DEMOGRAPHIC FEATURE RECOGNITION FROM OCCLUDED FACIAL IMAGES USING CONVOLUTIONAL NEURAL NETWORKS.

BY:

NEERAJA SANJAY

SHREYA DHARESHWAR

University at Buffalo The State University of New York

ABSTRACT

This project proposes the development of three custom CNN architectures (based on AlexNet, ResNet, and VGG) for accurate prediction of gender, age, and ethnicity from partially occluded facial images. The models are tailored for multi-task deep learning and aim to overcome the decline in performance typically observed in face demographics detection with occluded images. Evaluations include experiments on non-superimposed occluded and non-occluded facial images to assess the models' generalization and robustness. The results indicate that our models outperform existing models in the same context. The implications of this research are relevant in surveillance, marketing, and other real-world applications that require accurate face detection and recognition, particularly in environments with common facial occlusions.

INTRODUCTION

- This project aims on addressing the challenges of face detection and demographics analysis in the presence of occlusions. We recognize the limitations of existing models that struggle to accurately identify and analyze faces when masks or other obstructions are present.
- To overcome these limitations, we propose leveraging deep learning approaches, specifically CNN-based architectures, and adopting a multi-task learning approach to predict gender, age, and ethnicity from occluded facial images. We also emphasize the importance of retraining the models and updating their weights to optimize their performance.
- We utilize class activation maps to gain insights into the decision-making process of the models and visualize the regions of the face that contribute most significantly to the predicted face demographics

REVIEW OF RELATED RESEARCH WORK

- **1. Prerana Mukherjee, Vinay Kaushik, Ronak Gupta, Ritika Jha, Daneshwari Kankanwadi, and Brejesh Lall: Attribute Prediction in Masked facial images with deep multitask learning: 9th International Conference on Pattern Recognition and Machine Intelligence (PReMI 2021)**

In this paper, the authors propose a ResNet architecture CNN model to predict facial attributes from black and white masked images. **Gender** Accuracy: 88.03, **Ethnicity** Accuracy: 74.25 , **Age** Mean average error: 12.60. This paper elaborates the state-of-the-art method to detect facial attributes in occluded images. We have experimented with the model architecture of this model.

- **2. Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 (2017)**

This article provides a general overview of multi-task learning in deep neural networks, using hard parameter sharing and recent advances. It also offers guidelines for ML practitioners on choosing appropriate auxiliary tasks. We have incorporated concepts borrowed from this paper by using multi-task learning (MTL) for training a single network to predict age, gender, and ethnicity simultaneously, leveraging shared representations.

- **3. Age and gender Prediction using Deep CNNs and transfer learning – Vikas Sheoran, Shreyansh Joshi**

This paper talks about using CNNs to predict age and gender from NON occluded , black and white images. By adapting their architecture choices, we aim to improve the robustness for age and gender detection. Unlike the paper which uses transfer learning on non-occluded black and white images, we have trained the models from scratch on occluded colored images.

- **4. G. Wu, J. Tao and X. Xu, "Occluded Face Recognition Based on the Deep Learning," 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, 2019, pp. 793-797, doi: 10.1109/CCDC.2019.8832330**

The paper introduces a new method for recognizing occluded faces using deep learning, specifically a convolution neural network. The model is robust to changes in illumination, facial expression, and occlusion, achieving a recognition rate of up to 98.6% in experimental tests. The proposed method enables face recognition in complex environments and has practical applications.

INITIAL APPROACH

- In the initial approach, we developed a custom CNN based on the ALEXNET architecture for multi-task learning. However, this model did not yield satisfactory results, particularly in predicting ethnicity and age. Despite incorporating convolutional layers, batch normalization, and LeakyReLU activation, it struggled to capture the complex patterns required for accurate ethnicity classification. Therefore, we proposed an enhanced architecture based on RESNET presented in the state-of-the-art multi-task learning model.
- Our RESNET-based model utilized resized colored images of 50x50 resolution and employed retraining techniques. It consisted of convolutional layers, batch normalization, leaky ReLU activations, max pooling, average pooling, and three parallel branches for gender, ethnicity, and age predictions. The model was trained for two phases: an initial phase with a batch size of 8 for 39 epochs and a retraining phase with a batch size of 12 for 4 epochs. This model demonstrated superior performance compared to the state-of-the-art model for all three attributes, gender, ethnicity, and age.
- .While the results of our RESNET variation model were commendable, we noticed there was still room for improvement. To further enhance the model's precision in predicting ethnicity, we propose another methodology.

