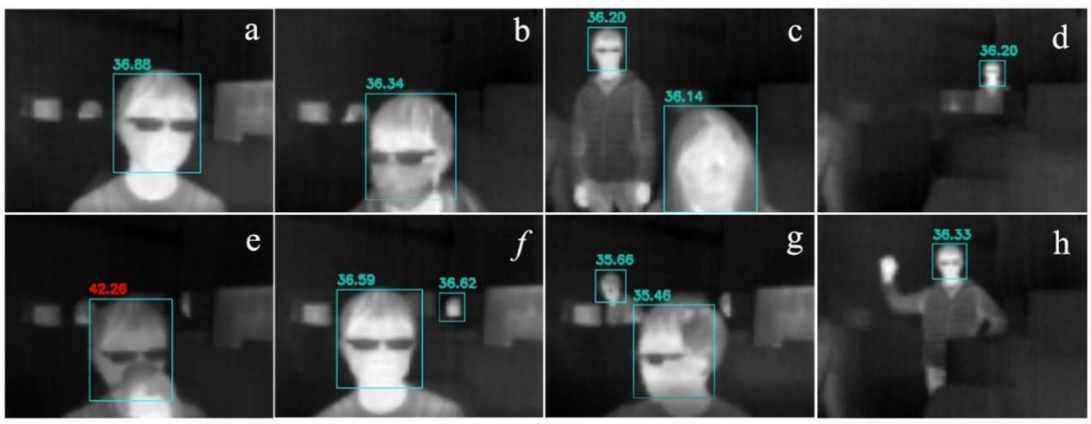


Fig 8 The real time temperature measurement experiment of temperature measurement system: (a) normal sitting and standing; (b) wear a mask; (c) interference testing at different distances; (d) remote temperature measurement experiment; (e) high temperature warning test; (*f*) remote multi-target temperature 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:(a) 正常坐立状态;(b) 佩戴口罩;(c) 不同距离干扰;(d) 远程测温;(e) 高温报警;(f) 远程多目标测温;(g) 侧脸;(h) 握拳干扰

surement values. In complementary human subject testing, we performed comparative evaluations against clinical forehead thermometers across 11 distinct measurement distances. The experimental results in Fig. 9(b) demonstrate that while uncorrected infrared measurements (blue curve) show distance-dependent attenuation, our distance-compensated algorithm (green curve) maintains consistent accuracy across all tested ranges. Notably, at close ranges (<1m), the ITC achieves comparable accuracy to clinical forehead thermometers. Beyond this range, our system demonstrates superior performance with an accuracy of ±0.3 ℃, while simultaneously maintaining excellent measurement repeatability and stability.

4 Conclusions

In this paper, we develop a dual-band infrared temperature measurement device (ITC) capable of measuring forehead temperature. The device integrates an infrared detector, an RGB sensor, and a humidity sensor for environmental monitoring. Furthermore, we propose four ultra-lightweight thermal infrared face detectors at different scales, designated as YOLO-Fastest-IR0 through YOLO-Fastest-IR3. Experimental results on our newly proposed RGBT-MLTF dataset demonstrate that the YOLO-Fastest-IR series outperforms existing algorithms (including YOLOv4 and YOLO-Fastest) in mobile and edge computing deployment scenarios. Specifically, while the tiny version achieves the fastest inference speed and maintains acceptable accuracy for infrared thermometry applications (albeit with a slight reduction in face localization precision compared to larger models), YOLO-Fastest-IR0 with its minimal network architecture shows limited detection capability due to insufficient network depth. Notably, the other variants demonstrate robust performance, achieving >95% average accuracy with > 90 FPS on Raspberry Pi 4B hardware. Comparative studies against YOLO-Fastest, YOLOv4, and YOLOv8s reveal that our proposed architectures significantly improve computational efficiency with minimal accuracy degradation. The network structure can be adaptively adjusted based on specific vision tasks to achieve optimal preci-

