

2019

On Matching Faces with Temporal Variations using Representation Learning

Daksha Yadav

West Virginia University, dayadav@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Yadav, Daksha, "On Matching Faces with Temporal Variations using Representation Learning" (2019). *Graduate Theses, Dissertations, and Problem Reports*. 3939.

<https://researchrepository.wvu.edu/etd/3939>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

On Matching Faces with Temporal Variations using Representation Learning

Daksha Yadav

Dissertation submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science

Afzel Noore, Ph.D., Chair
Victor Fragoso, Ph.D.
Keith B. Morris, Ph.D.
Mayank Vatsa, Ph.D.
Richa Singh, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2018

Keywords: Face Recognition, Temporal Variations, Facial Aging, Facial Plastic Surgery,
Machine Learning, Deep Learning

Copyright 2018 Daksha Yadav

Abstract

On Matching Faces with Temporal Variations using Representation Learning

by

Daksha Yadav

Developing automatic face recognition algorithms which are robust to intra-subject variations is a challenging research problem in the computer vision research community. Apart from the well-studied covariates such as pose and expression, temporal variations in the facial appearance also lead to a decline in the performance of face recognition systems. This research focuses on analyzing the temporal variations in facial features due to facial aging, facial plastic surgeries, and prolonged illicit drug abuse. The contributions of this dissertation are fivefold: (i) behavioral and neuroimaging studies are conducted to understand the human perception of faces affected by temporal variations, specifically facial aging; (ii) a novel generative adversarial network-based solution is proposed to match age-separated faces; (iii) the influence of temporal variations in faces altered by plastic surgery procedures is examined and a novel framework for detecting and verifying such faces is proposed; (iv) the effect of illicit drug abuse on face images is introduced and a new Illicit Drug Abuse Face (IDAF) database is created for the research community; and (v) a novel algorithm for single-image based detection of faces with plastic surgery and illicit drug abuse is proposed and its utilization as soft biometric in enhancing face recognition performance is demonstrated.

This research attempts to evaluate how humans perceive facial age and their ability to recognize age-separated faces. To accomplish this objective, two human studies (behavioral and neuroimaging) are conducted. The findings from these studies suggest that regular faces are processed differently from age-separated faces, highlighting the need for building specialized face recognition algorithms for processing such faces. Motivated by this observation, we propose a novel deep learning algorithm for matching faces with temporal variations caused by age progression. The proposed algorithm utilizes a unified framework which combines facial age estimation and age-separated face verification using generative adversarial networks. The key idea of this approach is to learn the age variations across time by conditioning the input image on the subject's gender and the target age group to which the face needs to be progressed. We demonstrate the efficacy of the proposed architecture on different facial age databases for age-separated face recognition.

We also analyze the temporal variations with respect to facial plastic surgeries. A novel solution is proposed to differentiate plastic surgery faces from regular faces by learning

representations of local and global plastic surgery faces using multiple projective dictionaries. Experimental results on the plastic surgery database show that the proposed framework is able to detect plastic surgery faces with a high accuracy of around 98%. To verify the identity of a person, the detected plastic surgery faces are divided into local regions of interest that are likely to be altered by a particular plastic surgery followed by distance metric calculation of feature representations. The proposed framework for face verification is combined with two commercial systems to demonstrate an improvement in face verification performance.

In this research, the impact of prolonged illicit drug abuse on face recognition is also introduced. Certain drugs, when taken continuously in large quantities, can cause physiological changes in the skin. For instance, the long-term effects of methamphetamine and heroin can cause severe weight loss and skin sores while addiction to opiates may lead to accelerated aging. We demonstrate that these physiological variations induced due to extensive substance abuse dramatically decrease the performance of current face recognition algorithms by increasing the intra-class distance between the facial appearance of a subject. This research also proposes a novel projective dictionary learning based illicit drug abuse face classification framework to effectively detect and separate faces affected by drug abuse from normal faces.

Lastly, two novel algorithms for single-image based detection of faces with temporal variations, specifically, plastic surgery and illicit drug abuse are proposed. In the proposed formulations, the variations in different local regions of these faces are analyzed by incorporating deep learning based multi-instance learning. The proposed approaches also utilizes multi one-shot metric to encode inter-class and intra-class variations leading to higher face image classification accuracy. Moreover, the classification scores from the proposed algorithms are utilized as soft biometric information to enhance the performance of existing face recognition algorithm.

Dedication

This dissertation is dedicated to my parents, who have always encouraged me to pursue my passion and achieve my dreams, especially this one . . .

Acknowledgements

I wish to utilize this opportunity to thank all the people who in one way or another contributed to the completion of this dissertation.

First and foremost, I gratefully acknowledge my advisor Dr. Afzel Noore for his continued guidance and motivation throughout this journey. His wise words motivated me to remain patient even during tough times and challenged me to think outside the box. I am also highly thankful to my sources of inspiration, Dr. Richa Singh and Dr. Mayank Vatsa for guiding me since my undergraduate degree. Their hard work, perseverance, and curiosity to work on novel research problems provoked me to choose the path of research. I also gratefully acknowledge the support and feedback provided by the other committee members, Dr. Victor Fragoso, and Dr. Keith Morris.

Additionally, I would like to take this chance to thank my collaborators. Their valuable support and help have been an essential component of my graduate life. I wish to thank my co-authors Maneet, Shruti, Prateekshit, Akshay, Shivangi, and Ekampreet for their hard work and countless sleepless nights we spent working on different deadlines. I also wish to thank Anush, Tejas, Rohit, Soumyadeep, Aakarsh, and Brian for priceless work discussions and invaluable life lessons. I also want to thank my friends in Morgantown for their support and love which made Morgantown my second home. My lab colleague and longtime friend Naman Kohli deserves a special mention. His determination and sagacious advice have been key driving factors in this arduous and exciting journey.

Finally, I am highly grateful to my parents for making immeasurable sacrifices for me. They have been my constant pillars of support and I will continue to strive every day to make them proud. I express gratitude to my brother, Adit and sister-in-law, Poonam for supporting my decisions and being my towers of strength. Thank you to all of you!

