Age and Gender Recognition for Masked Face Using YOLO-X and CNN in Smart Advertisement Systems

Handoko[1*], Aaron Berliano Handoko[2], and Darmawan Utomo[3]
[1−3]Teknik Komputer, Fakultas Teknik Elektronika dan Komputer, Universitas Kristen Satya Wacana
Jawa Tengah 50711, Indonesia
Email: [1]handoko@uksw.edu, [2]handoko.aaron@gmail.com, [3]darmawan.utomo@uksw.edu

***Abstract*—The conventional advertisement board often fails to attract its target customers effectively due to its limited ability to display content relevant to viewers. To address this, a Smart Personalized Advertisement (SAVER) board employing an age and gender recognition system is proposed. In the post-pandemic era, where many people wear face masks, developing effective smart advertising systems has become even more challenging. This study aims to evaluate and compare Convolutional Neural Network (CNN) architectures integrated with You Only Look Once-X (YOLO-X) for age and gender recognition in smart advertising applications that accommodate both masked and unmasked faces. The proposed framework first detects faces in an image using the YOLO-X model. The detected faces are then cropped based on bounding boxes and aligned to ensure consistent orientation. Subsequently, CNNs classify age groups and gender based on facial attributes. The detection results are used to determine which advertisements should be displayed. This study uniquely addresses the recognition of age and gender for both masked and unmasked faces and implements the solution in a real-time advertising system. The proposed system achieved 68% precision in delivering smart personalized advertisements, demonstrating its effectiveness in real-world public settings. In summary, this research contributes to the development of intelligent public display systems capable of delivering demographically aware content.**

***Index Terms*—Personalized Advertisement Board, Age and Gender Recognition, Masked and Unmasked Face, You Only Look Once-X (YOLO-X), Convolution Neural Network (CNN)**

I. Introduction

EFFECTIVE advertising is a key determinant of business success, particularly in today's increasingly competitive market. Previous studies have shown that effective advertisements should be creative, visually appealing, and relevant to the target audience [1–4]. In fact, relevant advertisements lead to more favourable attitudes from their audiences [5]. Personalized advertising technologies have been employed to address this demand by delivering content tailored to users based on their browsing behavior. Such technologies have been shown to influence purchase intentions and increase purchasing frequency [6]. However, their application remains limited to personal environments and is not yet widely implemented in public spaces.

Various approaches have been explored to achieve personalized advertising in public settings, particularly through the use of facial features such as gaze, gender, and age detection. Previous research has utilized a Convolutional Neural Network (CNN) to recognize age and gender for determining which advertisement to display, achieving 91.7% accuracy for age recognition and 95.5% for gender recognition [7]. CNN is also used for age and gender recognition, achieving 94% accuracy for gender detection and 86% for age group classification [6]. Then, another research has demonstrated that CNN outperforms Support Vector Machines (SVM) in terms of accuracy and robustness [8]. Furthermore, previous researchers have proposed CNN-based approaches for age and gender detection [9, 10]. In addition, another prototype of a smart advertisement board incorporates intention detection through eye movement, but its performance is limited, achieving only 68% accuracy in identifying user intention [11]. These findings suggest that facial features remain an effective means for implementing smart advertising systems in public environments.

The COVID-19 pandemic, however, introduces new challenges. Beyond its profound health impacts, the pandemic has triggered substantial social, cultural, and behavioral changes. One of the most visible transformations is the adoption of the "new normal," in which wearing masks in public becomes a necessity [12].

Received: June 06, 2025; received in revised form: Aug. 15, 2025; accepted: Aug. 15, 2025; available online: Oct. 13, 2025.
*Corresponding Author

Although the fear of COVID-19 contamination is starting to fade, the fact that a large number of people still prefer to wear masks in public places poses troublesome challenges for technologies that rely on facial feature recognition, especially face detection. However, relatively few studies have examined the performance of age and gender recognition systems on both masked and unmasked faces within real-time advertising applications.

The objective of the research is to develop a smart advertisement system that integrates You Only Look Once-X (YOLO-X) with CNN to enable accurate face recognition in both masked and unmasked conditions. It also aims to implement this system in a real-time prototype. Specifically, the proposed smart advertisement board is capable of dynamically displaying content tailored to the detected audience. The scope of the research includes: (1) a face detection system using YOLO-X trained on a publicly available dataset, and (2) a classification system using fixed age group categories (3–9, 10–19, 20–39, 40–59, $>$ 59) and gender categories (male and female). However, the research does not address the recognition of emotions or ethnicities, nor does it attempt to infer viewers' personal preferences, such as fashion or appearance.

A. Related Works

In recent years, many researchers have increasingly approached face detection as a general object detection task [13, 14]. However, with the introduction of regulations requiring the use of masks in public spaces, research on masked face detection has developed rapidly [15]. Detecting masked faces presents a distinct challenge as the presence of masks significantly reduces the performance of face detection systems, as they obscure critical facial features [16].

A variety of algorithms have been developed to address this problem, including YOLO [17–19], MobileNetV2 [20], and RetinaFace [21] algorithms. The recent YOLO algorithm, YOLO-X [22], has been widely applied to various object detection tasks, including road asset detection [23], foreign object detection [24], and masked face detection [25]. Prior studies have reported that YOLO-X outperforms MobileNetV2, achieving 95% accuracy in masked face detection [25]. These findings provide strong justification for employing YOLO-X as the face detection algorithm in the research, given its lightweight architecture and ability to perform effectively in real-time detection scenarios.

CNNs are among the most effective algorithms for image recognition and segmentation tasks. A CNN typically consists of three main types of layers: convolutional layers, pooling layers, and fully connected layers. Convolutional layers filter input images by convolving them with specified kernels to extract features. Pooling layers further refine these outputs by down-sampling the feature maps, while fully connected layers integrate the extracted features to generate final predictions [26].

Age and gender recognition is a machine learning task that relies on facial patterns as classification parameters [27]. Numerous studies have investigated this topic, as summarized in Table I. However, most existing models have not been evaluated on masked faces, with the exception of MobiFace and the FaceMaskNet-9 model. According to previous research, MobiFace achieves low accuracy results when detecting age and gender for masked faces [28]. While FaceMaskNet-9 shows promising results in age detection, it has only been tested using nine different individuals. Based on the previous research, YOLO-X can effectively distinguish between masked and unmasked faces [25]. Therefore, in this research, the researchers evaluate multiple age and gender recognition models combined with YOLO-X as the face detection algorithm.

