Occupant Attributes Recognition for Thermal Comfort in Passenger Car Cabin

Mingjie Liu
Chongqing University of Posts and Telecommunications
Chongqing, China
liumj@cqupt.edu.cn

Zhengyan He
Chongqing University of Posts and Telecommunications
Chongqing, China
s220331040@stu.cqupt.edu.cn

Fan Ren
Chang'an Automobile Intelligence Research Institute
Chongqing, China
renfan@changan.com.cn

Ping Liu
Chongqing University of Posts and Telecommunications
Chongqing, China
liuping_cqupt@cqupt.edu.cn

Jianbin Chen
Chang'an Automobile Intelligence Research Institute
Chongqing, China
chenjb3@changan.com.cn

Changhao Piao
Chongqing University of Posts and Telecommunications
Chongqing, China
piaoch@cqupt.edu.cn

***Abstract*—** Occupant attribute recognition is a key issue in the evaluation of thermal comfort in passenger car cabins. Different gender, ages, and dress make passengers have different perceptions of the temperature in the car. We developed an occupant attributes recognition system that utilizes the image taken in the cabin to recognize the gender, age, and clothing type of occupants. The system incorporates open-source object detection models and a multi-task learning network, which were trained on various publicly available datasets. To address the issue of low recognition accuracy caused by the lack of in-cabin scene images in existing public clothing datasets, we collected 31 images of occupants wearing different clothes. Through the analysis of experimental results, the proposed passenger attribute recognition system demonstrates high accuracy and fast recognition speed in real-world driving scenarios.

Keywords—Occupant attributes, Passenger car cabin, Object detection, Multi-task learning

I. INTRODUCTION

Ensuring thermal comfort within the passenger cabin is a critical criterion for evaluating the comfort of passenger cars, and it poses significant challenges [1]. The evaluation of thermal comfort is generally influenced by various factors, including human body characteristics (age, gender, weight, etc.), clothing, and environmental conditions [2]. Under the same room temperature conditions, females and older individuals tend to feel colder or hotter compared to males and younger people [3][4]. In other words, females and older individuals are more sensitive to temperature. People will choose to wear more or fewer clothes to protect themselves from the cold or heat. Therefore, we can utilize the temperature inside the vehicle and the occupant's age, gender, and clothing type to assess whether the current in-vehicle environment might cause discomfort for the occupant.

In recent years, numerous researchers have presented diverse techniques for facial attribute recognition, exhibiting commendable performance in age and gender recognition [5]. Face attribute recognition has evolved from single-attribute recognition to multi-attribute recognition. By adopting a multi-task learning approach incorporating parameter sharing, features are refined based on different tasks, improving accuracy [6][7]. Within the realm of clothing recognition, traditional machine learning methods like Support Vector Machine (SVM), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) possess inherent limitations. Conversely, convolutional neural network (CNN)-based approaches, including YOLO [8], SSD [9], and R-CNN [10], have gained widespread popularity in clothing recognition [11]. However, when utilizing publicly available clothing recognition datasets such as DeepFashion2 [12] and ACWS [13] to train models, the recognition performance falls short due to the specific perspectives of the car cabin, occlusions, and illumination. To address this issue, we have collected images of occupants to enhance the accuracy of clothing recognition within this specific environment.

In this paper, we present a novel approach for assessing thermal comfort in car cabins by proposing an occupant attributes recognition system. To enhance the accuracy of clothing recognition within the complex environment of car cabins, we collected a diverse set of clothing images from 31 individuals and used them as training data. Furthermore, we applied data augmentation techniques to augment the captured images, thereby increasing the size of our dataset. The