Table of contents

Dedication	iv
Acknowledgements	v
List of figures	x
List of tables	xiv
List of Publications	xvi
1 Introduction	1
1.1 Problem Statement	7
1.2 Related Work	7
1.3 Research Contributions	13
2 Understanding Human Capabilities in Processing Faces with Temporal Variations	16
2.1 Behavioral Study to Unravel Human Perception of Facial Aging	17
2.1.1 Methods	18
2.1.2 Results and Discussion	20
2.2 Neural Correlates of Face Verification due to Facial Aging	27
2.2.1 Methods	27
2.2.2 Preprocessing and Statistical Analysis	31
2.2.3 Results	32

2.3	Summary	34
3	Verifying Age-Separated Faces using Age Gap Reducer GAN	35
3.1	Proposed AGR-GAN	38
3.1.1	Objective Function	41
3.1.2	Implementation Details	43
3.2	Experimental Evaluation	43
3.2.1	Databases	44
3.2.2	Experimental Analysis	44
3.3	Summary	53
4	Recognizing Faces Altered by Plastic Surgery using Multiple Projective Dictionary	54
4.1	Multiple Projective Dictionary Learning Framework for Face Verification .	56
4.1.1	Preprocessing of Face Images	56
4.1.2	Proposed Dictionary based Detection of Plastic Surgery	56
4.1.3	Face Verification after Plastic Surgery Detection	61
4.2	Experimental Protocol	64
4.3	Results and Analysis	66
4.3.1	Parameter Setting	66
4.3.2	Plastic Surgery Detection	66
4.3.3	Face Verification Performance on Plastic Surgery Pairs	68
4.3.4	Performance of integrating the Proposed MPDL Framework with COTS	69
4.4	Summary	71
5	Introducing the Effect of Illicit Drug Abuse based Temporal Variations on Face Recognition	72
5.1	Effect of Illicit Drug Abuse on Face Recognition Algorithms	74
5.1.1	Creation of Illicit Drug Abuse Face (IDAF) Database	74

5.1.2	Face Recognition Algorithms for Evaluation	75
5.1.3	Experimental Scenarios	75
5.1.4	Experimental Results	76
5.2	Proposed Dictionary Learning based Illicit Drug Abuse Face Classification Framework	81
5.2.1	Pre-processing and Extraction of Facial Regions of Interest	81
5.2.2	Computation of Multiple Discriminatory Features	82
5.2.3	Feature Specific Paired Dictionary Learning	83
5.2.4	Illicit Drug Abuse Face Classification and Decision-Level Fusion	84
5.3	Experimental Results	85
5.3.1	Experimental Setup	85
5.4	Summary	86
6	Detecting Plastic Surgery and Illicit Drug Abuse Faces using Adaptive Deep Multi-One Shot Similarity based Multiple Instance Learning	88
6.1	Proposed Algorithms	90
6.2	Preliminaries	92
6.2.1	One-Shot Similarity	92
6.2.2	Multi-instance Learning	93
6.2.3	Laplacian Score	93
6.2.4	DenseNet Architecture	94
6.3	Proposed DMOSMIL Algorithm for Face Image Classification	95
6.4	Proposed Adaptive DMOSMIL Algorithm for Face Image Classification	97
6.5	Experiments	99
6.5.1	Implementation Details	99
6.5.2	Databases and Experimental Protocol	100
6.5.3	Feature Extractors for Comparative Analysis	100
6.5.4	Existing Metric Learning Algorithms for Comparative Analysis	101
6.5.5	Results	102

6.6	Utilizing Face Classification as Soft Biometric for Boosting Face Recognition	109
6.6.1	Experimental Protocol and Results	110
6.7	Summary	111
7	Conclusions and Future Work	113
	Bibliography	117

List of figures

1.1	Face images of two subjects captured under constrained environment with limited/no wild variations.	2
1.2	Examples of some intra-subjects variations which hamper the performance of face recognition algorithms.	2
1.3	Face images of two subjects showcasing the facial aging variations. The top row illustrates the variations during the formative years of the individual. The bottom row shows that more textural variations are evident during the later years.	3
1.4	Sample before and after plastic surgery face images of different subjects. . .	4
1.5	Sample before and after face images of different plastic surgeries illustrating local and global variations in the facial features.	4
1.6	Illustrating the significant effect of illicit drug abuse on faces. Noticeable variations can be seen in the facial features of the <i>after</i> images of these subjects.	6
2.1	Sample genuine and imposter age-separated face pairs illustrating notable variations between the two images of the respective face pairs.	17
2.2	Sample facial regions that are presented to participants for age group estimation.	19
2.3	Sample images presented to the participants for the task of recognizing age-separated images of individuals.	20
2.4	Protocol for face verification task. The response time is marked in red and the maximum limit on response time is 2.5 seconds. The screen turned blank after the participant provided response until 4.5 seconds.	29

2.5	Brain areas showing significant areas of activations in the group analysis for Age-Separated Face Pairs vs Baseline condition. Activations are reported at voxel-level threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ (FDR corrected).	33
3.1	Illustrating the problem of age-separated face recognition and the solution proposed in this research using Age Gap Reducer (AGR)-GAN.	37
3.2	Proposed Age-Gap Reducer Generative Adversarial Network (AGR-GAN) architecture.	40
3.3	Sample generated outputs by the proposed Age Gap Reducer-GAN across different age groups. Each row contains outputs for different subjects from different databases.	45
3.4	Illustrating the findings of ablation study to analyze the effect of removing the three loss functions in AGR-GAN. The age range of the input face image is 21-30 years)	49
3.5	Receiver Operating Characteristic (ROC) curves demonstrating the increase in the performance of FaceNet by using AGR-GAN outputs on CALFW database.	52
4.1	Proposed two-stage Multiple Projective Dictionary Learning (MPDL) framework to detect plastic surgery for face verification.	57
4.2	Facial regions extracted from pre-surgery and post-surgery images to perform face verification. The regions are (1) forehead, (2) right brow, (3) left brow, (4) left eye, (5) right eye, (6) left cheek, (7) right cheek, (8) nose, (9) lips, (10) chin, and (11) full face.	63
4.3	ROC for face verification on plastic surgery images only using HOG, UCLBP, LBP, and TPLBP face descriptors, De Marsico <i>et al.</i> [25] and Sun <i>et al.</i> [119], and Stage-2 of the proposed MPDL framework.	69

5.1	Sample images that demonstrate the significant effect of illicit drug abuse on faces. Noticeable variations can be seen in the facial features of the <i>after</i> images of these subjects.	73
5.2	CMC curves for <i>Regular Faces</i> and <i>Combined Faces</i> Scenario when COTS systems FaceVACS [21] and Luxand [77] are used. It is seen that introduction of illicit drug abuse images lowers the performance of COTS.	77
5.3	CMC curves for <i>Regular Faces</i> and <i>Combined Faces</i> Scenario when two face descriptors, LBP [86] and HOG [22] are used. It is seen that introduction of illicit drug abuse images lowers the performance of the face descriptors as well.	78
5.4	CMC curves for <i>Drug Faces</i> Scenario when two COTS, FaceVACS [21] and Luxand [77], and two face descriptors, LBP [86] and HOG [22] are used. It is seen that very few subjects have been correctly identified.	79
5.5	Proposed D ictionary learning based illicit D rug A buse face C lassification (DDAC) framework to classify faces affected by drug abuse.	80
5.6	Sample images from classes <i>drug-abuse faces</i> and <i>regular faces</i> which are correctly and incorrectly classified by the proposed DDAC framework.	84
6.1	Illustrating the variations in facial features due to facial plastic surgery.	89
6.2	Illustrating the variations in facial features due to prolonged illicit drug abuse.	89
6.3	Proposed Deep Multi-One Shot Similarity-based Multiple Instance Learning (DMOSMIL) algorithm for atypical face image classification.	91
6.4	Architecture of DenseNet based convolutional neural network.	94
6.5	ROC curves of the proposed DMOSMIL and ADMOSMIL algorithms on the plastic surgery face database.	105
6.6	Sample classification outputs by DenseNet feature $F_{ab=2}$ based DMOSMIL and ADMOSMIL algorithms on the plastic surgery face database. The label Positive is plastic surgery face and Negative is regular face.	105
6.7	ROC curves of the proposed DMOSMIL and ADMOSMIL algorithms on the proposed E-IDAF database.	107