II. Research Method

The proposed SAVER Board workflow is illustrated in the block diagram shown in Fig. 1. First, images captured from a webcam mounted above the advertisement board undergo preprocessing. If YOLO-X detects a face in the image, the model generates bounding boxes and facial landmark coordinates with each detected face. These outputs are then used to crop and align the faces. Second, the aligned faces are processed by CNN models for age and gender recognition. Based on the recognition outputs, the system displays the corresponding advertisement on the monitor.

A. Face Detection

The research adopts the YOLO-X architecture as described in previous research [25]. The model detects both masked and unmasked faces, along with five facial landmarks: the centers of the left and right eyes, the nose, and the left and right corners of the lips. To improve performance, the model is retrained using the Masked Face in the Wild (MAFA) [31], WiderFace [32], and CelebA datasets. Due to the lack of datasets for masked faces, artificial masks are applied to the image in the existing face dataset. The artificial masks are generated using the MaskTheFace model [33]. The final dataset comprises 214,541 masked faces and 280,983 unmasked faces, which are randomly split into training (80%), testing (10%), and validation (10%) subsets. Performance is evaluated using Accuracy, Precision, Recall, and F1-Score.

TABLE I
COMPARISON OF EXISTING AGE AND GENDER RECOGNITION ALGORITHMS.

Algorithm	Result	Advantage	Disadvantage
Convolutional Neural Network (CNN) [6]	95.5% accuracy in detecting gender and 91.7% in age	Having the best performance when compared with VGG-16 and Deep CNN	The system has not been evaluated with masked faces
VGG-16 [28]	93% accuracy in detecting gender and 61.8% in age group detection	Detecting age and gender with a low-resolution image	The algorithm has not been tested with masked faces
ResNet-34 [29]	94.9% accuracy in detecting gender and 60% in age	Being trained on a balanced dataset	The system has not been tested with masked faces
MobiFace [28]	76.6% and 45.9% accuracy in detecting gender and age group for masked faces, respectively	Having lightweight network architecture model	The accuracy of the model in detecting age is still under 50%
FaceMaskNet-9 [30]	94.96% accuracy in detecting age	Achieving high accuracy in detecting age for masked faces	The evaluation dataset consists of only nine subjects.

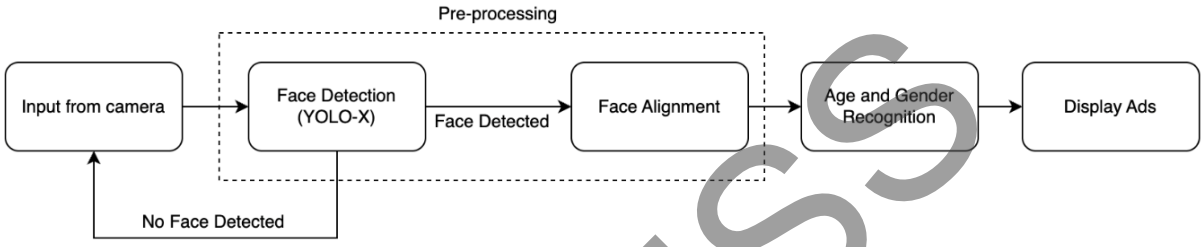


Fig. 1. The SAVER board framework.

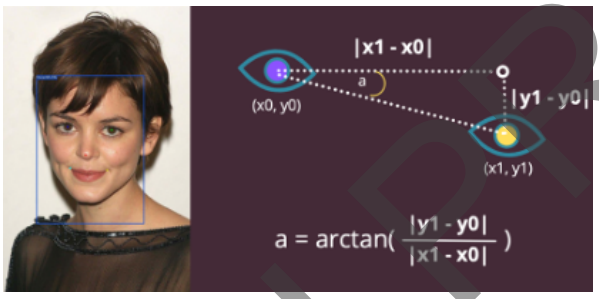


Fig. 2. Face Alignment using Arcus Tangent.

B. Face Alignment

Face alignment ensures that detected faces are consistently oriented, which improves age and gender classification. SAVER Board uses the left and right eye coordinates as alignment landmarks (Fig. 2). Detected faces are cropped based on bounding boxes, and the rotation angle is calculated using the arctangent equation (Eq. (1)) [34]. The faces are then rotated accordingly using the OpenCV library. The a is the angle of the two eyes. Meanwhile, (x_0, y_0) is location of the left eye in the image, and (x_1, y_1) is location of the right eye in the image.

$$a = arctan\frac{|y_1 - y_0|}{|x_1 - x_0|}. \quad (1)$$

C. Gender Recognition

Since the research emphasises on real-time detection, latency is critical for both gender and age recognition. High latency may degrade the user experience, making the personalised advertisement ineffective. Therefore, the classifier must produce high accuracy but also be lightweight enough to run effectively without the need for heavy computation. SmallerVGGNet and ResNet-34 are selected because they offer strong balance between simplicity and performance [27, 28]. SmallerVGGNet's architecture has the capability to reduce the number of parameters from their original VGG models while also maintaining good feature extractions [27]. While ResNet-34, employs residual connections that enable efficient training, which can maintain high accuracy without the need of excessive computational cost. The pre-trained SmallerVGGNet, trained on 2,200 face images, achieves a validation accuracy of 85%. The network consists of over 8.6 million parameters, including five convolutional layers, seven activation layers, pooling layers, and fully connected layers (see Fig. 3).

On the other hand, the researchers utilize a pre-trained ResNet-34 from previous research [29]. It has been trained on 108,000 face images evenly distributed between males and females. This model is implemented using ModelScope on Alibaba Cloud. ResNet-34 consists of 34 layers, primarily convolutional layers

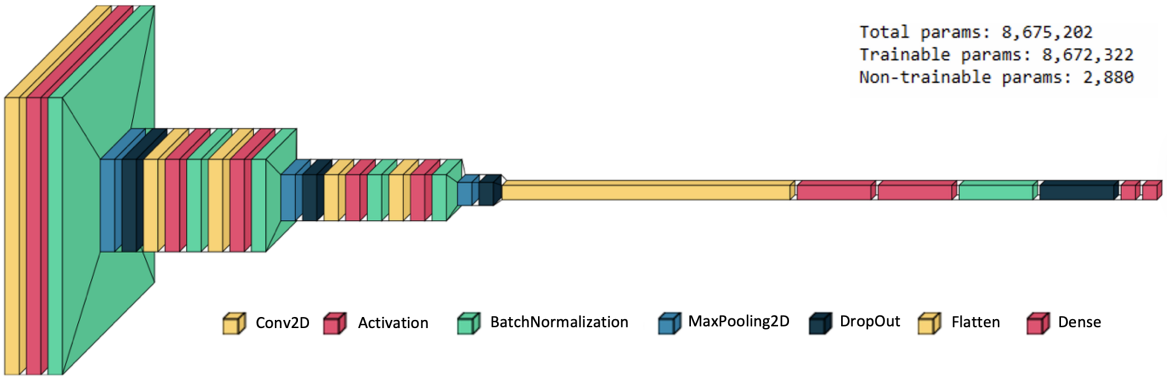


Fig. 3. SmallerVGGNet architecture.

with 3×3 filters, pooling layers, and fully connected layers. It contains nearly 21.8 million parameters, which is more than 2.5 times larger than SmallerVGGNet.