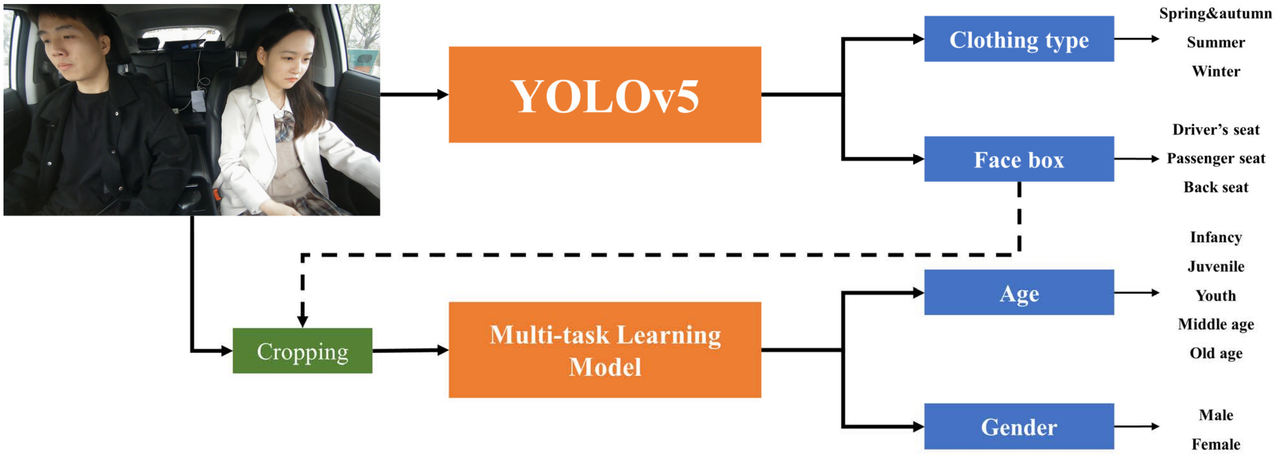

Fig. 1: Proposed occupant attribute recognition system architecture.

proposed system in this study leverages YOLOv5 and the multi-task learning model to develop an effective recognition system. Publicly available datasets were employed for the face attribute recognition component, and the dataset we collected was combined with publicly available clothing datasets and jointly utilized for training the clothing recognition model. Through comprehensive evaluations, our system demonstrates exceptional performance in recognizing key attributes such as gender, age, clothing type, and seating position (front or back seat) of the car occupants. Additionally, it exhibits robustness in accommodating various passenger postures.

The rest of the paper is organized as follows: Section II introduces our proposed occupant attribute recognition system for assessing thermal comfort in passenger cars. Section III discusses the performance of the system in the dataset and reality. The paper is concluded in Section IV with future research directions.

II. Occupant Attributes Recognition System

The occupant attribute recognition system proposed in this paper comprises two main components: facial attribute recognition and clothing attribute recognition. Facial attribute recognition focuses on recognizing the gender and age of the occupant. The overall system architecture as shown in Fig. 1. Initially, face detection and clothing recognition are performed using YOLOv5. Subsequently, a multi-task learning network is employed to recognize the occupant's gender and age based on face detection results. Moreover, we utilized the position of the face to determine the seating location of the occupant. The system's final output includes information such as the occupant's gender, age, clothing type, and location. To cater to the requirements of clothing recognition for different seasons, we have collected a comprehensive clothing dataset encompassing diverse clothing types worn by occupants in the car cabin.

A. Dataset

For face attribute recognition, we curated a collection of images representing various age groups from publicly available Asian face datasets such as Asian Face Age Dataset (AFAD) [14], UTK-Face [15], and MegaAge [16]. We categorized the age classification into seven levels: infancy (0-6), juvenile (7-14), youth (15-25), youth (26-35), youth (36-45), middle age (46-65), and old age (66 and above). Fig. 2 provides an overview of the data distribution after the final selection, comprising a total of 19,500 images.

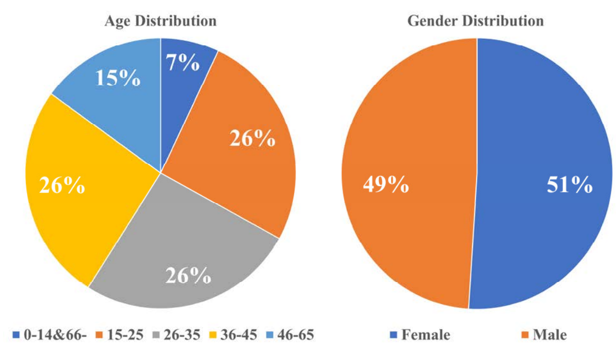

Fig. 2: Face attribute data distributions.

