Human Gender Identification Employing Convolution Neural Networks for Veiled Face Images

Waled H. Al-Arashi (1,*)
Makram Almaqtobi (2)
Abdulrahman Q. Al-fahidi (3)
Abdulghaffar M. Ali (4)
Abdulaziz Aborujilah (5)

Received: 02/10/2024
Revised: 11/10/2024
Accepted: 7/12/2024

[1] University of Science and Technology, Aden, Yemen.
[2,3,4] Electronic department, Faculty of Engineering, University of Science and Technology, Yemen.
[5] Information Systems department, college of Commerce and Business Administration, Dhofar university, Sultanan of Oman
* Corresponding Author Designation, Email: wal112002@yahoo.com

https://doi.org/10.20428/jst.v30i1.2486

Human Gender Identification Employing Convolution Neural Networks for Veiled Face Image

Waled H. Al-Arashi
University of Science and Technology
Aden, Yemen
wal112005@yahoo.com

Makram Almaqtobi
University of Science and Technology
Yemen
almaqtobi@gmail.com

Abdulrahman Q. Al-fahidi
University of Science and Technology
Yemen
abdualfohide@gmail.com

Abdulghaffar M. Ali
University of Science and Technology
Yemen
abdulghaffarabdulwasea@gmail.com

Abdulaziz Aborujilah
Dhofar university
Sultanan of Oman
abdulghaffarabdulwasea@gmail.com

***Abstract*—** The interest in biometrics and pattern recognition has significantly risen over the past few decades. Gender classification through facial analysis can greatly enhance the process of human identification, distinguishing between male and female individuals. . The study of biometric recognition, which centers on distinct facial attributes like the eyes, is a noteworthy and extensive domain of research, particularly concerning security and societal perspectives in Yemen and various Muslim nations.. This study seeks to advance the field of gender identification by introducing a novel model that leverages the capabilities of VGG19 and EfficientNetB7. Experimental results obtained from a veiled faces database indicate that the proposed model outperforms its counterparts, achieving an impressive accuracy rate of 99.1%, in contrast to VGG19's 97.3% and EfficientNetB7's 95.5%.

Keywords—Gender Recognition, Veiled Face Images, Convolution Neural Network, VGG19, EffecientNetB7.

I. INTRODUCTION

The field of pattern recognition has witnessed a remarkable surge of interest over the past decades, due to the increasing number of emerging applications that require efficient techniques for extracting, describing, and classifying patterns[1-2]. This increasing demand has led to an increased need for sophisticated theory to support the design of pattern recognition applications. The range of these applications spans multiple fields, necessitating the development of efficient and scalable algorithms for pattern recognition [2-3]. However, the field of automatic gender classification is garnering heightened interest, as gender provides valuable and distinct insights into the social activities of males and females [3-5]. Besides, gender recognition plays a significant role in enhancing human-robot interactions, as it significantly contributes to a more fulfilling user experience [6-7]. In addition, the prevalence of automatic gender recognition is increasingly associated with its application across various programs and devices, particularly due to the rise of social networking platforms and social media on the Internet [8]. Thus, precise gender recognition can enhance user experiences, and system security, and offer valuable insights into social interactions. Consequently, various algorithms have been introduced to tackle gender recognition from different angles. [9-12]. Nevertheless, , the application of face recognition and/or face features represents the most promising avenue within the field [13-15].

Nowadays, identifying a person's gender from facial data is likewise a difficult task, considering the rapidity of human metamorphosis. Aspects such as facial injuries, the emergence of facial hair, and the consequences of aging may contribute to a quick reduction in the effectiveness of even the most advanced therapeutic interventions [6]. In addition, additional factors including pose orientation and light also affect the system's accuracy. To cope with those details and attain dependable results, various methodologies and models have been formulated by researchers and have been incorporated into the body of literature. The conventional method utilized in face recognition, including facial gender discrimination, generally entails the stages of image capture, image processing, dimensionality reduction, and feature extraction [1, 5- 6, 16]. Furthermore, the effectiveness of the feature extraction process is dependent on a deep understanding of the specific application domain. Consequently, the choice of classifier, which is influenced by the feature extraction method employed, plays a crucial role in the effectiveness of the recognition system. Identifying a classifier that aligns effectively with the chosen feature extractor to achieve optimal classification performance presents a considerable challenge. Any alterations to the problem domain necessitate a comprehensive redesign of the system [5, 16-17]. Recently, the application of deep learning in the study of human gender has recently emerged as a crucial area of research, yielding superior outcomes with comparison to conventional techniques. [12, 18-19]. Many works use deep learning using the face or eye [18-20].

