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

PRESENTED BY:
NEERAJA SANJAY & SHREYA DHARESHWAR

INTRODUCTION

Deep learning has revolutionised tasks like object detection and face recognition. However, detecting demographic features from occluded facial images remains challenging. In this project, our objective is to develop a system for demographic feature recognition in **occluded** faces using **Convolutional Neural Networks** (CNNs) and **multi-task learning** approach. We explore the accuracy of popular CNN models, **ResNet**, **VGG16** and **AlexNet**, in predicting age, gender, and ethnicity from occluded facial images. 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. Furthermore, we evaluate the performance of the models on non-occluded faces to assess its generalization capabilities.

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.

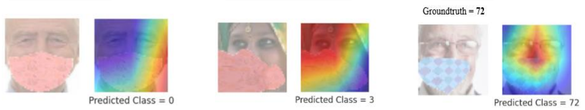
To tackle the challenge of occlusions, the dataset should contain images with different types of masks, simulating real-world scenarios. So, we superimposed masks on images into our data set using a tool called MaskTheFace. Masks of different types and colors are superimposed. **Surgical, N95, KN95, 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

After considering 2 other algorithms (ResNET) variation and Alexnet variation) We propose a model architecture based on VGG16 for multi-task learning, simultaneously predicting gender, ethnicity, and age from colored input images. 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.

GENDER **ETHNICITY** **AGE**



For gender prediction, the CAM indicates that the ear region is influential in distinguishing between male and female. Hair covering the side of the forehead and ear may be associated with females. Regarding ethnicity prediction, the forehead region plays a significant role. 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.

RESULTS

Model Architecture &	Type of Images	Resolution	Gender (Acc.)	Ethnicity (Acc.)	Age (MAE)
ResNet (SOTA)	**B&W**	**48x48**	**0.8913**	**0.7927**	**12.15**
AlexNet	Color	50x50	0.7943	0.5044	8.209
ResNet Variation	Color	50x50	0.9182	0.8203	5.537
VGG16 Variation (Optimal)	**Color**	**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.

VGG16 VARIATION EVALUATION

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

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 demonstrates high accuracy in predicting gender, achieving balanced precision, recall, and F1-scores for both male and female classes. Additionally, it exhibits good accuracy and balanced performance in predicting ethnicities, particularly for white, black, Asian, and Indian individuals, with precision, recall, and F1-scores ranging from 0.86 to 0.95, 0.81 to 0.95, and 0.85 to 0.90, respectively. Overall, the model shows promising results in identifying gender and ethnicities accurately.

The regression plot reveals a strong correlation between predicted and actual age values, indicating the model captures underlying age patterns with some variability.

EXPERIMENTS

▪**Testing on non-masked images**

Model Architecture	Method	Image Type	Result	Gender (Accuracy)	Ethnicity (Accuracy)	Age (MAE)
ResNet (SOTA)	**Existing**	**B&W**	**48x48**	**0.89**	**0.76**	**11.57**
ResNet Variation	Ours	Color.	50x50	0.94	0.70	5.63
VGG16 Variation (Optimal)	**Ours**	**Color.**	**50x50**	**0.95**	**0.89**	**5.68**

The removal of masks in non-masked facial images resulted in the VGG16 variation model outperforming the state-of-the-art model, indicating that the availability of complete and unobstructed facial information contributed to improved prediction accuracy.

▪**Testing on real time occluded facial images**

When comparing the RESNET and VGG16 variation models, the VGG16 model outperformed in attribute prediction in this task**.**

These Experiments prove that the model is robust and works well in generalized cases.

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

FUTURE SCOPE

To improve occlusion detection and segmentation in face demographic analysis models, techniques such as instance segmentation, generative adversarial networks (GANs), and the utilization of the FAN model for face detection can be implemented. Further enhancement of the models' performance, particularly in predicting ethnicity, can be achieved by exploring additional architectural variations, dataset augmentation techniques, and incorporating facial landmarks or contextual cues. Additionally, exploring other facial attributes can also be considered.

CONCLUSION

Our research findings have important implications for various real-world applications, including surveillance, marketing, and other contexts where face detection and recognition are essential, particularly in environments with frequent facial occlusion. Predicting demographic attributes from occluded images can help security personnel identify potential threats and optimize response strategies. Businesses can also utilize this information to deliver targeted ads and promotions, enhancing customer engagement and satisfaction. These practical applications underscore the significant practical value and potential impact of our model across diverse fields.

REFERENCES

[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).

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

[3] https://towardsdatascience.com/masktheface-cv-based-tool-to-mask-facedataset-1a71d5b68703

University at Buffalo The State University of New York

Department of Computer Science and Engineering
School of Engineering and Applied Sciences
cse-dept@buffalo.edu