DATASET

- We utilized the UTK Face Dataset, which consists of over 20,000 images of individuals from diverse ethnicities, including White, Black, Asian, Indian, and others. Each image is annotated with age, gender, and ethnicity information.
- To tackle the challenge of occlusions, the dataset should contain images with different types of masks, simulating real-world scenarios. So, we superimposed masks on images into our data set using a tool called MaskTheFace. Masks of different types and colors are superimposed. **Surgical, N95, KN95, Cloth, Gas**
- These superimposed **color** images are processed to get its flattened pixel values with 50x50 resolution, which are used for training the model. A total of 21k images are used to train the model.

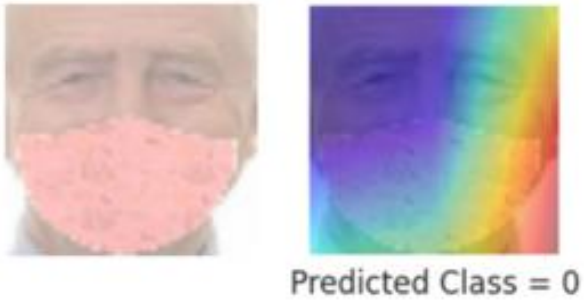
ALGORITHM:

- We propose a model architecture based on VGG16 for multi-task learning, simultaneously predicting gender, ethnicity, and age from colored input images. The model utilizes a common encoder module shared among all tasks, allowing it to learn shared representations while focusing on task-specific features. The gender, ethnicity, and age prediction components follow specific layers and operations tailored to each attribute.
- To train the model, task-specific loss functions are applied: binary cross-entropy for gender, categorical cross-entropy for ethnicity, and mean squared error for age. The model's performance is evaluated using accuracy for gender and ethnicity predictions and mean absolute error (MAE) for age estimation. The model was initially trained for 40 epochs using a batch size of 32. It was then retrained twice, with smaller batch sizes and fewer epochs. In the first retraining session, the model was trained for 5 epochs with a batch size of 17. In the second retraining session, the model was trained for 3 epochs with a batch size of 32.
- The proposed model surpasses the state-of-the-art performance in detecting age, gender, and ethnicity in occluded facial images. It also demonstrates good accuracy in non-occluded and real-time images with individuals wearing non-superimposed masks or sunglasses. The model's high accuracy showcases its robustness and generalization capabilities, effectively learning and utilizing facial features indicative of gender, ethnicity, and age in challenging and realistic conditions.

ALGORITHM:

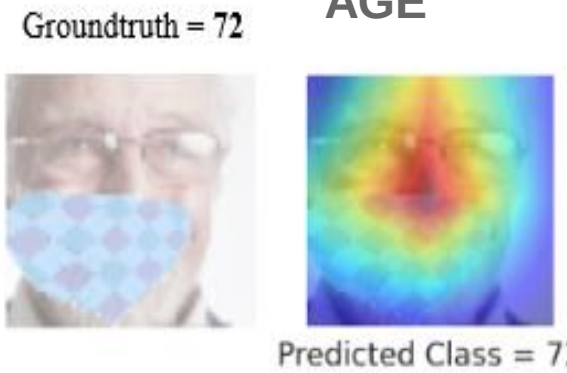
- Class Activation Maps (CAM) are utilized to provide interpretability to the prediction models.

GENDER



ETHNICITY

AGE



- For gender prediction, the CAM indicates that the ear region is influential in distinguishing between male and female. Visible ears may contribute to classifying an individual as male, while hair covering the side of the forehead and ear may be associated with females.
- Regarding ethnicity prediction, the forehead region plays a significant role. Different ethnic groups may exhibit distinct characteristics in the forehead, such as variations in skin tone, texture, or the presence of specific facial structures. For example, the presence of a "bindhi" on the forehead could indicate the individual's classification as Indian.
- For age prediction, the model emphasizes the bridge of the nose and the middle of the forehead. These areas often display visible signs of aging, including wrinkles, lines, and texture changes. The model may associate the presence of these age-related characteristics with older age groups.

