EMPCNet: Facial Attribute Recognition using Efficient Multi-Perspective Convolution for Human–Robot Interaction

Adri Priadana, Duy-Linh Nguyen, Xuan-Thuy Vo, Russo Mohammad Ashraf Uddin, and Kanghyun Jo
Department of Electrical, Electronic, and Computer Engineering, University of Ulsan, Ulsan, Korea
Email: priadana3202@mail.ulsan.ac.kr; ndlinh301@mail.ulsan.ac.kr; xthuy@islab.ulsan.ac.kr;
ashrafrusso@gmail.com; acejo@ulsan.ac.kr

***Abstract*—Human-Robot Interaction has become a significant field in robotics. In this domain, facial attributes are essential as they enable robots to understand human emotions, intentions, and preferences. In robot applications, which typically involve low-cost devices, efficient recognition technology is crucial for promising real-time operation by robots. This work proposes EMPCNet to perform facial attribute recognition, consisting of an Efficient Multi-Perspective Convolution (EMPC) block used to efficiently extract and capture various information from multiple perspectives using different kernel sizes and shapes of convolutional operations. The proposed network, which only utilizes a few parameters and low computational operations, achieves competitive performance on the CelebA and LFWA datasets. Additionally, when integrated with face detection, the proposed EMPCNet operates efficiently in real-time on an Intel Core i7-9750H CPU, achieving a frame rate of 21.27 frames per second (FPS) with an image input size of 224 × 224 consisting of a face area.**

***Index Terms*—Convolutional Neural Network, Efficient Multi-Perspective Convolution, Facial Attribute Recognition, Human–Robot Interaction, Real-time Recognition.**

I. Introduction

Human-Robot Interaction (HRI) has emerged as a pivotal domain within robotics. This technology reflects a growing trend toward integrating artificial intelligence (AI) with physical systems, enabling seamless communication and collaboration between humans and robots. Recent advancements in AI and robotics have propelled HRI beyond traditional industrial settings into various sectors, including healthcare [1], education [2], and retail [3]. Robots are increasingly becoming a part of modern society's life as companions and assistants.

Facial attributes play a crucial role in supporting HRI by enabling robots to perceive and interpret human emotions, intentions, and preferences. Through facial attribute recognition, robots can decipher facial expressions, age, gender, and other characteristics [4], [5], allowing them to adapt their behavior and responses to better engage with humans. This capability enhances the overall user experience and fosters more natural and intuitive interactions between humans and robots. It paves the way for broader acceptance and integration of robotic technologies into daily life. While facial attribute recognition appears to be a simple task of classifying images, it poses challenges due to its slow classification speeds, especially on a low-cost device. Additionally, the wide variety of facial appearances resulting from factors such as variations in illumination and viewpoint further complicates the task [6].

In recent years, Convolutional Neural Networks (CNNs) have been the cornerstone of facial attribute recognition. For instance, MCFA [7] introduced a novel multi-task learning of cascaded CNN to predict multiple facial attributes simultaneously. Another approach, SPLITFACE [8], proposed a deep CNN-based method to address partial occlusion in facial attribute recognition. DMM-CNN [6] offered a deep multi-task and multi-label CNN designed to extract features for objective and subjective attribute groups. These previously mentioned works used deep CNN, generating extensive parameters.

While CNNs have dominated, Transformer models emerged as challengers, promising superior recognition performance in computer vision tasks. MZTS [9] developed a multi-zone transformer, achieving high accuracy on the CelebA dataset. Despite their high accuracy, these approaches often come with heavy parameterization and computational requirements, making them unsuitable for low-cost or CPU devices. SCTE [10] introduced an efficient network with a squeeze channel transformer encoder for facial attribute recognition, yet it still entails significant parameters and computations. An efficient facial attribute recognition network is vital in HRI applications using a low-cost device to minimize implementation costs and resource usage. Achieving this requires a network capable of real-time operation with minimal memory and computational resources.