To train the gender recognition models, the researchers combine the MAFA, UTKFace, CelebA, and artificially masked CelebA datasets, filtering out unclear images. The dataset is split into training (80%), testing (10%), and validation (10%). Faces are cropped and aligned using the method described previously. Both SmallerVGGNet and ResNet-34 are fine-tuned using this dataset with a learning rate of 0.01 over 100 epochs.

D. Age Recognition

Age recognition in the SAVER Board is treated as a categorical problem rather than a numerical one to increase detection performance [35]. There are two CNN architectures that are evaluated in the research, including ResNet-34 and FP-Age. ResNet-34 is chosen due to its smaller architecture design and high accuracy in classification tasks. Meanwhile, previous research has stated that FP-Age performs well in estimating age based on facial attributes [33].

The architecture for ResNet-34 in age recognition is similar to that for gender detection [29]. The primary difference lies in the fully connected layers, which output eight distinct results for age-group recognition and two results for gender recognition. The model is also developed using ModelScope from Alibaba Cloud. Meanwhile, FP-Age is a CNN-based model with additional layers of an attention module [36]. The attention module is used to assign greater importance to facial attributes considered relevant, enabling the model to focus more effectively on these attributes. In previous research, the RoI-Tanh Polar Wrap has been used as the preprocessing method [37]. However, this method requires heavy computational power, which reduces the frame rate and poses a problem for real-time detection. In this research, the RoI-Tanh Polar Wrap is replaced with YOLO-X and arctangent face alignment. The output of FP-Age is the prediction of the actual age, which is then grouped into five age categories (3–9, 10–19, 20–39, 40–59, $>$ 59). Since individuals of similar ages often share advertising interests, age grouping is applied. Moreover, this approach also increases detection accuracy.

E. SAVER Board Prototype

The SAVER Board Prototype employs a webcam to capture images, a CPU to run detection, and a monitor to display advertisements. In real-world applications, there may be groups of people with different age groups and genders. People need time to read the advertisement. The effective duration of an advertisement is approximately five to six seconds [38]. Therefore, researchers apply at least five-second intervals before advertisements change. It is implemented using multi-threading, as shown in Fig. 4. When the system detects a face, two threads are run simultaneously. The first thread detects age and gender and stores the results in lists, while the second thread starts a five-second timer. After five seconds, the system selects the mode of age group and gender from the lists, which determines the advertisement displayed.

Furthermore, this method can increase the accuracy of age and gender recognition, as the outputs of age and gender are determined by the mode of age and gender classification in the duration of five seconds. The advertisements displayed follow the Advertisement Decision Board in Table II. For instance, if the system detects a male teenager, the SAVER Board will display advertisements related to online games. Furthermore, if the system detects adults with children

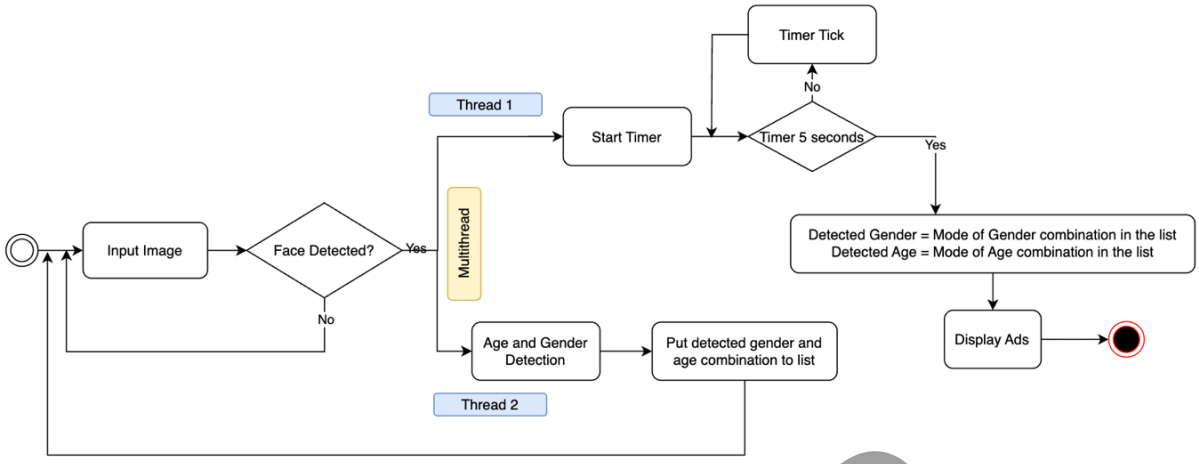


Fig. 4. Advertisement thread flowchart.

TABLE II
ADVERTISEMENT DECISION BOARD.

Advertisement	Children (3–9 years)		Teenagers (10–19 years)		Youth (20–39 years)		Adult (40–59 years)		Senior (>59 years)	
	F	M	F	M	F	M	F	M	F	M
Doll	✓									
Robot		✓								
Fashion			✓							
Online Game				✓						
Cosmetics					✓					
Gym Membership						✓				
Handbag							✓			
Cars								✓		
Calcium Milk									✓	
Film				✓						
Real Estate						✓				
School Equipment	✓			✓						
Family Holiday Package	✓			✓		✓		✓		

or teenagers, family holiday package advertisements will be displayed.