6.8	Sample classification outputs by the DenseNet feature $F_{db=2}$ based DMOSMIL and ADMOSMIL algorithms on the proposed E-IDAF database. The label Positive is illicit drug abuse and Negative is regular face.	107
6.9	CMC curves demonstrating the increase in face recognition performance by utilizing ADMOSMIL scores as soft biometric.	112
7.1	Types of covariates of face recognition: (a) covariates which impact the facial features due to temporal variations, (b) covariates due to heterogeneity of the face image domains, and (c) variations arising due to the interaction of the user with the image acquisition sensor.	113

List of tables

1.1	Overview of face age progression (generative) algorithms in the literature.	8
1.2	Overview of age-invariant face recognition (discriminative) algorithms in the literature.	9
1.3	Summary of face recognition algorithms designed for matching faces altered by plastic surgery.	11
2.1	Confusion matrix showing comparison between actual and predicted age group in the task of age estimation.	21
2.2	Analyzing the performance for estimating the age group of shown face stimulus. The results show that highest accuracy is obtained when the stimuli faces are in the range of 0 to 5 years.	21
2.3	Analyzing the effect of an individual facial region shown for predicting the age group of the stimuli. The results show that the chin region of children belonging to 0 to 5 years age is most prominent for age prediction.	25
2.4	Face recognition accuracy achieved with respect to stimuli age group and type of facial region shown. The values in the bold show which region is the most discriminating for recognizing the stimuli's belonging to the given age groups. It can be observed that in general, the whole face yields the highest accuracy whereas, for children and elderly people, binocular and chin regions are the most prominent respectively.	25

2.5	Regional brain activity for age-separated stimuli face pairs. Coordinates are listed according to MNI space. Activations are reported at voxel-level threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ (FDR corrected).	32
3.1	Age estimation (years) of faces generated by the proposed AGR-GAN on (a) MORPH, (b) CACD-VS, and (c) CALFW databases.	46
3.2	Equal error rate (%) for face verification performance of input test faces and faces from the 10 target age groups generated by the proposed AGR-GAN on (a) MORPH, (b) CACD-VS, and (c) CALFW databases.	47
3.3	Ablation study using aging model evaluation experiment to analyze the contribution of the age gap loss in the proposed AGR-GAN. Mean estimated age (years) of generated faces is reported.	50
3.4	Ablation study using identity preservation evaluation experiment to evaluate the contribution of the identity loss in the proposed AGR-GAN. Equal error rate (%) of face verification performance is reported.	50
3.5	Illustrating the increase in face recognition performance of FaceNet by using faces generated from the proposed AGR-GAN on: (a) MORPH, (b) CACD-VS, and (c) CALFW databases.	51
4.1	Regions of interest (ROIs) affected by different plastic surgeries.	62
4.2	Number of image pairs used for different surgeries from plastic surgery face database [116].	64
4.3	Accuracy (%) for plastic surgery detection using different classification algorithms.	67
4.4	Plastic surgery detection accuracy based on different plastic surgeries using Stage-1 of the proposed MPDL framework.	68

4.5	Performance of COTS-1 and COTS-2 with regular faces, combined database of plastic surgery and regular face images, and the proposed MPDL framework integrated with commercial systems.	70
5.1	Average detection accuracy (%) for illicit drug abuse face classification using different classification algorithms.	87
6.1	Classification accuracy of the proposed DMOSMIL, proposed ADMOSMIL, LDA, M-OSS, and MIL for plastic surgery detection.	104
6.2	Classification accuracy of the proposed ADMOSMIL and comparative algorithms for plastic surgery detection.	105
6.3	Classification accuracy of the proposed DMOSMIL, proposed ADMOSMIL, LDA, M-OSS, and MIL on the E-IDAF database for illicit drug abuse detection.	106
6.4	Classification accuracy of the proposed ADMOSMIL and comparative algorithms for illicit drug abuse detection.	107

List of Publications

* indicates research published as part of this dissertation

Journal Publications

- 1.* **D. Yadav**, N. Kohli, M. Vatsa, R. Singh, and A. Noore. On detecting atypical faces using ADMOSMIL: Case studies with plastic surgery, disguises, and illicit drug abuse. Under Review, 2018.
- 2.* **D. Yadav**, P. Pandey, N. Kohli, S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore, J. B-Lewis, G. Prabhakaran, and H. Mahajan. Neural correlates of face verification due to facial aging and disguises. Under Review, 2018.
3. N. Kohli, **D. Yadav**, M. Vatsa, R. Singh, and A. Noore. Supervised mixed norm autoencoder for kinship verification in unconstrained videos. *IEEE Transactions on Image Processing*, 28(3):1329-1341, 2019.
- 4.* N. Kohli, **D. Yadav**, and A. Noore. Multiple projective dictionary learning to detect plastic surgery for face verification. *IEEE Access*, 3:2572–2580, 2015.
- 5.* **D. Yadav**, R. Singh, M. Vatsa, and A. Noore. Recognizing age-separated face images: Humans and machines. *PloS one*, 9(12), 2014.

Book Chapters

1. N. Kohli, **D. Yadav**, M. Vatsa, R. Singh, and A. Noore. Deep face-representation learning for kinship verification. *Deep learning in biometrics*, CRC Press, 2018.

Conference Publications

- 1.* **D. Yadav**, N. Kohli, M. Vatsa, R. Singh, and A. Noore. Verifying Age-Separated Faces via Age Gap Reducer-GAN. Submitted to International Joint Conference on Artificial Intelligence, 2019.

2. N. Kohli, **D. Yadav**, M. Vatsa, R. Singh, and A. Noore. Face verification with disguise variations via deep disguise recognizer. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2018.
3. **D. Yadav**, N. Kohli, A. Akshay, M. Vatsa, R. Singh, and A. Noore. Fusion of hand-crafted and deep learning features for large-scale multiple iris presentation attack detection. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2018.
4. **D. Yadav**, N. Kohli, E. Kalsi, M. Vatsa, R. Singh, and A. Noore. Unraveling human perception of facial aging using eye gaze. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2018.
5. **D. Yadav**, N. Kohli, S. Yadav, M. Vatsa, R. Singh, and A. Noore. Iris presentation attack via textured contact lens in unconstrained environment. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1–8, 2018.
6. A. Agarwal, **D. Yadav**, N. Kohli, R. Singh, M. Vatsa, and A. Noore. Face presentation attack with latex masks in multispectral videos. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 275–283, 2017.
7. **D. Yadav**, N. Kohli, S. Nagpal, M. Singh, P. Pandey, M. Vatsa, R. Singh, A. Noore, Gokul Prabhakaran, and Harsh Mahajan. Region-specific fMRI dictionary for decoding face verification in humans. In *IEEE International Joint Conference on Neural Networks*, pages 3814–3821, 2017.
8. N. Kohli, **D. Yadav**, M. Vatsa, R. Singh, and A. Noore. Synthetic iris presentation attack using iDCGAN. In *IEEE/IAPR International Joint Conference on Biometrics*, pages 1–7, 2017.
9. **D. Yadav**, N. Kohli, M. Vatsa, R. Singh, and A. Noore. Unconstrained visible spectrum iris with textured contact lens variations: Database and Benchmarking. In *IEEE/IAPR International Joint Conference on Biometrics*, pages 1–7, 2017.