This work proposes a lightweight CNN, dubbed EMPCNet, consisting of an Efficient Multi-Perspective Convolution (EMPC) block. This network facilitates streamlined facial attribute recognition, achieving minimal parameter generation and operational overhead. Here is an overview of the contributions made in this study:

1) An EMPCNet, which incurs low computation and generates few parameters, is proposed for facial attribute recognition in human-robot interaction. EMPCNet achieves highly competitive accuracy compared to other networks on CelebA [11] and LFWA [12] datasets.
2) An Efficient Multi-Perspective Convolution (EMPC) block is introduced to efficiently capture various information based on many perspectives. This mechanism

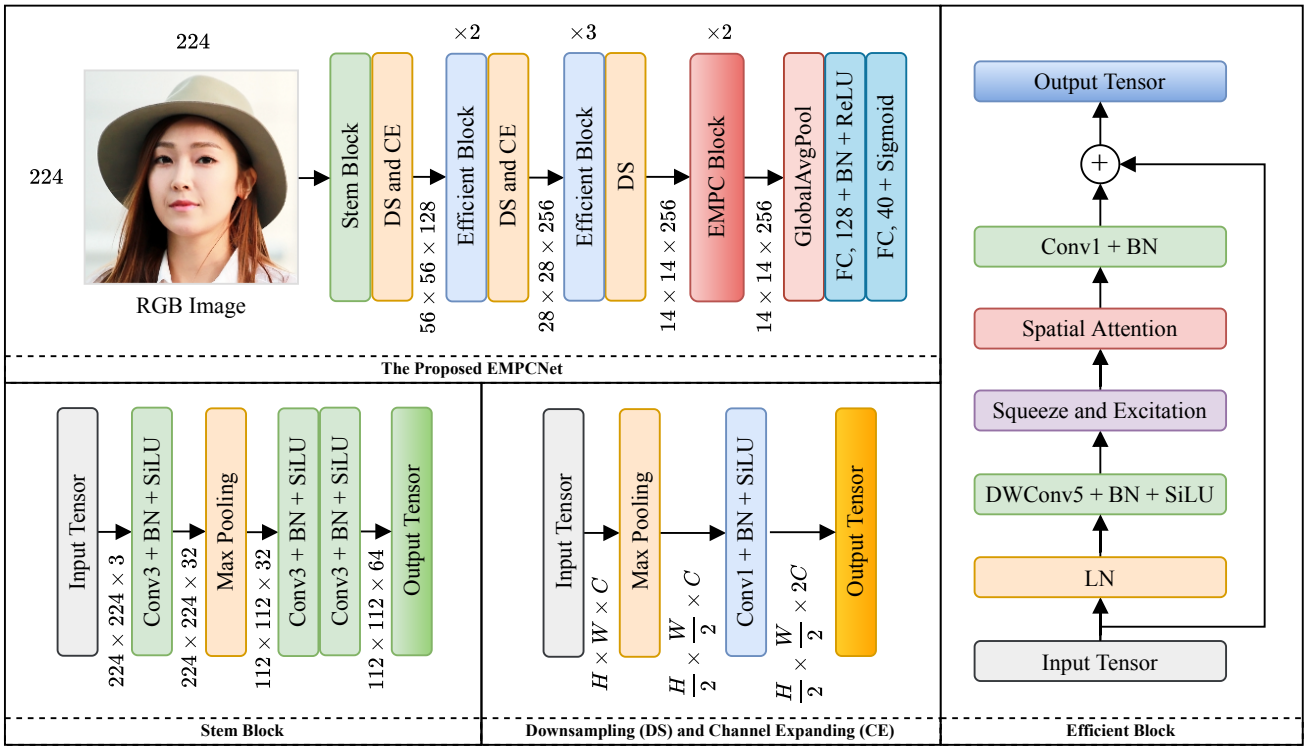

Fig. 1. The proposed EMPCNet, consisting of the proposed Efficient Multi-Perspective Convolution (EMPC) block. Conv, DWConv, and FC refer to Convolution, Depthwise Convolution, and Fully Connected layers, respectively. BN and LN denote Batch and Layer Normalization. SiLU and ReLU are Sigmoid Linear Unit and Rectified Linear Unit activation functions.

is achieved using different kernel sizes and shapes of convolution operation, supported by attention modules. It boosts the feature map quality, improving the recognition accuracy.