TABLE I. Clothing Types Distribution

Clothing Type	Number of Image Samples
Spring & autumn	7697
Summer	1345
Winter	236

To collect facial and clothing recognition data, a camera was installed beneath the car's rearview mirror to capture images of both the primary and secondary drivers. A total of 15 males and 16 females participated in the collection of the

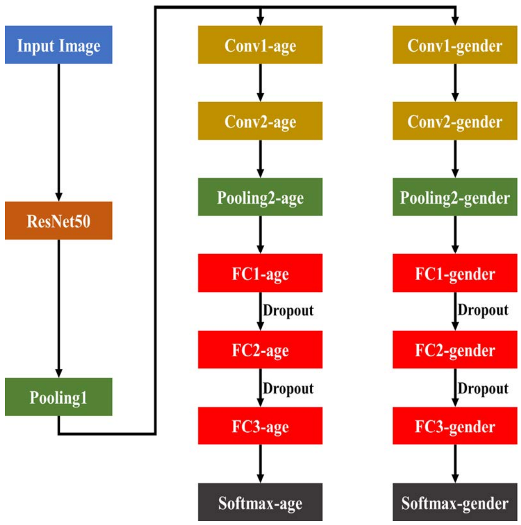

Fig. 3: Schematic diagram of multi-task learning model

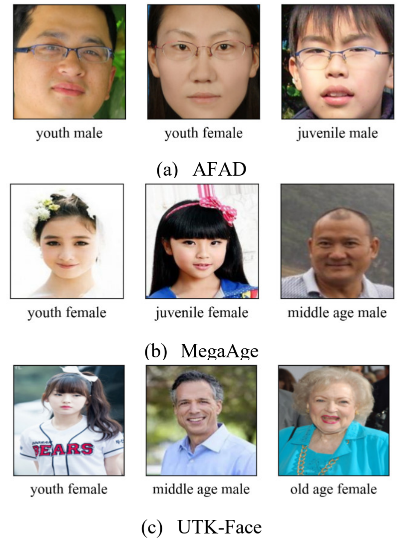

Fig. 4: Recognition results of the dataset.

dataset, each wearing a variety of clothing types to cover all seasons. The top left corner of Fig. 1 shows the image we collected, providing a clear view of the front and back seats. The collected images underwent data augmentation techniques and were labeled with face and clothing types. Table I shows the distribution of clothing types in the clothing dataset.

B. Network model

YOLOv5 model and multi-task learning model were trained using the provided datasets. The multi-task learning model utilized the cross-entropy loss as its loss function. To reduce overfitting and underfitting problems, the performance of the network can be improved by hyperparameter tuning [17]. In the proposed system, YOLOv5 is used to detect the coordinates of the face box in the image and recognize the type of clothing worn by occupants. The facial images were cropped from the images captured by the camera based on the coordinates of the face box. These cropped images were then used as input for the multi-task learning model to recognize the gender and age of occupants. Additionally, the coordinates of the face box are utilized to determine the seating position of occupants.

TABLE II. MULTI-TASK LEARNING MODEL PARAMETERS

Layer Name	Layer Type	Output Dimension
ResNet50	ResNet	2048*7*7
Pooling1	Avg-Pooling	2048*7*7
Conv1-age/gender	Convolution	1024*7*7
Conv2-age/gender	Convolution	1024*5*5
Pooling2-age/gender	Avg-Pooling	1024*1*1
FC1-age/gender	Fully Connection	1*1024
FC2-age/gender	Fully Connection	1*512
FC3-age	Fully Connection	1*7
FC3-gender	Fully Connection	1*2

The multi-task learning model is designed to fulfill the requirements of a single-input, single-model, multiple-output system with shared bottom-level parameters. Fig. 3 shows the proposed architecture of the multi-task learning model. The shared part of the network adopts ResNet-50 [18] while excluding the last pooling and fully connected layers. The non-shared part consists of two convolutional layers followed by several fully connected layers. The output of the model is normalized using softmax to obtain probability. Table II shows the parameter structure of the entire network. We train the multi-task learning model until convergence using SGD (Stochastic Gradient Descent) with a learning rate of 0.001, batch size 256, and a momentum of 0.75.

III. EXPERIMENTS AND RESULTS

To evaluate the performance of the system, we will conduct comprehensive tests focusing on the accuracy and speed of recognition in both the dataset and real-world scenes. In addition, we will evaluate whether the system exhibits false positives or false negatives when occupants are in different poses.

A. In Dataset

We tested the clothing recognition functionality in the collected dataset, and it achieved mAP in 92.82%. Likewise, we conducted tests on the face attribute recognition functionality using three different datasets: AFAD, UTK-Face, and MegaAge. Table III shows the performance of facial attribute recognition in different datasets.