F. Evaluation Matrix

The face detection, age, and gender recognition models are compared using Precision, Recall, F1-Score, and Accuracy metrics (Eqs. (2)–(5)). True Positive, True Negative, False Positive, False Negative, and Total Samples are represented as TP, TN, FP, FN, and TS, respectively. Since the dataset used is significantly unbalanced, accuracy alone can be misleading. With accuracy, a model may seem to achieve high accuracy while entirely failing to detect the minority class. Therefore, the researchers rely on F1-Score, which is the harmonic mean of precision and recall, to a more balanced performance on the minority class [39].

$$\text{Accuracy} = (TP + TN)/TS, \tag{2}$$

$$\text{Precision} = TP/(TP + FP), \tag{3}$$

$$\text{Recall} = TP/(TP + FN), \tag{4}$$

$$\text{F1-Score} = 2 \times \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{5}$$

III. RESULTS AND DISCUSSION

In this section, the performance of face detection, face alignment, and face recognition is evaluated and compared to determine the best model for the SAVER board prototype. All models are trained and tested on a server with an Intel CPU, an NVIDIA GeForce GTX 1050 Ti GPU and 16 GB of RAM. By reporting results on this hardware, the experiments approximate the performance constraint that the SAVER board may face in a real-world scenario.

TABLE III
YOLO-X PERFORMANCE IN DETECTING MASKED AND NORMAL FACES.

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Inference Time (ms)
93.9	99.6	88.1	93.5	302

ACTUAL		PREDICTION	
	No Mask	1192	4
	Mask	141	1047
		No Mask	Mask

Fig. 5. Confusion matrix of YOLO-X detection.

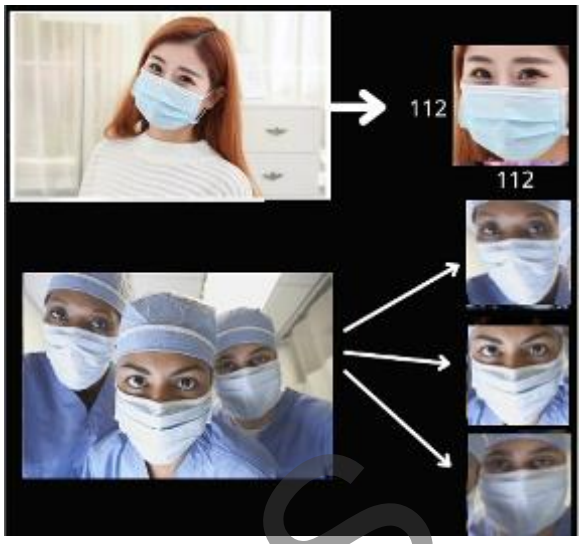


Fig. 6. Face alignment results.

TABLE IV
DETECTION PERFORMANCE WITH AND WITHOUT FACE ALIGNMENT.

Condition	Accuracy (%)	Average Inference Time (ms)
With Face Alignment	80	102.7
Without Face Alignment	60	103.2

A. Face Detection

The training procedure follows the previous research [25] but is retrained using a larger dataset, as stated previously. Based on the performance results in Table III and the confusion matrix in Fig. 5, YOLO-X can detect faces with and without masks, with 93.9% accuracy. The average inference time to detect a face is 302 milliseconds. Moreover, the landmark loss for the model is 0.102. These results indicate that the model maintains high detection accuracy even in detecting faces without masks, which are known to degrade performance in face detection. The inference time of 302 ms demonstrates that YOLO-X is efficient enough for near real-time time application. A low landmark loss further suggests that the model is accurately localising key facial points, which is critical for downstream tasks, such as alignment and recognition. Overall, YOLO-X is a promising candidate for deployment on the SAVER board.

B. Face Alignment

An example of face alignment is shown in Fig. 6. If more than one face is detected in a picture, the system will crop each face and align it separately. The significance of face alignment is also evaluated by comparing the detection performance with and without face alignment. The parameters assessed are inference time and the accuracy of detecting age and gender for 10 different subjects with obliquely angled faces. The result is considered a failure if the system incorrectly detects age or gender.

The researchers also conduct a test to determine how significant the face alignment is in detecting age and gender from face images. Here, the researchers test it using 20 different images with multiple faces to determine their age and gender. The accuracy of the age and gender detection is calculated for all the faces in the images. Then, the researchers compare the accuracy and also the average inference time for all the faces when it is aligned and not aligned. Out of 20 different tests, the system with face alignment performs better, as shown in Table IV. It is because the model is trained using aligned pictures. Hence, it can detect age and gender more accurately. There is no significant difference in detection time between with or without face alignment.

C. Gender Recognition

The evaluation and retraining of the gender recognition are based on ResNet-34 and SmallerVGGNet. The confusion matrix in Fig. 7 shows the gender detection results for both unmasked and masked faces with ResNet-34. The pre-trained ResNet-34 model is trained on a dataset comprising more than 100,000 faces, with a balanced representation of faces across classes. The average inference time for detecting gender using ResNet-34 is 25.21 ms for unmasked faces and 26.07 ms for masked faces.

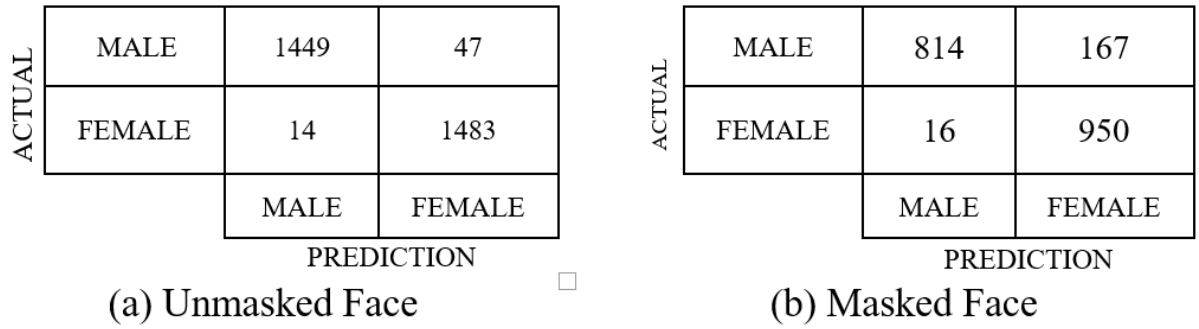


Fig. 7. Confusion matrix using ResNet-34.