3) The proposed facial attribute recognition network comprises approximately 883,368 parameters and 2,968 MFLOPs. By utilizing face detection as the initial process, this network performs at 21.27 frames per second on an Intel Core i7-9750H CPU in real-time without necessitating extensive memory or computational resources.

II. Proposed Network

The proposed EMPCNet consists of several blocks, followed by classification modules, as shown in Fig. 1. This network introduces an Efficient Multi-Perspective Convolution (EMPC) block positioned as the last block before the classification modules.

A. Efficient Multi-Perspective Convolution (EMPC) Block

CNNs have demonstrated promising performance in tasks related to classifying images, attributed to their capacity to capture local patterns within images, determined by the kernel size [13], [14]. To enhance their capability further, CNNs can use multi-kernel sizes or shapes, enabling them to capture features at various scales or shapes within input images. This approach allows the network to learn more diverse and various perspective representations with different kernel sizes, potentially improving its performance. Drawing inspiration from previous works such as Multi-Perspective Convolution Network (MPConvNet) [15], which utilized multi-kernel sizes, and InceptionNeXt [16], which employed multi-kernel shapes, this work introduces an Efficient Multi-Perspective Convolution (EMPC) block.

Different from MPConvNet [15] and InceptionNeXt [16], EMPC performs Layer Normalization (LN) on the input tensor $\mathbf{X}$ and then divides the result based on channel dimensions into four parallel branches with the same channel and performs different kernel sizes and shapes of depthwise convolution operations. It splits the LN's feature map output $\mathbf{X}'$ into four parts $[\mathbf{X}'_1, \mathbf{X}'_2, \mathbf{X}'_3, \mathbf{X}'_4]$ based on channel axes and involves small kernels (3×3), medium kernels (5×5), and two band kernels (11×1 and 1×11), respectively. Spatial attention consisting of average pooling operation across channels and sigmoid activation is used to enhance the quality of each branch feature map based on a spatial point of view. Subsequently, this block uses concatenation operation to combine all branches and performs LN, Squeeze and Excitation (SE) [17], and Feed Forward (FF), as shown in Fig. 2. SE involves global pooling and two times 1×1 convolutions with Rectified Linear Unit (ReLU) and sigmoid activation after the first and second convolution, respectively. FF encompasses two times 1×1 convolutions with Batch Normalization (BN) and Gaussian Error Linear Unit (GELU) activation after the first convolutions and only BN after the second convolution. This

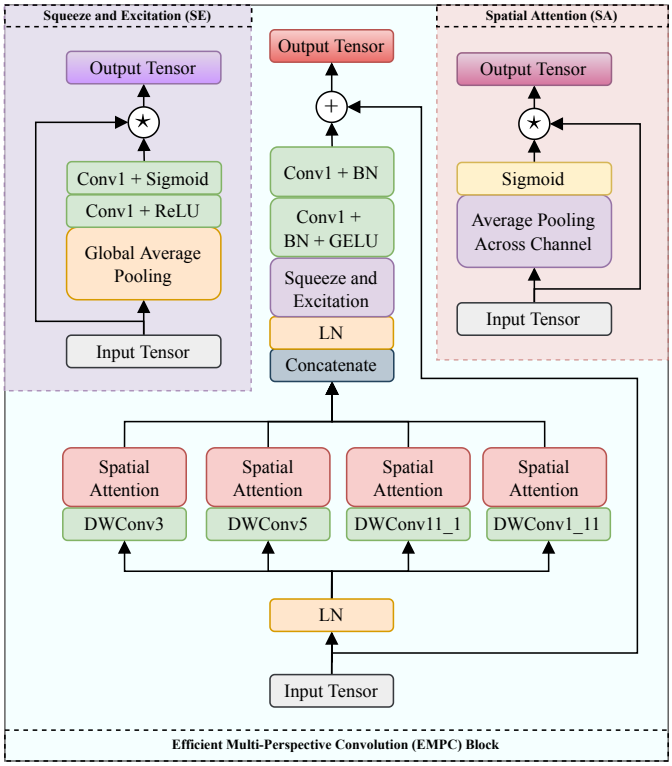

Fig. 2. The proposed Efficient Multi-Perspective Convolution (EMPC) block. Conv and DWConv refer to Convolution and Depthwise Convolution layers, respectively. BN and LN denote Batch and Layer Normalization. GELU is Gaussian Error Linear Unit activation.