Due to the nature of Yemeni and Islamic society, attention has shifted towards the eyes, as veiled women do not permit the capture of their facial images.

Therefore, this study relies on the application of deep learning techniques using eyes feature of the face. In [21], they introduce an innovative feature extraction method that focuses exclusively on the eye and eyebrow regions. The feature extraction techniques used are 2D-Wavelet Transform, Gray Level Co-occurrence Matrix, or Discrete Cosine Transform. Support Vector Machine (SVM) is utilized to obtain the results. The suggested approach achieved an accuracy rate of 99.49% for gender recognition utilizing 2D-Wavelet Transform, an accuracy rate of 98.49% with GLCM, and an accuracy rate of 99.62% employing DCT on the Faces94 database. These traditional approaches necessitate a full facial image for accurate classification, which can pose challenges in certain situations, particularly when the face is partially hidden as in our society. Additionally, contemporary feature extraction methods for determining facial gender are resource-intensive in terms of computation. Thus, deep learning is used instead. In [20], to get face features, the authors propose the creation of a set of sophisticated feature representations through the application of deep learning methodologies, referred to as Deep Hidden Identity features (DeepID). The features of DeepID are derived from the activations of the neurons in the final hidden layer of deep convolutional networks (ConvNets). A remarkable drawback of DeepID is its expanded complexity in comparison to other face recognition techniques, as it requires the training of distinct classifiers for each individual classification task. Due to, the paper presented in [21] investigates the application of pre-trained deep neural network (DNN) models in scenarios characterized by insufficient data availability. The findings of the research demonstrate that the EfficientNetb7 model achieves the highest performance, with an accuracy of 0.913, surpassing other models including Xception (0.876), InceptionResNetV2 (0.892), VGG16 (0.902), and Resnet50 (0.905). [18] is also used Deep convolutional neural network (CNN) using a pre-trained model VGG19 to extract the features from images of faces obscured by veils. The experimental results record accuracy values of 99.8% for gender recognition.

The subsequent sections of this paper are structured as follows. Section II discusses the most distinguished deep convolutional neural network models, particularly highlighting VGG19 and EfficientNetB7. In Section III, the proposed model is introduced, highlighting its architecture and the differences it exhibits when compared to the VGG19 model. Section IV presents the experiments along with the related discussion. The final section, Section V, provides the conclusion and outlines potential future work.

II. DEEP CONVOLUTIONAL NEURAL NETWORK (CNN) MODELS

A. VGG19 model

This architecture exemplifies a conventional deep Convolutional Neural Network (CNN) structure characterized by multiple layers, with the acronym VGG representing the Visual Geometry Group [23]. The term "deep" refers to the quantity of layers, where VGG-19 comprises 19 convolutional layers. It continues to be one of the most frequently utilized architectures for image recognition in contemporary applications. The essential elements of the VGG-19 architecture include:

1. ***Convolutional Layers****:* Utilization of 3×3 filters with a stride of 1 and padding of 1 to maintain spatial resolution.
2. ***Activation Function****:* The **ReLU** (Rectified Linear Unit) is employed following each convolutional layer to incorporate non-linearity.
3. ***Pooling Layers****:* Implementation of max pooling using a 2×2 filter and a stride of 2 to diminish the spatial dimensions.
4. ***Fully Connected Layers****:* The network concludes with three fully connected layers dedicated to classification.
5. ***Softmax Layer****:* The terminal layer is responsible for generating class probabilities.