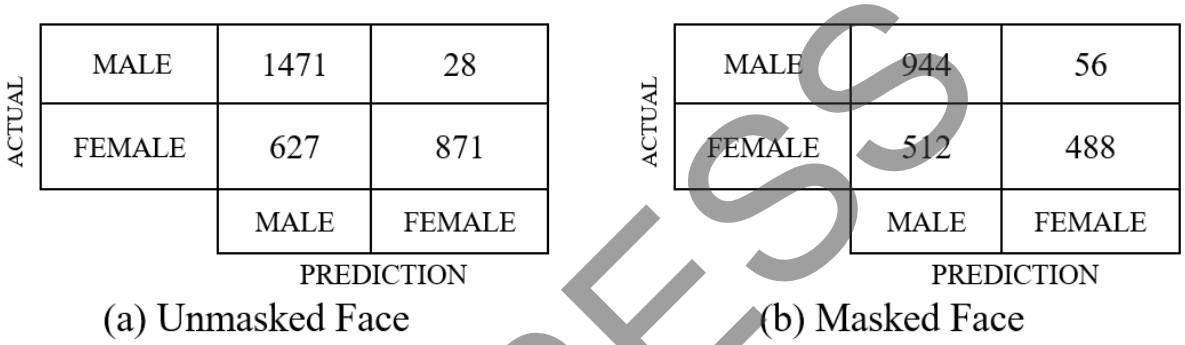


Fig. 8. Confusion matrix using SmallerVGGNet.

TABLE V
COMPARISON RESULTS OF GENDER DETECTION MODEL.

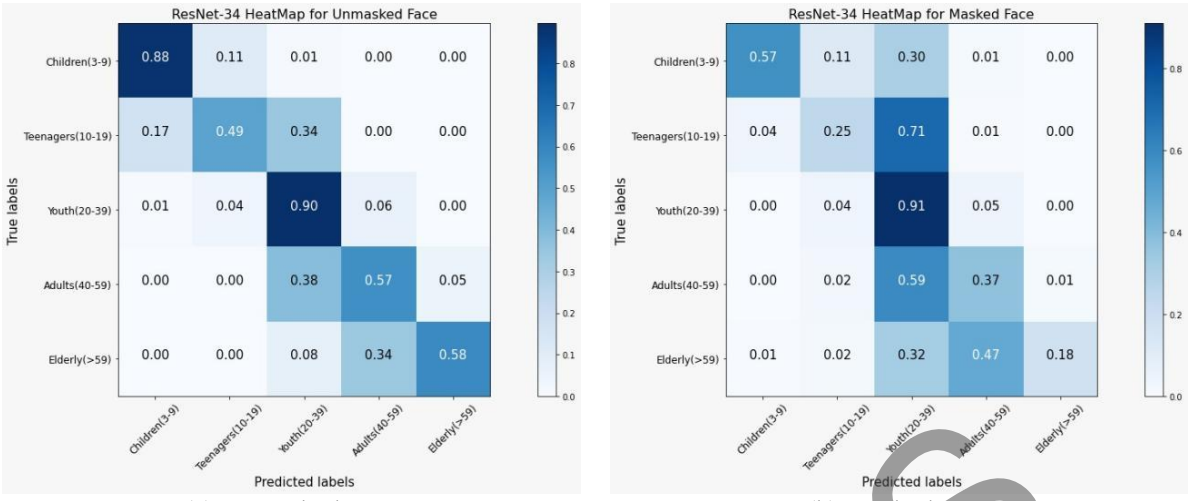
Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Inference Time (ms)
Gender Detection for Unmasked Face					
SmallerVGGNet	78.1	**96.9**	58.1	72.6	**3.31**
ResNet-34	**98.0**	**96.9**	**99.0**	**97.9**	25.21
Gender Detection for Masked Face					
SmallerVGGNet	71.6	**89.79**	48.8	63.2	**3.25**
ResNet-34	**90.6**	85.00	**98.3**	**91.2**	26.06

Meanwhile, for SmallerVGGNet, training requires an average of five minutes per epoch, with a total of 100 epochs. Figure 8 shows that the model is more likely to predict faces as male rather than female for both masked and unmasked faces, as indicated by the number of False Negatives. This bias results from the imbalanced dataset used for retraining, which contains more male faces than female. The inference time is 3.31 milliseconds for unmasked faces and 3.25 milliseconds for masked faces.

D. Gender Detection Model Comparison

Based on Table V, ResNet-34 achieves better Accuracy, Recall, and F1-Score compared to the SmallerVGGNet model, with differences of 19.9%, 40.9%, and 25.3% for unmasked faces and 19.0%, 49.5%, and 28.0% for masked faces, respectively. On the other hand, SmallerVGGNet can detect faces approximately eight times faster than ResNet-34. It is because SmallerVGGNet has more than 2.5 times fewer parameters in total compared to ResNet-34. The number of parameters also impacts the performance of detection, as more parameters result in more trainable layers. The dataset used to train the pre-trained ResNet-34 model is also larger than that used for SmallerVGGNet, with faces distributed equally in each class.

According to Table V, a masked face significantly affects both gender detection models. The bold number



(a) Unmasked Face (b) Masked Face

Fig. 9. Heatmap using ResNet-34.

shows the best performance model for the metrics. On average, ResNet-34 performs 6.78% worse when detecting a masked face compared to an unmasked face. Moreover, the performance of SmallerVGGNet decreases by an average of 13.68% when applied to masked faces. Then, the inference time for detecting masked and unmasked faces has an insignificant difference. The downside of using the ResNet-34 model is that it requires more resources to perform the preprocessing method and prepare the model. Although SmallerVGGNet is faster for gender detection, ResNet-34 shows superior performance in the gender detection model for the SAVER Board due to its higher Accuracy, Recall, and F1-Score. An average of 25.64 milliseconds is still acceptable for real-time detection, which typically takes around 6–7 seconds.