block also utilizes skip connection as a residual mechanism. The overall EMPC block is defined as follows:

$$\mathbf{X}'(\mathbf{X}) = \mathrm{LN}(\mathbf{X}), \tag{1}$$

$$\begin{aligned}\mathrm{EMPC}(\mathbf{X},\mathbf{X}') = \mathbf{X} + \mathrm{FF}(\mathrm{SE}(\mathrm{LN}(\mathrm{Concat}[\\ \mathrm{SA}(\mathrm{DW}_{3\times3}(\mathbf{X}'_{\mathbf{1}})),\\ \mathrm{SA}(\mathrm{DW}_{5\times5}(\mathbf{X}'_{\mathbf{2}})),\\ \mathrm{SA}(\mathrm{DW}_{1\times11}(\mathbf{X}'_{\mathbf{3}})),\\ \mathrm{SA}(\mathrm{DW}_{11\times1}(\mathbf{X}'_{\mathbf{4}}))])),\end{aligned} \tag{2}$$

$$\mathrm{SA}(\mathbf{Y}) = \mathbf{Y} * \mathrm{APAC}(\sigma(\mathbf{Y})), \tag{3}$$

$$\mathrm{SE}(\mathbf{Z}) = \mathbf{Z} * \mathrm{GAP}(\sigma(\mathrm{C}(\delta(\mathrm{C}(\mathbf{Z}))))), \tag{4}$$

$$\mathrm{FF}(\mathbf{A}) = \mathrm{BN}(\mathrm{C}(\pi(\mathrm{BN}(\mathrm{C}(\mathbf{A}))))), \tag{5}$$

where C indicates 1×1 convolution operation and $\mathrm{DW}_{\mathrm{m\times n}}$ indicates $m\times n$ kernel size of depthwise convolution operations. $\mathbf{X}$, $\mathbf{X}'$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{A}$ are inputs of function. FF, SE, Concat, SA, APAC, and GAP refer to Feed Forward, Squeeze and Excitation, Concatenation, Spatial Attention, Average Pooling Across Channel and Global Average Pooling, respectively. σ, δ, and π denote sigmoid, ReLU, and GELU activations. LN and BN are Layer and Batch Normalization, respectively.

B. Overall Network

The network proposed in this study comprises one Stem block, two Efficient blocks, and one EMPC block, as depicted in Fig. 1. These blocks are connected by Downsampling (DS) and Channel Expanding (CE) modules. DS employs max-pooling while CE utilizes a 3 × 3 convolution with BN and ReLU activation, which serve for spatial downsampling and channel number expansion, respectively. However, in the final part, only the DS module is utilized. As shown in Fig. 1, the Stem block consists of a max-pooling and three of 3×3 convolution with BN and ReLU activation, which downsamples the spatial dimension and expands the channel of the input image from $224\times224\times3$ to $112\times112\times64$ pixel.

The Efficient (Eff) block is equipped with a 5×5 depthwise convolution with BN and Sigmoid Linear Unit (SiLU) activation to efficiently capture information across a wider field. Additionally, it includes a 1×1 convolution with BN as a projection to integrate information across channels. Between those convolutions, SE and SA are also used to enhance the feature map quality. Layer Normalization (LN) is preceded before the depthwise convolution operation, and skip connections are employed as a residual mechanism. The overall Efficient block is formulated as follows:

$$\mathrm{Eff}(\mathbf{X}) = \mathrm{BN}(\mathrm{C}(\mathrm{SA}(\mathrm{SE}(\sigma'(\mathrm{BN}(\mathrm{DW}_{5\times5}(\mathbf{X}))))))), \tag{6}$$

where σ' represents the SiLU activation function. Following the Efficient block, the classification module operates in the final section of the network, comprising two fully connected (FC) layers. BN and ReLU activation are applied to the first FC layer, while sigmoid activation is used after the second FC layer, enabling multi-label classification similar to the prior facial attribute recognition approach [10].