10. D. Yambay, B. Becker, N. Kohli, **D. Yadav**, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, L. Verdoliva, L. He, Y. Ru, H. Li, N. Liu, Z. Sun, and T. Tan. Livdet iris 2017 - Iris liveness detection competition. In IEEE/IAPR International Joint Conference on Biometrics, pages 733–741, 2017.
- 11.* **D. Yadav**, N. Kohli, P. Pandey, R. Singh, M. Vatsa, and A. Noore. Effect of illicit drug abuse on face recognition. In IEEE Winter Conference on Applications of Computer Vision, pages 1–8, 2016. **[Best Paper Award]**
12. N. Kohli, **D. Yadav**, M. Vatsa, R. Singh, and A. Noore. Detecting medley of iris spoofing attacks using DESIST. In IEEE International Conference on Biometrics Theory, Applications and Systems, pages 1–6, 2016.

Chapter 1

Introduction

Processing and recognizing faces is a fundamental cognitive ability of humans which is essential for identifying and interacting in social space [96]. This cognitive skill makes social interaction extremely effortless and provides an evolutionary advantage to humans as a species [45]. Even though the human face contains discriminatory information for unique identification, it may undergo different transformations due to inter-subject or intra-subject variations. The inter-subject variations can be associated with factors such as race or gender and intra-subject transformations can be attributed to the change in expressions, pose, and temporal variations. It has been established that the face processing network in the human brain is robust to the majority of the real-world transformations such as the change in view angle [42, 101] and illumination variations [70, 118]. This remarkable ability of the visual cortex system has motivated computer vision researchers to utilize these cues as the building blocks to artificially replicate the face processing abilities.

In the computer vision community, developing automatic face recognition algorithm which is robust to intra-subject variations is a challenging research problem. As reported in the Face Recognition Vendor Test conducted by National Institute of Standards and Technology [40], the performance of current face recognition algorithms is remarkable on faces acquired in controlled conditions with limited variations (as seen in Figure 1.1). A thorough evaluation of various commercial face recognition systems revealed that their performance on face images acquired in the *wild* was lower as compared to the controlled

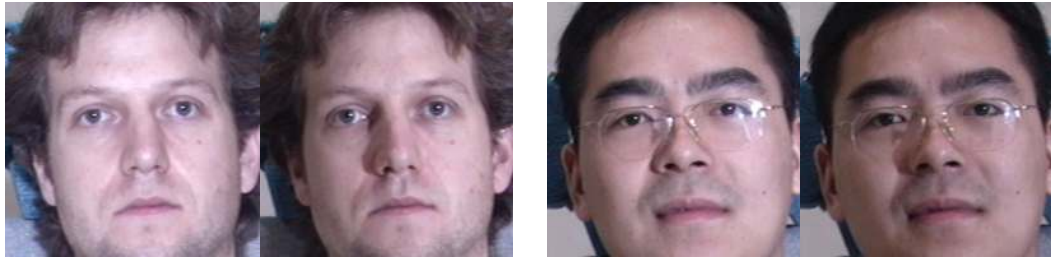


Figure 1.1: Face images of two subjects captured under constrained environment with limited/no wild variations.

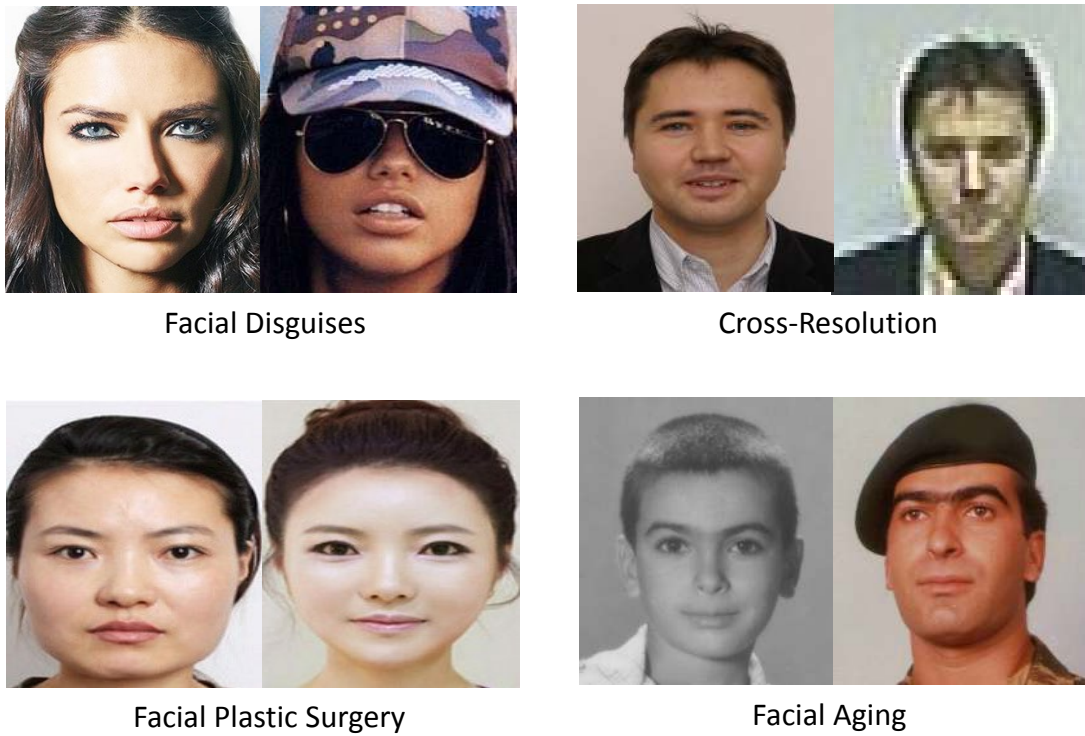


Figure 1.2: Examples of some intra-subjects variations which hamper the performance of face recognition algorithms.

faces. Figure 1.2 showcases some challenges associated with face images captured in the wild. To mitigate the impact of wild variations, a multitude of algorithms have been proposed