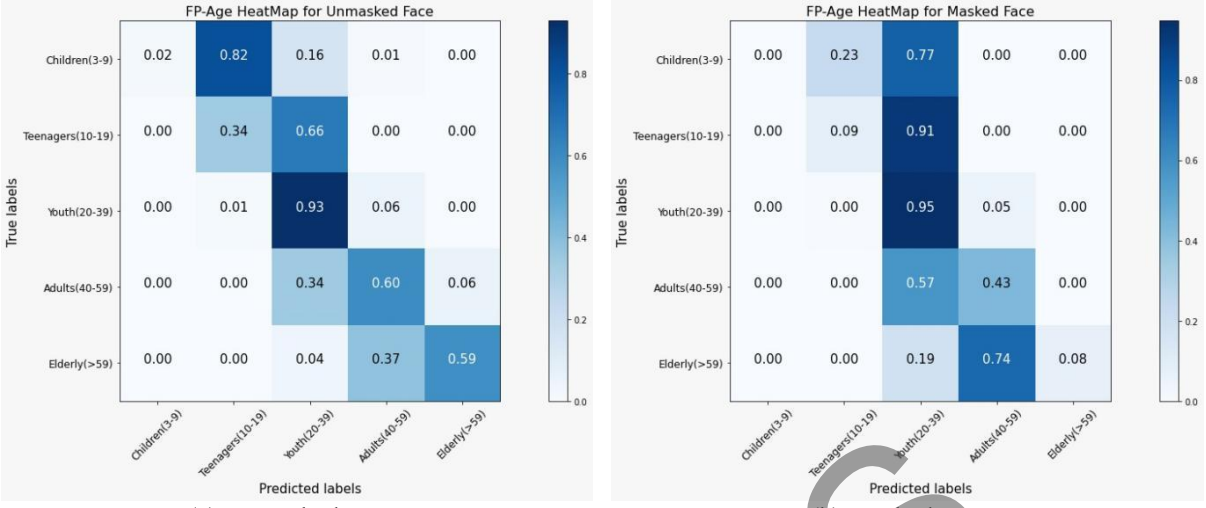
E. Age Recognition

The models evaluated for age recognition are ResNet-34 and FP-Age. Both models are tested on masked and unmasked face images across several age groups. ResNet-34 can determine the age group for both masked and unmasked faces, with accuracies of 75.6% and 64.0%, respectively. As shown in Fig. 9a, the model classifies 34% of teenagers into the youth age group. Misclassification mainly occurs for faces aged 17 to 19 years old, which are close to the age group threshold. Moreover, some male teenagers already have dense facial hair, a feature typically associated with older age groups. In age recognition with a masked face (Fig. 9b), the model’s accuracy drops by 7.4%, since significant facial features, including lower facial hair, are covered by the mask. The average detection time for age recognition using ResNet-34 is 25.54 ms for unmasked faces and 27.07 ms for masked faces.

Then, the average inference time for age recognition using FP-Age is 432 ms for unmasked faces and 436 ms for masked faces, with weighted average accuracies of 72.1% and 60.5%, respectively. Figure 10 shows that the model struggles to classify children’s faces, as they tend to be classified into the teenager age group. It occurs because the pre-trained model uses only a small number of children’s faces for training. When the face is masked, the model’s accuracy drops by a substantial 11.6%. It is because the attention module of FP-Age relies heavily on the nose and mouth coordinates, which are covered by masks [36].

F. Age Detection Performance Comparison

The age recognition model with the best Accuracy, Precision, Recall, and F1-Score is ResNet-34, as shown in Table VI (the best score of each metric is presented in bold). The pre-trained ResNet-34 model used in the research is trained on more than 100,000 faces with an equally distributed number of samples in each class. Due to the limited dataset available, FP-Age cannot compete with ResNet-34. FP-Age also has the longest inference time for detecting age groups, because it predicts the actual age first, which requires passing through numerous layers. It is clear that masks affect age detection, with an average performance decrease of 8.83% and 6.75% for ResNet-34 and FP-Age, respectively. Based on the result, the researchers choose



(a) Unmasked Face (b) Masked Face

Fig. 10. Heatmap using FP-Age.

TABLE VI
COMPARISON RESULTS OF AGE DETECTION MODEL.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Inference Time (ms)
Age Detection for Unmasked Face					
ResNet-34	**75.9**	78.0	**75.9**	**76.52**	**25.21**
FP-Age	72.1	**78.4**	72.1	74.60	432
Age Detection for Masked Face					
ResNet-34	**64.0**	75.4	**64.0**	67.6	**25.21**
FP-Age	60.5	**80.9**	60.5	**68.3**	436

ResNet-34 as the age recognition model for SAVER Board. The model achieves the best performance in terms of Accuracy, Precision, Recall, and F1-Score, as well as the fastest average inference time.

G. Prototype Testing

The SAVER Board prototype was tested using 224 scenarios. A condition is considered correct if the system displays the appropriate advertisement, in other words, if it correctly classifies both the user's age and gender. The testing dataset consists of 30 videos, 23 real-time demos (with actual people), and 171 real-time demos using printed faces. The scenarios include 60% male and 40% female subjects, with 35% wearing masks and 65% without masks.

Figure 11 shows the system's detection results for the entire evaluation. By using a multithreading method and the ResNet-34 model, the system achieves 68% accuracy in detecting 13 different classes with an inference time of 50 ms. The system demonstrates accurate detection across all test cases. This performance suggests that while the model can handle multi-class classification efficiently, it is also able to maintain low latency suitable for real-time applications. However, the age recognition task remains a more challenging problem due to the small intra-class differences and overlapping facial features among adjacent age ranges. Misclassification in age groups is likely caused by dataset imbalance and the difficulty of learning caused by the similarity of features for similar age brackets. Nevertheless, the fast inference time highlights the potential for on-device deployment.

Then, Fig. A1 in Appendix illustrates various scenarios that are captured by the SAVER Board. Each picture shows the webcam camera on the left and its corresponding advertisement on the right. For evaluation purposes, each detected faces are annotated with its bounding box (border with red line), and the five facial landmarks are marked. Additionally, the confidence of the age and gender detection is displayed on the top left. This result allows the researchers to evaluate the performance of the system in detecting facial landmarks and recognizing the audience's age and gender in a real-world scenario.