III. Implemental Configuration

This work employs the CelebA and LFWA datasets to train the proposed network using the Tensorflow and Keras framework on an NVIDIA GTX1080Ti 11GB GPU. To enhance generalization and prevent overfitting, various data augmentation techniques, such as rotation, rescaling, and shifting, are implemented during training, similar to the previous work [10]. 10^{-3} is assigned as the initial learning rate, with a 75% reduction in the absence of improvement after five epochs. This training utilizes the Adam optimizer with the Cosine Similarity loss function. This work uses a batch size of 32 and trains for 30 epochs on the CelebA dataset, while it operates a batch size of 16 and trains for 60 epochs on the LFWA dataset. The performance of the proposed EMPCNet is evaluated in real-time scenarios using an Intel Core i7-9750H CPU @ 2.60GHz with 20GB RAM.

IV. Experimental Results

A. Evaluation on Datasets

1) CelebA: The CelebFaces Attributes (CelebA) [11] dataset comprises 202,599 face images, each annotated with multiple labels, covering poses and background variations across 40 binary attributes. This dataset provides aligned and cropped versions, divided into training (162,770), validation (19,867), and testing (19,962) image sets. The proposed EMPCNet, with only 883,368 parameters and 2,968 MFLOPs,

TABLE I
THE ASSESSMENT OUTCOMES ON THE CELEBA [11] DATASET.

Networks	Input Dimension (Pixel)	Data Augmentation	Number of Parameters (Million)	Average Accuracy (%)
SPLITFACE [8]	196 × 196	Yes	26.09	90.61
MCFA [7]	224 × 224	No	260.00	91.23
SOP [18]	224 × 224	No	4.99	91.26
MCN-AUX [19]	224 × 224	No	16.00	91.29
SCTE [10]	224 × 224	Yes	2.11	91.50
MZTS [9]	224 × 224	No	85.83	**91.66**
DMM-CNN [6]	224 × 224	No	43.59	**91.70**
EMPCNet (ours)	**224 × 224**	**Yes**	**0.88**	**91.52**

TABLE II
THE ASSESSMENT OUTCOMES ON THE LFWA [12] DATASET.

Networks	Input Dimension (Pixel)	Data Augmentation	Number of Parameters (Million)	Average Accuracy (%)
MCFA [7]	196 × 196	Yes	260.00	83.63
SPLITFACE [8]	196 × 196	Yes	26.09	85.82
MCN-AUX [19]	224 × 224	Yes	16.00	86.31
SCTE [10]	224 × 224	Yes	2.11	86.45
DMM-CNN [6]	224 × 224	No	43.59	**86.56**
MZTS [9]	224 × 224	No	85.83	**86.73**
EMPCNet (ours)	**224 × 224**	**Yes**	**0.88**	**86.48**

achieves an average accuracy of 91.52% on this dataset, ranking third among other networks. Table I shows the accuracy of all networks, in which the proposed EMPCNet is slightly below the top-ranked model by 0.18% and the second-ranked model by 0.14%. Nevertheless, the proposed EMPCNet boasts significantly fewer parameters.

2) LFWA: The Labeled Faces in the Wild Attributes (LFWA) [12] dataset comprises 13,143 face images, each labeled with multiple attributes, covering a wide range of lighting conditions, poses, ages, occlusions, and expressions. This dataset includes annotations for 73 binary attributes. Similar to the previous studies [6], [9], [10], this work utilizes the identical set of 40 attributes from this dataset as employed in the CelebA dataset. The dataset is partitioned into training (6,572) and testing (6,571) image sets. The proposed EMPCNet achieves an average accuracy of 86.48% on this dataset, ranking third among other networks. This performance is slightly lower than the top-ranked model by 0.25% and the second-ranked model by 0.08%, as illustrated in Table II.