The VGG-19 architecture is composed of five convolutional layer blocks, succeeded by three fully connected layers, as illustrated in Figure 1. The Figure also shows the details of each layer.

https://doi.org/10.20428/jst.v30i1.2486

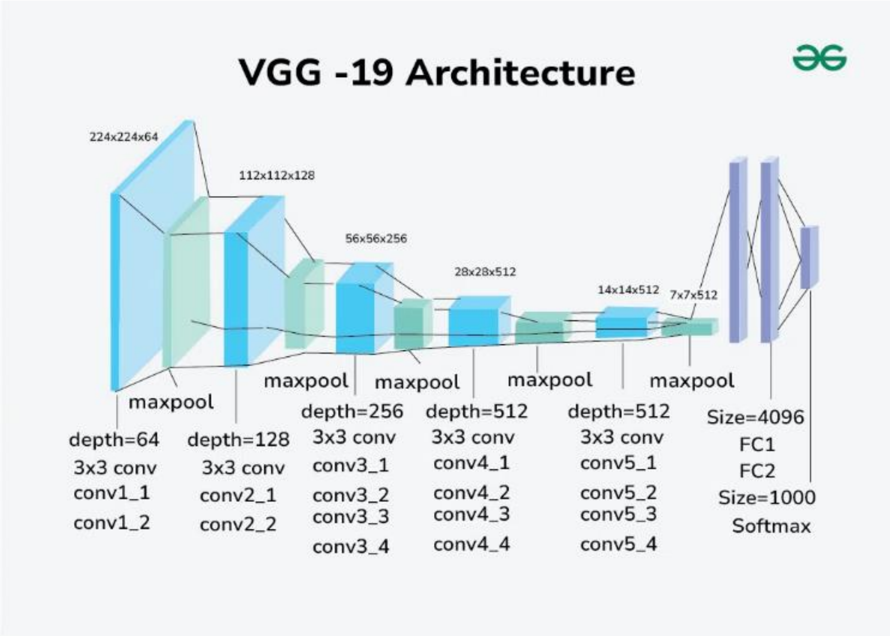

Figure 1: VGG-19 Architecture

B. EfficientNetB7 model

The EfficientNet models [24] are constructed using fundamental and highly effective compounded scaling techniques. This approach enables the scaling of the ConvNet baseline to any desired resource constraints while preserving the model's utility for transferring learning datasets. EfficientNet comprises versions ranging from B0 to B7, with each version featuring a different number of parameters, varying from 5.3 M to 66 M. EfficientNetB7 consists of several layers that work together to collaboratively function to extract features from images and acquire valuable representations for conveying information. These layers consist of:

1. ***Convolutional Layers****:* These layers serve the purpose of extracting fundamental features from images by analyzing their characteristics. EfficientNetB7 incorporates multiple convolutional layers designed to capture various features at different scales.
2. ***Pooling Layers****:* These layers help to decrease the data size and streamline operations. They play a crucial role in minimizing dimensionality while managing the loss of information.
3. ***Fully Connected Layers****:* These layers convert the extracted features into final predictions. In EfficientNetB7, the fully connected layers comprise multiple interconnected layers that translate the features into interpretable predictions.

The EfficientNet7 architecture is composed of seven blocks as Figure 2 illustrates.

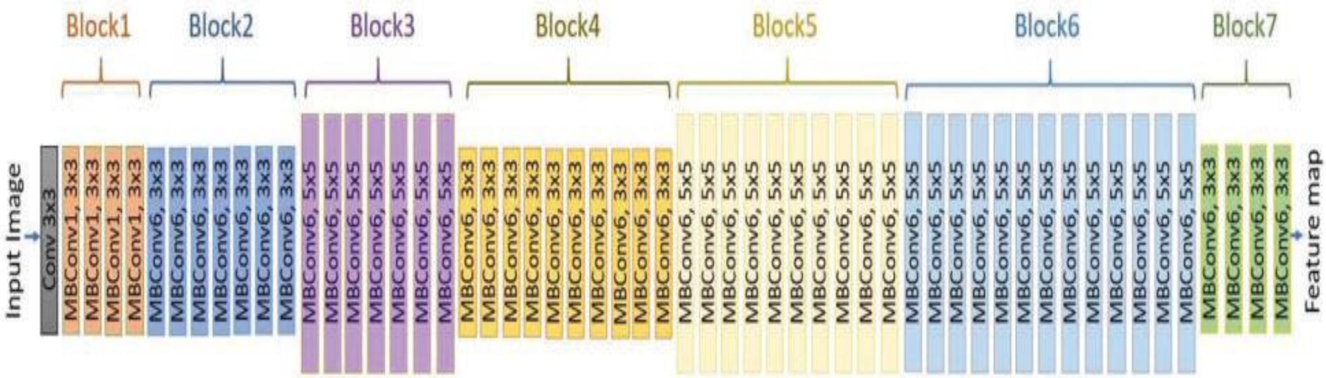

Figure 2: Efficientnetb7 model architecture [25]