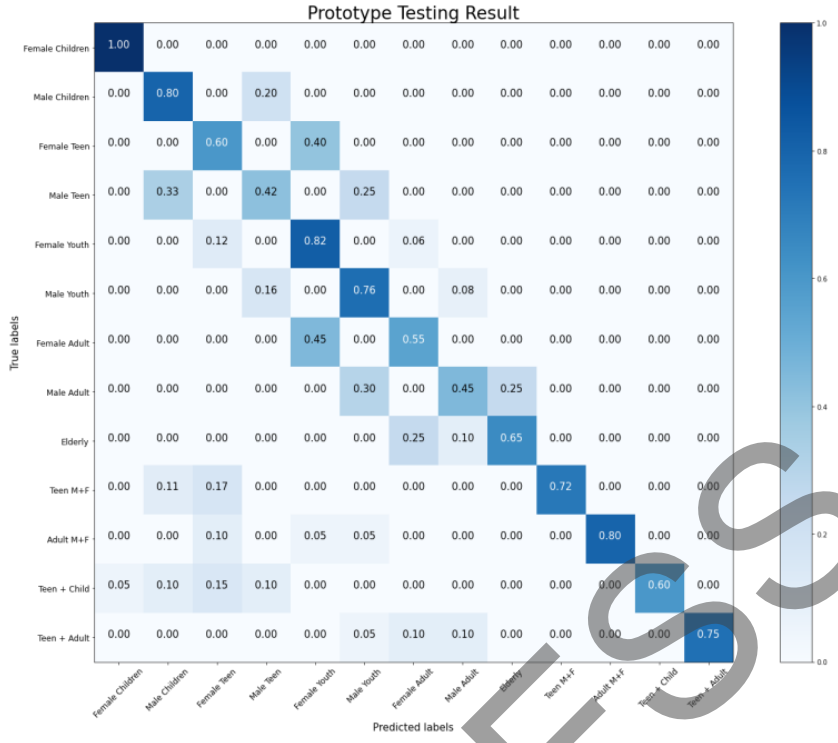


Fig. 11. Heatmap of prototype testing.

IV. Conclusion

The research addresses the challenge of delivering personalized advertisement systems in public spaces using real-time demographic recognition, especially under masked conditions. The advertisement system first captures the user’s picture in front of the board. Then, the picture goes through the YOLO-X model for face detection. Once faces are detected, they are cropped based on their bounding boxes and then aligned. The aligned faces then go through the age and gender recognition model, which classifies the faces into age groups and genders. The age and gender combinations then determine the advertisement the board will display. The research contributes to the field of computer vision by integrating YOLO-X and CNN for masked and unmasked face recognition in a smart advertising system. The system achieves 68% accuracy in displaying 13 different types of advertisements.

The findings demonstrate that ResNet-34 is the most effective CNN model for age and gender recognition under both masked and unmasked conditions, supporting the feasibility of the SAVER Board system in real-time public applications. The results highlight that balancing the model complexity and inference speed is critical for real world deployment, especially when running on resource constrained embedded hardware. For future development, face spoofing prevention has the potential to transform public advertising by making it more adaptive, responsive, and user aware.

The research faces several limitations. First, the datasets are primarily composed of non-Asian ethnicities, resulting in a lack of representation of Asian facial attributes. This underrepresentation contributes to biased outputs and less reliable demographic profiling. Using more diverse and ethical datasets for training will improve system performance. Second, technological constraints also affect system performance and reliability. The system often captures blurred or unclear images, reducing the model’s ability to recognize audience age and gender accurately. Incorporating infrared imaging for low-light detection can increase the system’s reliability in such conditions.

Author Contribution

Designed algorithm for SAVER board, H; Prepared and performed analysis of the data, H; Wrote research method for SAVER board and abstract, H; Designed the algorithm and wrote the software for age and gender detection, A. B. H.; Run the experiment and collected all data, A. B. H.; Contributed to the analysis of the data and created tables, A. B. H.; Wrote the research methods, result and discusion sections, A. B.

H.; Contributed to the introduction, related work and conclusion parts, D. U.; and Compared some age and gender algorithms, D. U.

Data Availability

The data that support the findings of the research are available at https://doi.org/10.48550/arXiv.1411.7766 and https://doi.org/10.48550/arXiv.1511.06523. These data were derived from the following resources available in the public domain: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html and http://shuoyang1213.me/WIDERFACE/

References

[1] S. Rosengren, M. Eisend, S. Koslow, and M. Dahlen, "A meta-analysis of when and how advertising creativity works," *Journal of Marketing*, vol. 84, no. 6, pp. 39–56, 2020.

[2] J. L. Hayes, G. Golan, B. Britt, and J. Applequist, "How advertising relevance and consumer—Brand relationship strength limit disclosure effects of native ads on Twitter," *International Journal of Advertising*, vol. 39, no. 1, pp. 131–165, 2020.

[3] L. V. Shulga, J. A. Busser, B. Bai, and H. Kim, "Branding co-creation with consumer-generated advertising: Effect on creators and observers," *Journal of Advertising*, vol. 52, no. 1, pp. 5–23, 2023.

[4] A. Kallevig, W. Zuem, M. Willis, S. Ranfagni, and S. Rovai, "Managing creativity in the age of data-driven marketing communication: A model for agencies to improve their distribution and valuation of creativity," *Journal of Advertising Research*, vol. 62, no. 4, pp. 301–320, 2022.

[5] S. Chen, Y. Wu, F. Deng, and K. Zhi, "How does ad relevance affect consumers' attitudes toward personalized advertisements and social media platforms? The role of information co-ownership, vulnerability, and privacy cynicism," *Journal of Retailing and Consumer Services*, vol. 73, 2023.

[6] M. K. Benkaddour, S. Lahlali, and M. Trabelsi, "Human age and gender classification using convolutional neural network," in *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH)*. Boumerdes, Algeria: IEEE, Feb. 9–10, 2021, pp. 215–220.

[7] M. Alhalabi, N. Hussein, E. Khan, O. Habash, J. Yousaf, and M. Ghazal, "Sustainable smart advertisement display using deep age and gender recognition," in *2021 International Conference on Decision Aid Sciences and Application (DASA)*. Sakheer, Bahrain: IEEE, Dec. 7–8, 2021, pp. 33–37.

[8] O. Singh and K. Mourya, "Gender and age detection using machine learning algorithm," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 3, pp. e32–e–35, 2024.

[9] D. K. Srivastava, E. Gupta, S. Shrivastav, and R. Sharma, "Detection of age and gender from facial images using CNN," in *Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2022*. Telangana, India: Springer, March 28–29, 2023, pp. 481–491.

[10] S. Naaz, H. Pandey, and C. Lakshmi, "Deep learning based age and gender detection using facial images," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. Chennai, India: IEEE, May 9–10, 2024, pp. 1–11.

[11] F. Murtadho, D. W. Sudiharto, C. W. Wijiutomo, and E. Ariyanto, "Design and implementation of smart advertisement display board prototype," in *2019 International Seminar on Application for Technology of Information and Communication (ISemantic)*. Semarang, Indonesia: IEEE, Sep. 21–22, 2019, pp. 246–250.