B. Ablation Study

This analysis, conducted on the CelebA dataset, involves removing the proposed EMPC block from the network and evaluating the average accuracy to analyze the influence of each component. We also assess how the recognition performance is affected by the individual attention modules inside the EMPC block. Table III demonstrates that applying individual Spatial Attention (SA) and Squeeze and Excitation (SE) modules on the proposed EMPC block increases the average accuracy by 0.07% and 0.08%, respectively. Moreover, the proposed EMPC block with attention modules inside, which only adds a few extra parameters and low operations, can enhance the average accuracy by 0.18%.

TABLE III
THE ABLATION STUDY OF THE PROPOSED EMPC BLOCK ON CELEBA [11] DATASET

Settings	Number of Parameters	MFLOPs	Average Accuracy (%)
w/o EMPC	705,960	2,910	91.34
EMPC w/o SE and SA	850,600	2,967	91.41
EMPC w/ SA	850,600	2,967	91.48
EMPC w/ SE	883,368	2,967	91.49
EMPC w/ SE and SA	883,368	2,968	91.52

w/ indicates with
w/o indicates without

C. Runtime Efficiency

HRI necessitates facial attribute recognition coupled with face detection as the initial process to execute real-time recognition on a CPU device in the practical implementation. According to this scenario, this work explores the runtime performance of the proposed EMPCNet across three different input sizes: 224×224, 112×112, and 56×56 pixels. In this procedure, smaller input sizes result in lower Floating Point Operations Per Second (FLOPs), enabling faster recognition. Consequently, this comes at the cost of reduced average accuracy due to decreased information acquisition from smaller input images, as demonstrated in Table IV. Additionally, we integrate the proposed face attribute recognition with LWFCPU [20] as an efficient face detector to acquire face area. Afterward, this face area is expanded (covering the area around

TABLE IV
RUNTIME EFFICIENCY OF THE EMPCNET WITH DIFFERENT INPUT DIMENSION ON A CPU DEVICE

Input Dimension (Pixel)	Number of Parameters	MFLOPs	Average Accuracy (%)	FAR (FPS)	FD + FAR (FPS)
224×224	883,368	2,968	91.52	23.25	21.27
112×112	883,368	742	91.02	64.53	51.52
56×56	883,368	185	90.00	119.72	82.16

FAR denotes Facial Attribute Recognition
FD + FAR denotes Facial Attribute Recognition integrated with Face Detection

the face), cropped, and resized to a specific size suitable for the input size of the proposed face attribute recognition network. Utilizing a 224×224 size as the face area input image and implementing it on an Intel Core i7-9750H CPU @ 2.60GHz with 20GB RAM, the proposed EMPCNet achieves a recognition speed of 23.25 FPS for human face attribute recognition. Moreover, it achieves a recognition speed of 21.27 FPS when integrated with face detection. These results affirm that the proposed network is capable of deployment on a robot with a CPU device to support HRI.

D. Qualitative Results and Discussion

Fig. 3 illustrates the outcome of facial attribute recognition performed by the proposed EMPCNet, trained on the CelebA dataset and integrated with LWFCPU [20] as an efficient face detector. In this illustration, the blue bounding box indicates a male face, while the green bounding box denotes a female face. The red bounding box outlines the face's region of interest obtained from the face detection process. Despite the CelebA dataset containing only a male gender label, the recognizer can deduce the presence of a female face according to the probability value of the male gender label. A minimal or near-zero probability value for the male class indicates the detected face is likely female. Due to display limitations, the proposed recognizer showcases only the six most probable facial attributes, if space permits. These attributes are arranged in descending order of probability value, with the attribute having the highest probability positioned at the bottom of the screen. The proposed recognizer successfully recognizes various facial attributes such as wearing hats, eyeglasses, bags under eyes, smiles, wearing lipstick, straight hair, black hair, bangs, gender, young, etc. By recognizing these facial attributes, robots can adapt their responses and provide recommendations to better engage with humans, thereby increasing the quality of HRI.