The EfficientNet7 method's compound scaling is based on the principle of maintaining equilibrium with a constant value to adjust the ratios of depth, width, and resolution. Where α is the Depth, β is the Width, and γ is the Resolution. The definition is established through the utilization of the grid search algorithm, while Φ denotes the computational resources of the network.

The encoding and decoding layers utilize distinct kernel sizes; specifically, the encoding layers employ a kernel size of 3 with filter sizes of 16, 32, 64, 128, and 256, all maintaining a channel count of 3. Conversely, the decoding layers feature a kernel size of 4 and filter sizes of 256, 128, 64, 32, and 16, also with 3 channels. The increase in the number of channels corresponds to the ascending number of channels across successive spatial resolution levels. The implementation of encoding and decoding blocks enhances the robustness of the model. The total number of slices ranges from 150 to 220.

III. THE PROPOSED MODEL

A. Proposed Architecture

The proposed model architecture is designed to categorize images by employing techniques associated with convolutional neural networks. It is composed of multiple layers that collaboratively function to extract features from images and identify patterns for information representation. It consists of the following layers:

1. ***Convolutional Layers:***
   - Four convolutional layers (Conv2D) layers are used for extracting features from images, and make up the model.
   - The initial layer is equipped with 32 filters, the subsequent layer has 64 filters, while both the third and fourth layers are configured with 128 filters, each having a dimension of 4 * 4.
   - The ReLU activation function is used to activate the output.
   - Following each Convolutional layer, a MaxPooling2D operation with a dimension of (3 * 3) is implemented to reduce the size of the image for subsequent layers.
2. ***Flatten Layers:***
   - It transforms data into a flattened structure, changing it from a two-dimensional to a one-dimensional arrangement.
   - t serves the purpose of preparing data for fully connected layers.
3. ***Fully Connected Layers:***
   - The model architecture comprises three fully interconnected Dense layers.
   - The first layer contains 512 units that employ the ReLU activation function.
   - The second layer incorporates a dropout factor of 0.5 to prevent overfitting and minimize random transitions among the units.
   - The final layer comprises a quantity of units that corresponds to the various categories of images to be classified, employing a softmax activation function to align the probabilities of the distinct classifications.

B. The distinction between the proposed and the VGG19 models

This section gives a comparison of the proposed model with the prominent model in the literature, VGG19, which reveals several key differences. A comparison of VGG19 is done here because it is proven in literature [6, 17-19, 23 and 26] and the result of this work is that VGG- models give better results compared with other models. The differences can be outlined as follows:

1. Model Depth:

- **VGG19** is characterized by its 19-layer architecture, which classifies it as a very deep learning model.
- **The proposed model**, in contrast, is constructed with merely four recursive convolution layers, which simplifies its structure while ensuring the quality of classification remains intact. This simplification arises from the fact that fewer features are required for gender classification compared to other applications for which VGG19 was originally designed.

2. Number and Size of Filters:

- **VGG19** employs a variety of filter sizes across its layers, with the last layer utilizing over 128 filters.
- **The proposed model**, however, is structured with a limited number of filters, incorporating 32 filters in the initial layer, 64 in the subsequent layer, and 128 filters in both the third and fourth layers, which effectively minimizes resource requirements.

3. Use of Dropout:

- **VGG19** does not incorporate Dropout in its architecture.
- **The proposed model,** on the other hand, incorporates a Dropout rate of 0.5 within the second layer of the fully connected layers. This is safeguarding the model against overfitting and reducing the occurrence of random transitions among the units.