TABLE III. FACIAL ATTRIBUTE RECOGNITION ACCURACY

Dataset	Age	Gender
AFAD	69.04%	83.67%
UTK-Face	60.48%	86.16%
MegaAge	61.20%	80.02%

Fig. 5: Recognition results of the real-world scenes.

From Table III, we can conclude that the average accuracy rates for age and gender recognition across different datasets are 63.57% and 83.28%. It is worth noting that the dataset used for training is predominantly concentrated in the youth stage (15-45). Consequently, the accuracy of recognition is higher within this age interval. Fig. 4 shows the recognition results of the dataset, with the corresponding results displayed at the bottom of each image.

B. In Real-World Scene

The occupant attribute recognition system captures real-time occupant images through the in-vehicle camera and feeds the images into the model to obtain outputs that contain occupant attributes. We recorded three videos, each approximately 1 minute, to evaluate the accuracy and speed of the system in real-world scenes. Fig. 5 shows the recognition results during the operation of our system. When all videos were tested, the overall accuracy of recognizing the occupant attributes reached 91%. The system runs at 56 FPS on a GeForce GTX 1050Ti with 4 GB of memory.

IV. Conclusion

In this paper, an occupant attribute recognition system is proposed to provide a basis for evaluating thermal comfort in the passenger car cabin. The system consists of YOLO and multi-task learning model, enabling simultaneous recognition of occupant gender, age, clothing type, and seating position. The conducted tests demonstrate that the proposed system exhibits enhanced accuracy and real-time performance in occupant cabin environments.

In future work, the system will be enhanced by adding the capability to recognize additional passenger attributes such as body type (e.g., overweight, slim), identity, and pose.

Acknowledgment

This work was supported by National key R&D program (No.2022YFE0101000).

References

[1] A. Lajunen, Y. Yang and A. Emadi, "Review of Cabin Thermal Management for Electrified Passenger Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 6, pp. 6025-6040, 2020.

[2] A. Vartires, I. Colda, and S. M. Toropoc, "Research on thermal comfort and indoor air pollution in a passenger car," 2017 International Conference on ENERGY and ENVIRONMENT (CIEM), pp. 148-152, 2017.

[3] S. Del Ferraro, S. Iavicoli, S. Russo, and V. Molinaro, "A field study on thermal comfort in an Italian hospital considering differences in gender and age," Applied Ergonomics, vol. 50, pp. 177–184, 2015.

[4] S. Karjalainen, "Thermal comfort and gender: a literature review," Indoor Air, vol. 22, no. 2, pp. 96–109, 2011.

[5] W. Zhang, S. Guan, C. Wang, Y. Zhang, and X. Zhou, "Research on Face Detection and Face Attribute Recognition based on Deep Learning," 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 18-22, 2021.

[6] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," arXiv.org, 2017.

[7] H. Liu, X. Shen, and H. Ren, "FDAR-Net: Joint Convolutional Neural Networks for Face Detection and Attribute Recognition," 2016 9th

International Symposium on Computational Intelligence and Design (ISCID), pp. 184-187, 2016.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[9] W. Liu et al., “SSD: Single Shot MultiBox Detector,” Computer Vision – ECCV 2016, pp. 21–37, 2016.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.

[11] Eshwar S G, Gautham Ganesh Prabhu J, A. V. Rishikesh, Charan N A, and Umadevi V, "Apparel classification using Convolutional Neural Networks," 2016 International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1-5, 2016.

[12] Y. Ge, R. Zhang, X. Wang, X. Tang, and P. Luo, "DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5332-5340, 2019.

[13] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, “Apparel Classification with Style,” Computer Vision – ACCV 2012, pp. 321–335, 2013.

[14] Z. Niu, M. Zhou, L. Wang, X. Gao and G. Hua, "Ordinal Regression with Multiple Output CNN for Age Estimation," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4920-4928, 2016.

[15] “UTKFace,” UTKFace. https://susanqq.github.io/UTKFace/

[16] Y. Zhang, L. Liu, C. Li, and C c. Loy, “Quantifying Facial Age by Posterior of Age Comparisons,” arXiv:1708.09687 [cs], 2017

[17] A. Hafner, P. Peer, Ž. Emeršič and M. Vitek, "Deep Iris Feature Extraction," 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 258-262, 2021

[18] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.