RESULTS

Model Architecture	Method	Type of Images	Resolution	Gender (Accuracy)	Ethnicity (Accuracy)	Age (MAE)
RESNET (State of the art)	***EXISTING***	***BLACK AND WHITE***	***48 x 48***	***0.8913***	***0.7927***	***12.15***
ALEXNET	OURS	COLORED	50x50	0.79430	0.5044	8.209
RESNET VARIATION	OURS	COLORED	50x50	0.9182	0.8203	5.537
VGG16 VARIATION	**OPTIMAL - OURS**	**COLORED**	**50x50**	0.9509	0.8733	6.598

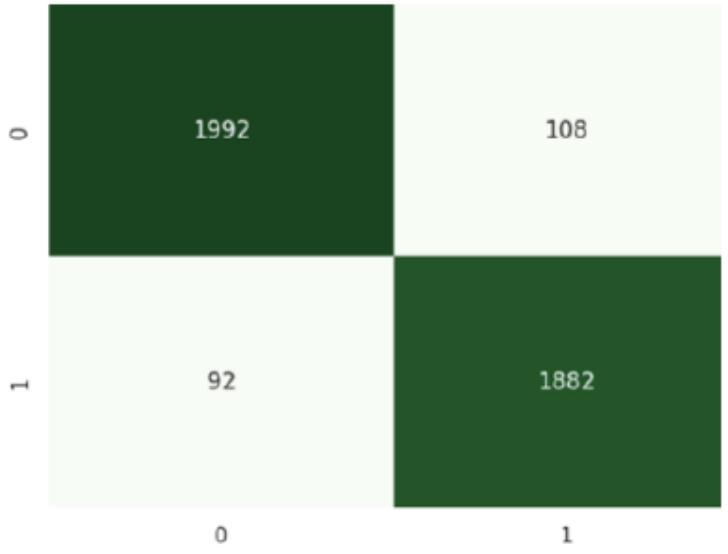
- From the given table it is clear that the VGG16 Variation model was found to be the most optimal choice, exhibiting competitive performance across all evaluated metrics in gender, ethnicity, and age prediction tasks.

RESULTS - VGG16 Variation Evaluation

GENDER

Class	Precision	Recall	F1 score
Male	0.96	0.95	0.95
Female	0.95	0.95	0.95

These results indicate that the model has a high accuracy in predicting both male and female genders, with balanced precision, recall, and F1-score values for both classes.

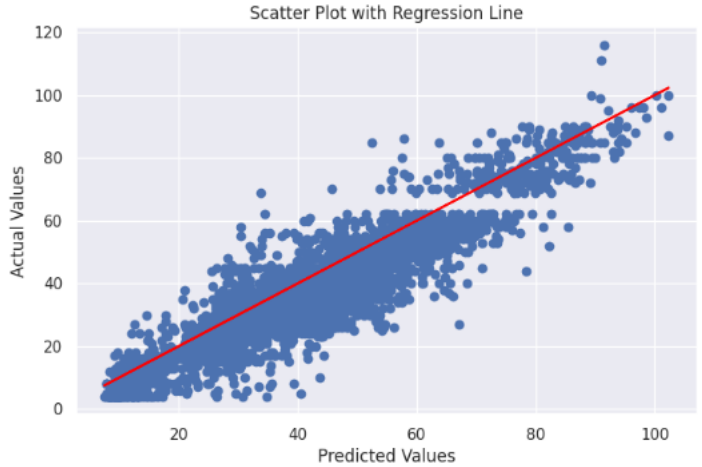


ETHNICITY

Class	Precision	Recall	F1 score
White	0.86	0.95	0.90
Black	0.89	0.86	0.88
Asian	0.95	0.82	0.88
Indian	0.89	0.81	0.85
Others	0.78	0.65	0.71

The model shows good accuracy and balanced performance in predicting the ethnicities of white, black, Asian, and Indian individuals, with precision values ranging from 0.86 to 0.95, recall values ranging from 0.81 to 0.95, and F1-scores ranging from 0.85 to 0.90. These results indicate that the model is effective in identifying samples from these ethnicities and provides promising results overall.