[12] D. Sundawa, D. S. Logayah, and R. A. Hardiyanti, "New normal in the era of pandemic COVID-19 in forming responsibility social life and culture of Indonesian society," in *IOP Conference Series: Earth and Environmental Science*, vol. 747. East Java, Indonesia: IOP Publishing, Sep. 12, 2021, pp. 1–10.

[13] Y. Feng, S. Yu, H. Peng, Y. R. Li, and J. Zhang, "Detect faces efficiently: A survey and evaluations," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 1–18, 2021.

[14] D. Qi, W. Tan, Q. Yao, and J. Liu, "YOLO5Face: Why reinventing a face detector," in *European Conference on Computer Vision*. Tel Aviv, Israel: Springer, Oct. 23–27, 2022, pp. 228–244.

[15] B. Wang, J. Zheng, and C. L. P. Chen, "A survey on masked facial detection methods and datasets for fighting against COVID-19," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 3, pp. 323–343, 2021.

[16] D. Fitousi, N. Rotschild, C. Pnini, and O. Azizi, "Understanding the impact of face masks on the processing of facial identity, emotion, age, and gender," *Frontiers in Psychology*, vol. 12, pp. 1–13, 2021.

[17] R. Liu and Z. Ren, "Application of YOLO on

mask detection task," in *2021 IEEE 13th International Conference on Computer Research and Development (ICCRD)*. Beijing, China: IEEE, Jan. 5–7, 2021, pp. 130–136.

[18] S. Abbasi, H. Abdi, and A. Ahmadi, "A face-mask detection approach based on YOLO applied for a new collected dataset," in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*. Tehran, Iran: IEEE, March 3–4, 2021, pp. 1–6.

[19] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 19 753–19 768, 2021.

[20] S. A. Sanjaya and S. A. Rakhmawan, "Face mask detection using MobileNetV2 in the era of COVID-19 pandemic," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*. Sakheer, Bahrain: IEEE, Oct. 26–27, 2020, pp. 1–5.

[21] X. Fan and M. Jiang, "RetinaFaceMask: A single stage face mask detector for assisting control of the COVID-19 pandemic," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Melbourne, Australia: IEEE, Oct. 17–20, 2021, pp. 832–837.

[22] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021. [Online]. Available: https://arxiv.org/abs/2107.08430

[23] T. Panboonyuen, S. Thongbai, W. Wongweeranimit, P. Santitamnont, K. Suphan, and C. Charoenphon, "Object detection of road assets using transformer-based YOLOX with feature pyramid decoder on Thai highway panorama," *Information*, vol. 13, no. 1, pp. 1–12, 2021.

[24] M. Wu, L. Guo, R. Chen, W. Du, J. Wang, M. Liu, X. Kong, and J. Tang, "Improved YOLOX foreign object detection algorithm for transmission lines," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–10, 2022.

[25] A. B. Handoko, V. C. Putra, I. Setyawan, D. Utomo, J. Lee, and I. K. Timotius, "Evaluation of YOLO-X and MobileNetV2 as face mask detection algorithms," in *2022 IEEE Industrial Electronics and Applications Conference (IEACon)*. Kuala Lumpur, Malaysia: IEEE, Oct. 3–4, 2022, pp. 105–110.

[26] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," 2020. [Online]. Available: https://arxiv.org/abs/2009.07485

[27] A. Saxena, P. Singh, and S. N. Singh, "Gender and age detection using deep learning," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. Noida, India: IEEE, Jan. 28–29, 2021, pp. 719–724.

[28] F. Alonso-Fernandez, K. Hernandez-Diaz, S. Ramis, F. J. Perales, and J. Bigun, "Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images," *IET Biometrics*, vol. 10, no. 5, pp. 562–580, 2021.

[29] K. Karkkainen and J. Joo, "Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Virtual, Jan. 5–9, 2021, pp. 1548–1558.

[30] R. Golwalkar and N. Mehendale, "Age detection with face mask using deep learning and FaceMaskNet-9," 2020. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3733784

[31] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with LLE-CNNs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2682–2690.

[32] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5525–5533.

[33] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," 2020. [Online]. Available: https://arxiv.org/abs/2008.11104

[34] M. F. Karaaba, O. Surinta, L. R. B. Schomaker, and M. A. Wiering, "In-plane rotational alignment of faces by eye and eye-pair detection," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP-2015)*, 2015, pp. 392–399.

[35] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Gait-based human age estimation using age group-dependent manifold learning and regression," *Multimedia tools and applications*, vol. 77, no. 21, pp. 28 333–28 354, 2018.

[36] Y. Lin, J. Shen, Y. Wang, and M. Pantic, "FP-age: Leveraging face parsing attention for facial age estimation in the wild," *IEEE Transactions on Image Processing*, vol. 34, pp. 4767–4777, 2022.

[37] ——, "RoI Tanh-polar transformer network for

face parsing in the wild," *Image and Vision Computing*, vol. 112, 2021.

[38] D. G. Goldstein, R. P. McAfee, and S. Suri, "The effects of exposure time on memory of display advertisements," in *Proceedings of the 12th ACM conference on Electronic commerce*. California, USA: Association for Computing Machinery, June 5–9, 2011, pp. 49–58.

[39] M. H. Nguyen, "Impacts of unbalanced test data on the evaluation of classification methods," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, pp. 497–502, 2019.

APPENDIX

The Appendix can be seen in the next page.

(a) Boy & Robot Cartoon (b) Girl & Doll

(c) Male Youth & Cars (d) Female Youth & Cosmetic

(e) Male and Female Senior & Calcium Milk (f) Family & Holiday Package

Fig. A1. SAVER Board prototype results.

Figure 12. SAVER Board prototype results.

IV. CONCLUSION

The research addresses the challenge of delivering personalized advertisement systems in public spaces using real-time demographic recognition, especially under masked conditions. The advertisement system first captures the user's picture in front of the board. Then, the picture goes through the YOLO-X model for face detection. Once faces are detected, they are cropped based on their bounding boxes and then aligned. The aligned faces then go through the age and gender recognition model, which classifies the faces into age groups and genders. The age and gender combinations then determine the advertisement the board will display. The research contributes to the field of computer vision by integrating YOLO-X and CNN for masked and unmasked face recognition in a smart advertising system. The system achieves 68% accuracy in displaying 13 different types of advertisements.

The findings demonstrate that ResNet-34 is the most effective CNN model for age and gender recognition under both masked and unmasked conditions, supporting the feasibility of the SAVER