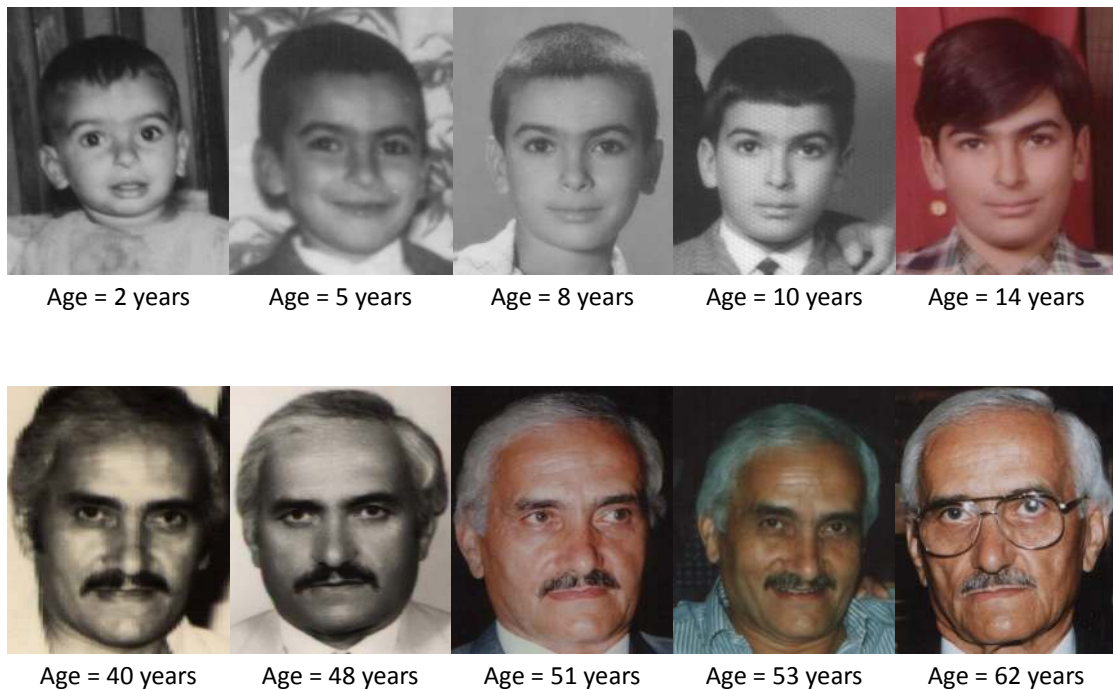


Figure 1.3: Face images of two subjects showcasing the facial aging variations. The top row illustrates the variations during the formative years of the individual. The bottom row shows that more textural variations are evident during the later years.

to handle variations in illumination, pose, expressions [54, 99, 124, 149]. Apart from these traditional covariates, temporal variations (or time-based changes) such as age progression, plastic surgery, and illicit drug abuse also contribute to wild variations.

During the lifetime of an individual, temporal variations alter the facial appearance in diverse ways. Different studies have reported that every person has a personalized aging pattern depending on numerous factors including ethnicity, environmental conditions, and stress level [30, 80]. Moreover, the aging pattern varies across different age groups. During the formative years of a person, the variations in the shape of a face are more prominent while in the later stages of life, texture variations such as wrinkles and pigmentation are more visible [98, 107]. Figure 1.3 shows face images of an individual with age variations which articulate these observations.



Figure 1.4: Sample before and after plastic surgery face images of different subjects.



Figure 1.5: Sample before and after face images of different plastic surgeries illustrating local and global variations in the facial features.

Additionally, temporal/time-based variations are also evident when the face undergoes plastic surgery. Plastic surgery procedures are becoming prevalent due to their low cost and aesthetic appeal. According to the American Society of Plastic Surgeons, more than \$16 billion was spent on cosmetic plastic surgery in 2016 [3]. Figure 1.4 shows the alterations in the facial features caused due to such plastic surgeries. Local plastic surgeries are targeted

towards modifying a specific local facial region in order to enhance its aesthetic appeal or to rectify an anomaly. These surgeries lead to localized alterations but the overall face may appear to be the same. Examples of local plastic surgeries are lip augmentation where lips are made fuller, Rhinoplasty where the nose is altered, and Blepharoplasty where the appearance of eyelids is improved. Global plastic surgeries, on the other hand, result in large regions of the facial structure being modified by altering the facial geometry as well as the skin texture, which may lead to facial rejuvenation. As observed in Figures 1.4 and 1.5, such surgical alterations increase the intra-class distance between before and after face images of the subjects which contribute to a decrease in the face matching performance.

As part of this dissertation, we demonstrated that prolonged illicit drug abuse has the temporal influence leading to significant alterations in the facial features. Illicit drug abuse has become one of the primary health and social concerns in today's world. According to the World Drug Report [4], it is estimated that a total of 246 million people aged 15-64 have used illicit drugs, mainly substance belonging to cannabis, cocaine or amphetamine-type stimulants. The problem of illicit drug abuse is becoming more apparent, considering that 1 out of 10 drug users is a *problem drug user*, who is suffering from drug dependence. Certain drugs, when taken continuously in large quantities, can cause physiological changes in the skin. For instance, the long-term effects of methamphetamine (meth) and heroin can cause severe weight loss and skin sores. Reece [106] noted the evidence of accelerated aging due to the addiction to opiates. A sample of such images is shown in Figure 1.6 where the accelerated aging and formations of scars are very evident. It was also showcased that these physiological variations induced due to extensive substance abuse dramatically decrease the performance of current face recognition algorithms by increasing the intra-class distance between the facial appearance of a subject.



(a) Faces before using illicit drugs

(b) Faces after using illicit drugs

Figure 1.6: Illustrating the significant effect of illicit drug abuse on faces. Noticeable variations can be seen in the facial features of the *after* images of these subjects.

1.1 Problem Statement

In this dissertation, the problem of matching faces with temporal variations is defined as verifying if the given pair of face images belong to the same individual when the faces have been impacted by temporal/time-based physiological variations. In this context, temporal variations can be caused due to age progression, facial plastic surgery procedures, or prolonged illicit drug abuse. The key challenge associated with this problem is the increased intra-class variations as well as numerous factors such as genetics and environmental settings that can impact the facial features. Moreover, this research problem also suffers from the lack of any large-scale labeled face databases which capture these variations in the facial appearance during the lifetime of different subjects.

1.2 Related Work

In this section, we will discuss the relevant studies published in the literature for the three aspects of faces with temporal variations: facial aging, plastic surgery, and illicit drug abuse.

The approaches for face recognition across age progression are categorized as generative and discriminative [68]. Generative methods involve inducing the changes in the input facial images to incorporate aging variations and project the images at a common age. Table 1.1 summarizes age progression algorithms in the literature.

Ramanathan and Chellappa [104] focused on young faces and developed a craniofacial growth based approach which modeled growth related shape variations observed in the formative years. They modeled the facial growth using different growth parameters defined over facial landmarks which are popular in anthropometric studies. Ramanathan et al. [105] proposed an aging model which incorporated facial growth statistics (facial measurements across different ages), demographic information, and alternate wrinkle patterns to estimate other wrinkle patterns that may have been observed on the subject.

Park et al. [93] utilized a 3-dimensional facial aging model to compensate for age variations evident in 2-dimensional face images. They evaluated their approach on three face

Table 1.1: Overview of face age progression (generative) algorithms in the literature.

Authors	Approach	Verification (V) / Identification (I)
Ramanathan and Chellappa [104]	Craniofacial growth based modeling	I
Ramanathan et al. [105]	Craniofacial growth and wrinkle based modeling	I
Park et al. [94]	3D aging modeling technique	V
Wang et al. [127]	Tensor space based analysis	I
Shu et al. [113]	Multiple aging dictionaries for different age groups	V
Sagonas et al. [110]	Expressing face image as a superposition of age and common components	V
Antipov et al. [9]	Age progression with conditional generative adversarial networks	I
Sagonas et al. [111]	Joint and individual variance explained method	V

aging databases to demonstrate its efficacy. Wang et al. [127] simulated facial aging by using super-resolution in tensor space and active appearance model. The synthetic faces generated using this approach are evaluated using different metrics such as perceived age group and identity information of the synthetic image.