V. CONCLUSION

This work proposes EMPCNet to perform facial attribute recognition. This work offers a lightweight CNN, consisting of an Efficient Multi-Perspective Convolution (EMPC) block used to capture various information based on many perspectives using different kernel sizes and shapes of convolution operation, supported by attention modules to boost the feature map quality. As a result, the proposed EMPCNet achieves competitive performance compared to other methods on the CelebA and LFWA datasets. Additionally, it demonstrates effective real-time performance on a CPU setup, achieving 23.25 FPS for recognizing facial attributes and 21.27 FPS when combined with face detection utilizing a 224×224 size of the face area input image. For further work, this recognizer will be directly implemented or embedded in a robot to support HRI applications.

ACKNOWLEDGMENT

This result is supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE)(2021RIS-003)

REFERENCES

[1] F. Soljacic, T. Law, M. Chita-Tegmark, and M. Scheutz, "Robots in healthcare as envisioned by care professionals," *Intelligent Service Robotics*, pp. 1–17, 2024.

[2] Y. Cui, X. Song, Q. Hu, Y. Li, P. Sharma, and S. Khapre, "Human-robot interaction in higher education for predicting student engagement," *Computers and Electrical Engineering*, vol. 99, p. 107827, 2022.

[3] I. Roozen, M. Raedts, and A. Yanycheva, "Are retail customers ready for service robot assistants?" *International Journal of Social Robotics*, vol. 15, no. 1, pp. 15–25, 2023.

[4] M. D. Putro, D.-L. Nguyen, and K.-H. Jo, "A fast cpu real-time facial expression detector using sequential attention network for human–robot interaction," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7665–7674, 2022.

[5] H. A. Younis, N. I. R. Ruhaiyem, A. A. Badr, A. K. Abdul-Hassan, I. M. Alfadli, W. M. Binjumah, E. A. Altuwaijri, and M. Nasser, "Multimodal age and gender estimation for adaptive human-robot interaction: A systematic literature review," *Processes*, vol. 11, no. 5, p. 1488, 2023.

[6] L. Mao, Y. Yan, J.-H. Xue, and H. Wang, "Deep multi-task multi-label cnn for effective facial attribute classification," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 818–828, 2022.

[7] N. Zhuang, Y. Yan, S. Chen, and H. Wang, "Multi-task learning of cascaded cnn for facial attribute classification," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 2069–2074.

[8] U. Mahbub, S. Sarkar, and R. Chellappa, "Segment-based methods for facial attribute detection from partial faces," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 601–613, 2020.

[9] S. Chen, X. Zhu, D.-H. Wang, S. Zhu, and Y. Wu, "Multi-zone transformer based on self-distillation for facial attribute recognition," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 2023, pp. 1–7.

[10] A. Priadana, M. D. Putro, J. An, D.-L. Nguyen, X.-T. Vo, and K.-H. Jo, "Facial attribute recognition using lightweight multi-label cnn-transformer architecture for intelligent advertising," in *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*, 2023, pp. 1–7.

[11] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

Fig. 3. The qualitative results of the proposed facial attribute recognizer.

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 966–11 976.

[14] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 16 133–16 142.

[15] A. Priadana, M. D. Putro, and K.-H. Jo, "An efficient face gender detector on a cpu with multi-perspective convolution," in *2022 13th Asian Control Conference (ASCC)*, 2022, pp. 453–458.

[16] W. Yu, P. Zhou, S. Yan, and X. Wang, "Inceptionnext: When inception meets convnext," *arXiv preprint arXiv:2303.16900*, 2023.

[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.

[18] Y. Lu, A. Kumar, S. Zhai, Y. Cheng, T. Javidi, and R. Feris, "Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1131–1140.

[19] E. M. Hand and R. Chellappa, "Attributes for improved attributes: A multi-task network for attribute classification," *arXiv preprint arXiv:1604.07360*, 2016.

[20] M. D. Putro, D.-L. Nguyen, and K.-H. Jo, "Lightweight convolutional neural network for real-time face detector on cpu supporting interaction of service robot," in *2020 13th International Conference on Human System Interaction (HSI)*, 2020, pp. 94–99.