AGE



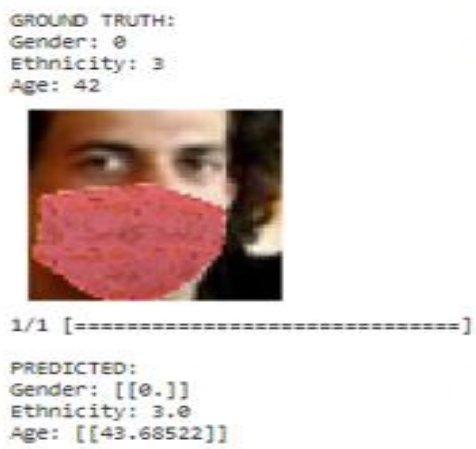
The regression plot demonstrates a strong correlation between the predicted age values and the actual age values, indicating that the model captures the underlying patterns in the age data. While there is some scatter around the line, the model is still able to provide accurate age predictions with a certain level of variability.

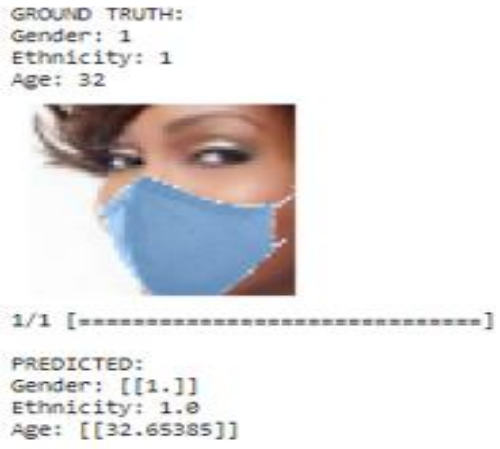
COMPARISON WITH THE STATE OF THE ART

- Comparing our ResNet variation to the state-of-the-art ResNet, we observed improvements in all components This implies that our RESNET model provides more precise age predictions, resulting in smaller deviations from the actual age values.
- Our VGG16 variation, on the other hand, exhibits even more impressive results compared to the state-of-the-art model and our RESNET model. The VGG16 model provides the most optimal results.
- One important aspect to consider is the resolution of the images used to train the model. In our case, the VGG16 variation and the ResNet variation was trained on colored images with a resolution of 50x50 pixels, while the state-of-the-art ResNet was trained on black and white images with a resolution of 48x48 pixels.
- Even though we used the same architecture as the state-of-the-art model to develop our ResNet model, the superior performance of our ResNet and VGG16 variation models can be attributed to various factors, such as the utilization of colored images with a slightly higher resolution, the use of colored images, retraining of the models to adapt to the specific task, and potentially optimized model architectures. These factors enable our models to leverage richer visual information, learn task-specific features, and improve overall prediction accuracy.

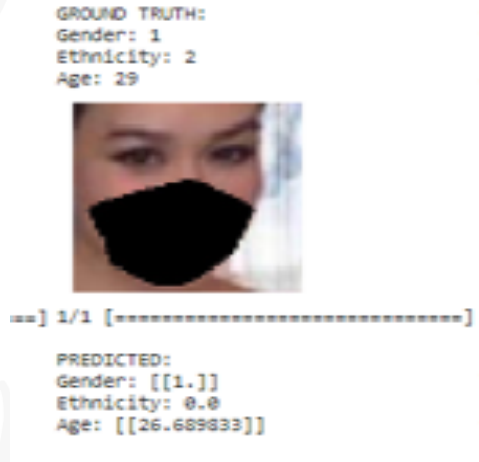
INTERPRETATION

- The VGG16 variation model demonstrates overall successful predictions, particularly in gender classification with low errors. However, it faces challenges in accurately predicting ethnicity for Asian individuals, often misclassifying them as White ethnicity. Additionally, it struggles to distinguish between Indian and Asian ethnicities, indicating difficulties in capturing subtle facial characteristics that differentiate these groups.