Shu et al. [113] learned age-group specific dictionaries for age-separated face verification by generating aging faces. They defined the aging layer to model the aging characteristics and the personalized layer to capture the personalized characteristics of a subject. Sagonas et al. [110] recovered low-rank age and low-rank common components from different age groups. Next, these components are used to age/rejuvenate an input face a in bidirectional manners.

Recently, generative adversarial networks (GANs) are being utilized to generate synthetic images using convolutional neural networks. Due to their popularity, different approaches based GANs have been proposed for facial age simulation. For instance, Antipov et al. [9] used GANs conditioned on age group with identity-preserving optimization. Wang et al. [126] changed faces across different ages by learned the intermediate transition states using a recurrent neural network. However, a major drawback of this approach was that it required

Table 1.2: Overview of age-invariant face recognition (discriminative) algorithms in the literature.

Authors	Approach	Verification (V) / Identification (I)
Ling et al. [69]	Gradient orientation pyramid with SVM	V
Biswas et al. [12]	Scale invariant feature transform with feature drift	V
Li et al. [68]	Multi-feature discriminant analysis with scale invariant feature transform and multi-scale local binary patterns	I
Yadav et al. [139]	Bacteria foraging based fusion	I
Yadav et al. [138]	Human perception based fusion scheme	I
Chen et al. [17]	Cross-age reference coding based approach	I
Li et al. [67]	Local pattern selection descriptor with a hierarchical model based on two-level learning	I
Wen et al. [130]	Latent factor based convolutional neural networks	I

multiple face images at various ages per subject. Zhang et al. [147] used a conditional adversarial autoencoder to learn the face manifold. This face manifold was traversed to realize age progression and age regression simultaneously. Duong et al. [29] developed different approaches for the short-term age progression and long-term age progression. For short-term age progression, they mapped the data densities of two neighboring age groups using ResNet blocks. For long-term age progression, the synthesis was performed by chaining short-term stages. However, it did not consider any personality/identity information which resulted in varying identity in the generated faces. Yang et al. [140] proposed a GAN based approach to progress face images with <30 age to older age groups.

On the other hand, non-generative (discriminative) methods do not involve any changes but they find the age-invariant signatures from the input faces and use it for recognition (summarized in Table 1.2). Ling et al. [69] proposed the use of gradient orientation pyramid to learn a robust representation of faces as they age. In conjunction with support vector machines, they modeled the age-invariant face recognition problem as a two-class problem. Biswas et al. [12] analyzed the drifts of facial features with age progression for the matching

task. Lanitis et al. [62, 63] proposed utilizing the training images for finding the relationship between the coded face representation and the facial age of the subject. This relationship is then utilized for estimating the age of a facial image and simulating the facial appearance at any given age.

In the algorithm proposed by Yadav et al. [139] for face recognition across age progression, they mitigated the effect of facial changes caused due to aging by combining the local binary pattern features of global and local facial regions at match score level, by means of the bacteria foraging fusion technique. Li et al. [68] proposed a discriminative model for age-invariant recognition. They developed an approach involving the use of scale-invariant feature transform, multi-scale local binary pattern as local descriptors, and multi-feature discriminant analysis. Guo et al. [43] studied the relationship between face recognition accuracies and age intervals on MORPH-II, a face database. They observe that when the age gap between the gallery and probe images is more than 15 years, the performance decreases much more as compared to within 15 years. They also observe that the use of soft biometric features can help in improving face recognition across age progression. Chen et al. [17] developed a cross-age reference coding framework to learn the low-level features of a face image with an age-invariant reference subspace.

More recently, deep learning based approaches have been proposed to build age-invariant face recognition systems. Wen et al. [130] developed a new model which constructed a latent identity analysis module to train the convolutional neural network parameters. Li et al. [66] created a new distance metric optimization approach for training the deep-learning framework for age-invariant face recognition. Xu et al. [135] utilized encoders to model the aging and de-aging process.

Next, we present a summary of papers published for matching faces altered by plastic surgery. Table 1.3 summarizes the algorithms published for mitigating the impact of plastic surgery on face recognition. The effect of plastic surgery on face recognition algorithms was first demonstrated by Singh et al. [115, 116]. They identified plastic surgery as a challenging research area in face recognition and also presented the results of face recognition algorithms

Table 1.3: Summary of face recognition algorithms designed for matching faces altered by plastic surgery.

Authors	Approach	Verification (V) / Identification (I)
Singh et al. [116, 115]	Demonstrated the performance of existing face recognition algorithms	I
Bhatt et al. [11]	Used evolutionary algorithm to compute optimal weights for face granules	I
Jillela and Ross [53]	Fused information from the face and the ocular region	I
Aggarwal et al. [5]	Used Sparse representation of part-wise facial regions	I
Sun et al. [119]	Used Structural Similarity Index to model variations	I
De Marsico et al. [24, 25]	Employed fractals and localized version of correlation measure for region-based approaches	I
Gupta et al. [44]	Utilized invariant scattering transform based feature extraction	I
Suri et al. [120]	Combined supervised classifier and task-independent network to encode basic visual cues such as color, shape, and texture	I
Kohli et al. [57]	Used multiple projective dictionary learning to detect plastic surgery for face verification	V

on the plastic surgery database. They observed that six existing face recognition techniques show a significant decrease in the performance if they are applied to plastic surgery database. These findings motivated researchers to develop algorithms which can mitigate the effect of variations in faces due to various plastic surgery procedures.

Bhatt et al. [11] proposed an evolutionary algorithm to compute optimal weights for face granules during the face matching process. Their algorithm focused on extracting discriminatory information from face patches obtained at different levels of granularity. Next, an evolutionary approach using genetic algorithm was utilized to find the optimal feature extractors for each face granule as well as weights for each face granule for face matching. Jillela and Ross [53] proposed a computationally less intensive approach where they fused information from the face and the ocular region. Their proposed scheme improved face

recognition performance without using any training based approach. Aggarwal et al. [5] utilized a sparse representation of part-wise facial regions to tackle variations arising due to plastic surgery. To handle the limited number of images for training, images having facial regions most similar to each subject in the gallery were utilized. Sun et al. [119] used Structural Similarity Index (SSIM) to model variations arising due to plastic surgery procedures. They developed a weighted fusion scheme where different face patches were assigned a weight based on SSIM. They reported a performance comparable with the state-of-the-art algorithm.