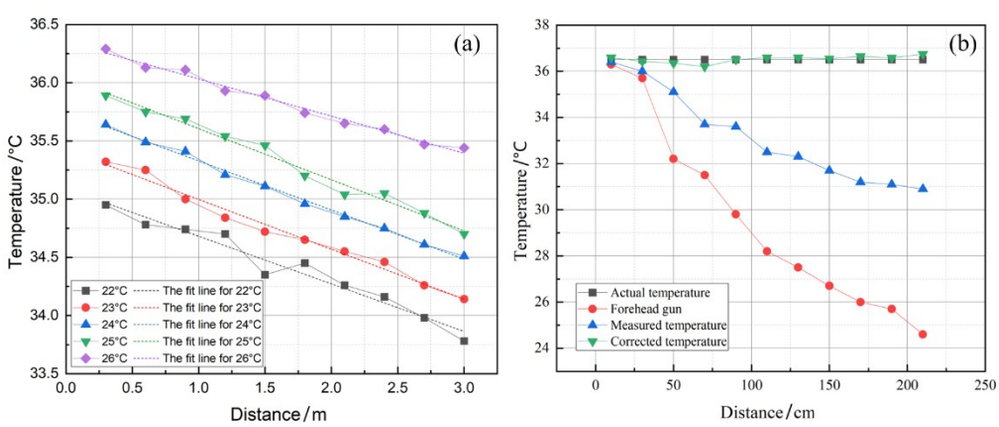


Fig. 9 Analysis of temperature measurement accuracy of ITC: (a) the variation relationship of different temperatures of blackbody under different environmental temperatures and distances; (b) temperature correction experiment
图 9 红外测温系统的测温精度分析:(a) 不同环境温度及距离下黑体各温度点的变化关系;(a) 温度修正实验

sion-speed trade-offs. Experimental results confirm both the effectiveness of our ITC device and the superior performance of the proposed face detection framework.

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). Cluj-Napoca: IEEE, 2018: 1-6.

[2] Ring E F J, Jung A, Zuber J, et al. Detecting fever in polish children by infrared thermography[C]. 9th International Conference on Quantitative InfraRed Thermography (QIRT). Krakow: QIRT Council, 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). Bandung: IOP Publishing, 2021: 1096.

[4] Ye X, Gao S, Li F. ACE-STDN: an infrared small target detection network with adaptive contrast enhancement[J]. Journal of Infrared and Millimeter 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[EB/OL]. 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. San Jose: IEEE, 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). Singapore: IEEE, 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]. 5th International Conference on Machine Vision and Applications (ICMVA). Prague: ICMVA, 2022: 16-21.

[12] Miao Z, Zhang Y, Li W. Real-time infrared target detection based on center points[J]. Journal of Infrared and Millimeter 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). Honolulu: IEEE, 2018: 4371-4374.

[14] HELEN Dataset[EB/OL]. http://www.ifp.illinois.edu/~vuongle2/helen/

[15] IBUG Dataset[EB/OL]. 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]. IEEE International Conference on Computer Vision Workshops (ICCVW). Sydney: IEEE, 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[EB/OL]. https://archive.ics.uci.edu/ml/datasets/Iris/

[19] NVIE Dataset[EB/OL]. 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]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2021, 3(2): 1-12.

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

[22] Friedrich G, Yeshurun Y. Seeing people in the dark: face recognition in infrared images[M]. Biometric Authentication. Berlin: Springer, 2002: 41-50.

[23] Reese K W, Zheng Y, Elmaghraby A. A comparison of face detection algorithms in visible and thermal spectrums[C]. International Conference on Advances in Computer Science and Application (CSA). Cairo: CSA, 2012: 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]. International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS). Antwerp: Springer, 2017: 518-529.

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

[26] Wang X, Chen J, Wang P, et al. Infrared human face auto locating based on SVM and a smart thermal biometrics system[C]//Sixth International Conference on Intelligent Systems Design and Applications. Jinan: IEEE, 2006.

[27] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego: IEEE, 2005.

[28] Marković N, Furht B, Pokrić M, et al. Object detection with pixel intensity comparisons organized in decision trees[J]. Pattern Recognition, 2014, 47(9): 2936-2946.

[29] Kwasniewska A, RumińSki J, Rad P. Deep features class activation map for thermal face detection and tracking[C]. 10th International Conference on Human System Interactions (HIS). Ustron: IEEE, 2017: 41-47.

[30] Santos G, Margal R, Ferreira A, et al. Face detection in thermal images with YOLOv3[C]. International Symposium on Visual Computing (ISVC). Lake Tahoe: Springer, 2019: 89-99.

[31] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[32] Yolo-Fastest[EB/OL]. https://github.com/dog-qiuqiu/Yolo-Fastest

[33] YOLOv8[EB/OL]. https://github.com/ultralytics/ultralytics

[34] Kopaczka M, Kolk R, Schock J, et al. A thermal infrared face database with facial landmarks and emotion labels[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 68(5): 1-6.