SUCCESS CASES



FAILURE CASES

- These errors may be attributed to the model's limited ability to capture age-related visual cues in these individuals. While the model achieves high accuracy, the few gender prediction errors could be influenced by data distribution biases. In terms of age prediction, errors occur for individuals around the age of 75, particularly bald individuals, where the model tends to underestimate their age. These errors may be attributed to the model's limited ability to capture age-related visual cues in these individuals.

EXPERIMENTS

TESTING ON NON-MASKED IMAGES

Model Architecture	Method	Type of Images	Resolution	Gender (Accuracy)	Ethnicity (Accuracy)	Age (MAE)
RESNET (State of the art)	*EXISTING*	*BLACK AND WHITE*	*48 x 48*	*0.89*	*0.76*	*11.57*
RESNET VARIATION	OURS	COLORED	50x50	0.94	0.70	5.63
VGG16 VARIATION	**OPTIMAL - OURS**	**COLORED**	**50x50**	***0.95***	***0.89***	***5.68***

- In the evaluation of non-masked facial images, our VGG16 variation model demonstrated superior performance compared to the state-of-the-art model. This improvement can be attributed to the removal of masks, which provided the model with more complete and unobstructed facial information, leading to enhanced prediction accuracy.

- The ResNet variation model performed well in gender prediction and overall ethnicity classification but had errors in age estimation and fine-grained ethnicity classification. On the other hand, the VGG16 variation model demonstrated better accuracy in predicting age and showed improved ethnicity classification due to its deeper architecture and ability to capture intricate features specific to different ethnicities.
- Overall, the VGG16 variation model outperformed the state-of-the-art model in gender prediction, ethnicity prediction, and age estimation for both masked and non-masked facial images.

EXPERIMENTS

- **TESTING ON REAL TIME OCCLUDED FACIAL IMAGES**

- To assess the real-world applicability and robustness of our trained model, we evaluated its performance on facial images without superimposed occlusions using the Real World Occluded Faces (ROF) dataset.
- The VGG16 variation model showcased effectiveness in this challenging scenario, accurately predicting facial attributes even in the presence of non-superimposed real occlusions. Out of the 50 images from the ROF dataset, the model correctly predicted all three attributes in 41 images, resulting in a cumulative accuracy of 82%.
- However, errors were observed specifically for sunglass occlusions, leading to inaccuracies in age prediction due to the occlusion of the nose bridge. Blurred images also contributed to errors in ethnicity prediction. When comparing the RESNET and VGG16 variation models, the VGG16 model outperformed in attribute prediction in this task.

FAILURE CASES - **REAL TIME OCCLUDED FACIAL IMAGES**

COMPARISON BETWEEN RESNET VARIATION AND VGG16 VARIATION - **REAL TIME OCCLUDED FACIAL IMAGES**

FUTURE SCOPE

- Further, we plan on implementing techniques to enhance occlusion detection and segmentation in face demographic analysis models, to automatically detect and segment occluded regions in facial images, enabling the models to focus on the visible parts of the face for more accurate predictions and improved handling of occlusions.
- By exploring advanced computer vision methods, such as instance segmentation or generative adversarial networks (GANs) we plan to successfully implement these techniques. Additionally, we plan to focus on further enhancing the models' performance, especially in predicting ethnicity, by exploring additional architectural variations, dataset augmentation techniques, or incorporating other sources of information, such as facial landmarks or contextual cues.
- We intend to explore our work for learning other facial attributes as well.

CONCLUSION

- The VGG16 variation model achieved high accuracy rates for gender, ethnicity, and age prediction, both in occluded and non-occluded facial images, indicating its robustness and generalization capabilities.
- The findings of our research have significant implications in surveillance, marketing, and other real-world applications where face detection and recognition are crucial, especially in environments with frequent facial occlusion.
- Predicting demographic attributes from occluded images can aid security personnel in identifying potential threats and optimizing their response strategies.
- Additionally, businesses can leverage this information to deliver targeted ads and promotions that resonate with specific demographics, resulting in enhanced customer engagement and satisfaction. These applications highlight the practical value and potential impact of our model in diverse fields.

THANK YOU