De Marsico et al. [24, 25] proposed face recognition against occlusion and expression variations (FARO) and face analysis for commercial entities (FACE) algorithms based on fractals and localized version of a correlation measure to employ region-based approaches for plastic surgery invariant face recognition algorithm. In their results, they reported a recognition rate for different surgeries and demonstrated better performance than state-of-the-art algorithms. As part of this dissertation, we proposed a novel solution to discriminate plastic surgery faces from regular faces by learning representations of local and global plastic surgery faces using multiple projective dictionaries and using compact binary face descriptors [57]. Gupta et al. [44] used invariant scattering transform based feature extraction to learn plastic surgery invariant facial features. They also used principal component analysis and linear discriminant analysis to reduce dimensionality, reduce intra-class variations, and increase inter-class variations. Suri et al. [120] combined supervised learning with a task-independent network to encode visual cues such as color, shape, and texture for matching faces with plastic surgery.

Lastly, we describe the literature related to the third aspect of the face with temporal variations, i.e. illicit drug abuse which is a comparatively newer field of research. In this dissertation, we introduced the Illicit Drug Abuse Face dataset and presented the effect of illicit drug abuse as another challenge of face recognition [137]. Using experimental results, we showed the deterioration in the performance of commercial face recognition algorithms as well as commonly used face descriptors when illicit drug abuse face images were added to the

database of regular faces. Pandey et al. [92] used scattering transform and autoencoder-style mapping function with scattering transform for face recognition under illicit drug abuse variations.

1.3 Research Contributions

This research focuses on developing novel algorithms to mitigate the effect of temporal variations on face recognition accuracy. We focus on three factors related to temporal variations, i.e. facial aging, plastic surgery, and illicit drug abuse. Specifically, the research contributions of this dissertation are:

- **Unraveling and understanding the human skills in processing faces with temporal variations:** An extensive human study is conducted to understand the behavioral and neural correlates of how humans process and match faces with temporal variations. In this research, we investigate which facial cues are utilized by humans for estimating the age of people belonging to various age groups. We also analyze how various facial regions such as binocular and mouth regions influence age estimation and recognition capabilities. We also conduct the first study to utilize fMRI for analyzing the neural correlates of age-separated face verification task by humans to investigate the evidence of a face age-processing module in the brain. These studies are described in-depth in Chapter 2 of this dissertation. The findings from these studies indicate that humans process age-separated faces differently as compared to regular faces. This highlights the need for specialized automatic face recognition algorithms to handle such faces.
- **Matching age-separated face images using the proposed age gap reducer generative adversarial network:** Next, we propose a novel solution for matching faces with temporal variations induced by facial aging. In this, a unified solution is proposed which incorporates facial age estimation and age-separated face verification using GANs. The key idea of this approach is to model the age variations across time by conditioning the input image on the individual's gender as well as the target age group

to which the input face needs to be progressed. Additionally, the training critic simultaneously learns the age group of an input image while estimating how realistic the faces appear. An additional constraint is placed on the loss function of the proposed GAN to account for keeping the consistent identity across age progression by embedding the generated face closer towards the input image in a face space. This algorithm is presented in Chapter 3 of this dissertation.

- **Recognizing faces affected by plastic surgery using multiple projective dictionary learning:** In Chapter 4 of this dissertation, we focus on another aspect of temporal variations on faces, i.e. plastic surgery. We propose a novel solution to discriminate plastic surgery faces from regular faces by learning representations of local and global plastic surgery faces using multiple projective dictionaries and by using compact binary face descriptors. To verify the identity of a person, the detected plastic surgery faces are divided into local regions of interest that are likely to be altered by particular plastic surgery. The cosine distance between the compact binary face descriptors is computed for each ROI in the detected plastic surgery faces. In addition, we compute the human visual system feature similarity score based on phase congruency and gradient magnitude between the same regions of interest. The cosine distance scores and the feature similarity scores are combined to learn a support vector machine model to verify if the faces belong to the same person. We integrate our proposed framework for face verification with two commercial systems to demonstrate an improvement in verification performance on a combined database of plastic surgery and regular face images.
- **Introducing illicit drug abuse as a face recognition covariate with temporal variations:** In Chapter 5 of this dissertation, we introduce the illicit drug abuse face dataset and present the effect of illicit drug abuse as another challenge of face recognition. Experiments are performed to show the deterioration in the performance of commercial face recognition algorithms as well as commonly used face descriptors when illicit

drug abuse face images are added to the database of regular faces. The results clearly demonstrate the need to further study and mitigate the effect of illicit drug abuse on face recognition algorithms. We also propose a detection framework to seamlessly classify in real-time using multiple dictionaries if a given face image is a *regular* face or a *drug-abuse* face and improve the face recognition performance. This framework can act as a crucial pre-processing step in mitigating the effect of such images. The proposed framework gives a classification accuracy of 88.81% when applied on a combined database of illicit drug abuse faces as well as regular faces.

- **Detecting faces affected by temporal variations using deep adaptive multi-one shot similarity based multiple instance learning:** In faces affected by temporal variations, different local regions may be impacted differently. For instance, various plastic surgeries target different local regions to enhance its appearance and different disguises such as sunglasses occlude different local regions. Similarly, for faces affected by illicit drug abuse, sores may appear in the forehead region and wrinkles may be more prominent in the chin area. Therefore, it is imperative to examine the contribution of different local regions for detecting these face images. In this research, we present a novel single-image based algorithm for detecting such faces which incorporates deep features in a multi-instance learning framework to capture the variations across different local facial regions. To encode the diversity in these faces, we incorporate Laplacian Score to determine k representatives from these faces and multi-one shot similarity metric learning approach is employed to determine the similarity of an input face with the k representatives, which increases the discriminability power of the classifier. We utilize the scores from this proposed algorithm as a soft biometric for enhancing the face recognition performance of current systems. This algorithm is described in Chapter 6 of this dissertation.

Chapter 2

Understanding Human Capabilities in Processing Faces with Temporal Variations

In day to day interactions, humans are able to perceive and verify face identities effortlessly. However, with the increasing focus on decoding the face processing network of the human brain, there is a limited emphasis on analyzing the process of face verification. Understanding how the brain perceives and makes a determination if the given two face images belong to the same person or not can provide valuable insight about identity processing mechanism in the brain. Correctly verifying faces may be required in critical scenarios such as law enforcement, airport security, and border control where officers daily match the images in the presented ID or passport to the person in front of them.

Given a pair of face images, if both the face images belong to the same person, the face pair is classified as a genuine; otherwise, it is termed as an imposter (as shown in Figure 2.1). In an ideal case, it is expected that the face images in the genuine face pair should be as similar to each other as possible. However, in real-world scenarios, the genuine pair face images may have variations such as age progression, disguises, expression changes, and variable illumination. Studies have demonstrated that these variations may affect the face recognition performance of humans [26, 89, 90, 138]. Some researchers have explored these covariates independently due to their relevance in high-stakes scenarios as mentioned above.



Figure 2.1: Sample genuine and imposter age-separated face pairs illustrating notable variations between the two images of the respective face pairs.

In this research, we focus on the facial aging aspect of temporal variations to understand the correlates of how humans perform such a complex face verification task. With this aim, two studies are conducted: (i) behavioral study to analyze human skills of age-separated face matching and (ii) fMRI based neuroimaging study to understand the neural correlates responsible for this task. The experiment design and the analysis from both these studies are explained subsequently.