IV. EXPERIMENTS AND DISCUSSION

To assess the effectiveness of the proposed method, the Veiled-Persons Identification (VPI-New) Database [18, 26] is utilized. This is because standard face recognition databases generally consist of images depicting complete

https://doi.org/10.20428/jst.v30i1.2486

faces, presented from either a frontal or profile viewpoint, except the VPI-Old database. The VPI-New database has been developed to facilitate the identification of individuals while classifying their gender, age, and facial expressions, particularly eye smiles, in images where faces are veiled. The images were recorded with the subject in a seated position on a chair. The camera was situated quite close to the subject, and it shifted slightly after each capture, resulting in images taken from different angles and with a range of eye-level perspectives. It contains 41 males and 109 females (the age of the participants spans from 8 to 78 years old). The images were obtained from two separate sessions, where images of different genders were collected. The images were obtained from multiple environments and under uncontrolled lighting circumstances. Rather than being shot in one specific location, they were taken in more than four different sites, utilizing a range of lighting effects to enhance realism. Figure 5 shows some examples of VIP-new images.

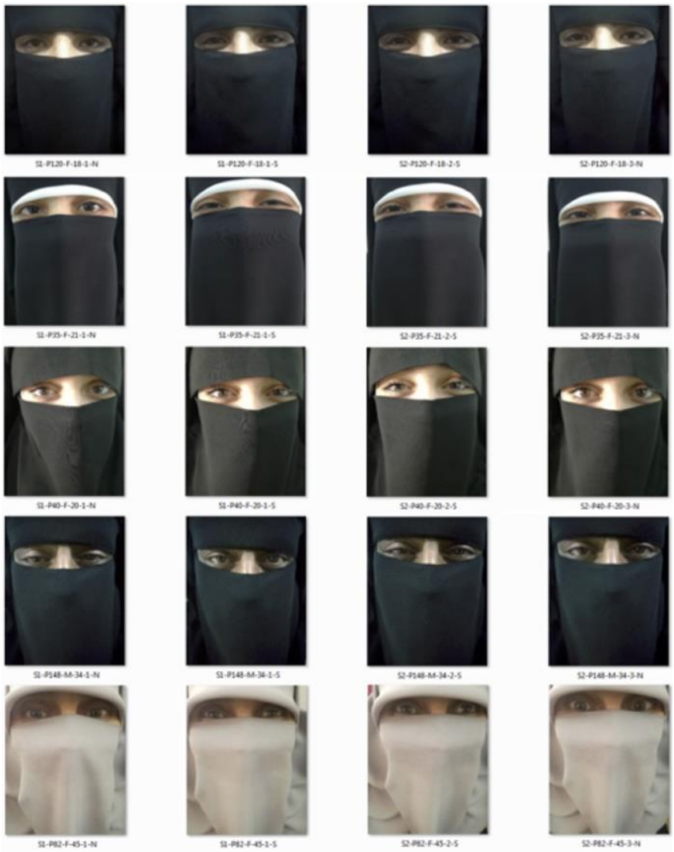

Figure 3: Sample images from VPI-New database

The database is partitioned such that 80% is reserved for training and 20% for testing. The images are resized to a standard size of 150×150 pixels. The number of epochs used to train the models is 30.

In Table 1, a comparison is made between the proposed model and the VGG19 and EfficientNetB7 architectures.

Table 1: Comparison between the proposed model and VGG19 and Efficientnetb7

The model	*Image Size*	*Epochs*	*Accuracy (%)*
VGG19	224×224	30	97.3
EfficientNetB7	224×224	30	95.5
The proposed model	150×150	30	99.1

As illustrated in Table I, the proposed method exhibits a remarkable accuracy of 99.1%, highlighting its effectiveness. Despite utilizing images of lower resolution, specifically 150×150, the proposed method outperforms both VGG19 and EfficientNetB7 in terms of accuracy. Additionally, the results reveal that VGG19 achieves superior accuracy when compared with EfficientNetB7.

V. CONCLUSION AND FUTURE WORKS

This study introduces an innovative model that leverages the strengths of both VGG19 and EfficientNetB7. One of the key advantages of this model is its reduced complexity, characterized by a model depth (lower number of hidden layers) and number and size of Filters. Despite its simplified architecture, the proposed model demonstrates superior performance in terms of accuracy. It achieves an impressive accuracy rate of 99.1% when evaluated on a veiled face database, highlighting its effectiveness in this specific application.

The results indicate that incorporating VGG19 and EfficientNetB7 into a simpler framework not only preserves high accuracy but also improves the model's applicability in real-world scenarios. The equilibrium achieved between complexity and performance establishes the model as a highly promising candidate in the domain of image recognition.

In future research, the proposed method will undergo testing across various facial databases to highlight its advantages compared to alternative models.

VI. References

[1] Guo G. *Gender Classification*. Springer, Verlag London; 2014 Jan.

[2] Nihad A, Saleh A. Fusion of multiple simple convolutional neural networks for gender classification. *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*. IEEE; 2020. p. 251-6.

[3] Jain AK, Duin RP. Introduction to Pattern Recognition. 1997.

[4] Feng L, et al. Human gender classification: a review. *Int J Biometrics*. 2016;8:275-300.

[5] Preeti R, Khanna P. Gender Classification Techniques: A Review. *Advances in Computer Science, Engineering & Applications: Proceedings of the Second International Conference on Computer Science, Engineering and Applications (ICCSEA 2012)*. Springer Berlin Heidelberg; 2016. Volume 1.

[6] Althnian A, et al. Face gender recognition in the wild: an extensive performance comparison of deep-learned, hand-crafted, and fused features with deep and traditional models. *Appl Sci*. 2020;11(1):89-105.

[7] Scheuerman MK, Paul JM, Brubaker JR. How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proc ACM Hum Comput Interact*. 2019;3:1-33.

[8] Alfonso G, et al. A multi-agent system for the classification of gender and age from images. *Comput Vis Image Underst*. 2018;172:98-106.

[9] Mainul I, Ali M. A Comparative Analysis of Machine Learning and Deep Learning Models for Gender Classification from Audio Speech. *Int J Comput Digit Syst*. 2024;16(1):1-11.

[10] Sarah G. *Body image: Understanding body dissatisfaction in men, women and children*. Routledge; 2021.

[11] Xuelong L, Maybank S, Tao D. Gender recognition based on local body motions. *IEEE Int Conf Syst Man Cybern*. IEEE; 2007. p. 3881-6.

[12] Mei W, Weihong D. Deep face recognition: A survey. *Neurocomputing*. 2021;429:215-44.

[13] Giorgio B, et al. Gender Classification via Graph Convolutional Networks on 3D Facial Models. *Proc 39th ACM/SIGAPP Symp Appl Comput*. 2024. p. 482-9.

[14] Ganjar T, et al. Classifying Gender Based on Face Images Using Vision Transformer. *JOIV Int J Inform Vis*. 2024;8(1):18-25.

[15] Rami H, David S, Josep D. Multi-Task Faces (MTF) Data Set: A Legally and Ethically Compliant Collection of Face Images for Various Classification Tasks. *arXiv preprint*. arXiv:2311.11882; 2023.

[16] Arora S, Bhatia M. A robust approach for gender recognition using deep learning. *9th Int Conf Comput Commun Netw Technol (ICCCNT)*. 2018. p. 1-6.

[17] Amit D, Kumar R, Bhan V. Gender recognition through face using deep learning. *Procedia Comput Sci*. 2018;132:2-10.

[18] Hassanat A, et al. Deep learning for identification and face, gender, expression recognition under constraints. *arXiv preprint*. arXiv:2111.01930; 2021.

[19] Zhang X, et al. Gender recognition on RGB-D image. *IEEE Int Conf Image Process (ICIP)*. IEEE; 2020. p. 1836-40.

[20] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2014. p. 1891-8.

[21] Arshad M. Detection Of Human Gender From Eyes Images Using DNN Approach. *J Teknoinfo*. 2013;2(6):2441-5.

[22] Alrashed H, Berbar M. Facial gender recognition using eyes images. *Int J Adv Res Comput Commun Eng*. 2023;17(2):336-40.

[23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*. arXiv:1409.1556; 2014.

[24] Mingxing T. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv preprint*. arXiv:1905.11946; 2019.

[25] Baheti B, et al. Eff-UNet: A novel architecture for semantic segmentation in unstructured environment. *Proc CVPR*. Seattle, WA, USA; 2020. p. 358-9.

[26] Hassanat AB, et al. Classification and gender recognition from veiled-faces. *JOIV Int J Biometrics*. 2017;9(4):347-64.