2.1 Behavioral Study to Unravel Human Perception of Facial Aging

The objective of this research is to study the process of facial aging from the perspective of human cognition. As per our knowledge, this research is the first one to analyze various aspects of facial aging such as which facial cues are utilized by humans for estimating the age of people belonging to various age groups, how various facial regions influence age estimation and recognition capabilities.

2.1.1 Methods

Ethics Statement

The study was conducted at Amazon's Mechanical Turk (MTurk), which is an online crowd-sourcing platform. At MTurk, only individuals who are above 18 years of age can register and work as participants. We follow the policies of MTurk which clearly transfer the rights of any survey to the requester and it is informed to the participants at the time of their online registration. The participants' consent to fill and submit the survey is taken as their willingness to participate in our study. Further, at the beginning of the study, we also inform the participants that their responses would be used for research and analysis purposes. All the procedures used in the current study are approved by the IIIT-Delhi Ethics Board.

Participants

Amazon's MTurk is a platform that enables researchers to conduct research by offering features such as a unified participant compensation model, participants having diverse demographics, an efficient procedure of study design, participant enrollment, and data gathering. MTurk allows the researchers or the 'requesters' to post tasks such as surveys, studies, and experiments which are, in turn, completed by the participants or 'workers'. The participants are paid an amount fixed by the requester upon successful completion of the task. Research conducted by Buhrmester et al. [14] on the effectiveness of MTurk suggests that MTurk not only offers a rich pool of diverse participants but can also be used for economically acquiring a large amount of good quality data over a short span of time. In our study, 482 individuals participated, out of which there were

- 366 Indian adults ($M^1 = 33.45$ years, $SD^2 = 11.67$ years, 149 males, 217 females),
- 81 Caucasian adults ($M = 35.39$ years, $SD = 10.74$ years, 43 males, 38 females),
- 29 Asian (non-Indians) adults ($M = 28.13$ years, $SD = 6.93$ years, 6 males, 23 females),

¹M = Mean Age

²SD = Standard Deviation in Age



Figure 2.2: Sample facial regions that are presented to participants for age group estimation.

- 3 African adults ($M = 30.33$ years, $SD = 8.17$ years, 2 males, 1 female), and
- 3 participants with undisclosed ethnicity ($M = 27.12$ years, $SD = 1.7$ years, 1 male, 2 females).

The responses from all the participants have been analyzed in the study in order to preserve the diversity in the responses.

Stimuli

The stimuli faces have been selected from 36 male and 18 female subjects from the FG-Net Facial Aging Database [61] and IIIT-Delhi Facial Aging Database [123]. The creators of FG-Net database have allowed the use of these images for research purposes. Out of a total of 54 distinct subjects, there is an equal number of Indian and Caucasian subjects. The number of images per subject varies from 1 to 4. The chosen images represent the unconstrained nature of real-world conditions.

For evaluation, 10 sets of assignments are created and one set is randomly assigned to every participant. Each set contains three questions.

1. The first question contains five facial images and the participants are asked to estimate the age group from the given face image. Similar to a previous research [15], the age of face stimuli belongs to one of the following 10 age groups: 0 – 5, 6 – 10, 11 – 20, 21 – 30, 31 – 40, 41 – 50, 51 – 60, 61 – 70, 71 – 80, and >80.



Figure 2.3: Sample images presented to the participants for the task of recognizing age-separated images of individuals.

2. Five images of various facial regions such as the T-region, binocular region, chin region, eyes portion masked, and T-region masked are shown to the participants. They have to estimate the age group corresponding to every facial part individually. Figure 2.2 shows some example images that are presented to the participants belonging to each facial region. These images also belong to one of the 10 earlier mentioned age groups.
3. In the last set of questions, five pairs of age-separated images are shown to the participants and they are asked to determine if the pair of images belong to the same individual or not. Some sample images are shown in Figure 2.3.

Procedure

Each participant is randomly assigned one of the 10 sets. The participant is supposed to answer the three questions in the Stimuli section. There is no time constraint on the participant to submit the responses. Each participant sees a face image and an identity only once to ensure there is no bias. In all the questions, a mixture of stimuli from different ethnic groups and ages is presented to each participant.

2.1.2 Results and Discussion

The analysis of responses obtained are classified into four categories and key observations are discussed in this section.

Table 2.1: Confusion matrix showing comparison between actual and predicted age group in the task of age estimation.

Stimuli Age Group	Predicted Age Group									
	0–5	6–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	>80
0–5	211	32	0	1	0	0	0	1	0	0
6–10	8	193	43	2	0	0	0	0	0	0
11–20	0	1	121	104	17	2	1	0	1	0
21–30	0	2	77	110	48	5	0	0	0	0
31–40	0	1	0	8	107	59	12	1	0	0
41–50	0	0	0	0	17	63	51	35	8	0
51–60	0	0	0	1	6	47	105	71	7	0
61–70	0	0	0	0	0	4	31	36	5	0
71–80	0	0	1	2	3	15	56	80	52	15
>80	0	0	0	0	0	6	12	12	3	16

Table 2.2: Analyzing the performance for estimating the age group of shown face stimulus. The results show that highest accuracy is obtained when the stimuli faces are in the range of 0 to 5 years.

Analysis of Perceptual Discrimination of Age Groups by Humans							Auto- matic Algo
Age Group of Face Stimuli	Sensitiv- ity (%)	Speci- ficity (%)	d'	Stimu- lus Entropy (bits)	Noise Entropy (bits)	Info Entropy (bits)	Face++ (%)
0–5	86.12	99.52	2.7500	0.3782	0.0612	0.3171	100
6–10	78.46	97.86	2.3971	0.3790	0.1072	0.2718	100
11–20	48.99	92.80	1.4453	0.3798	0.1735	0.2063	60
21–30	45.46	93.00	1.3422	0.3758	0.1978	0.1780	20
31–40	56.92	94.77	1.6776	0.3275	0.1595	0.1680	80
41–50	36.21	92.13	1.0658	0.3132	0.2127	0.1005	20
51–60	44.30	90.36	1.3085	0.3717	0.2208	0.1509	20
61–70	47.37	89.20	1.3981	0.1839	0.1226	0.0613	20
71–80	23.21	98.59	0.6292	0.3608	0.2030	0.1578	0
>80	32.65	99.20	0.9543	0.1347	0.0856	0.0491	0

Age Group Prediction Accuracy

The responses on predicting age group based on the face stimuli presented to participants are summarized in a stimulus-response confusion matrix shown in Table 2.1. The confusion

matrix is used to determine various performance measures of participants to accurately predict the age group category of the face stimulus shown. The performance is evaluated in terms of:

1. sensitivity,
2. specificity, and
3. information entropy.

Sensitivity (or accuracy) represents the true positive performance [8]. However, it alone may not fully represent the performance of the participants. We are also interested in the performance of the participants in accurately predicting if a face stimulus does not belong to a particular age group. This information can be obtained from *specificity* [8] which represents the true negative performance. Table 2.2 summarizes the sensitivity and specificity values for each age group. It shows that the age groups for which the participants were able to best estimate the face stimuli were age groups 0 – 5 and 6 – 10 with an accuracy of 86.12% and 78.46% respectively. In contrast, the two lowest age group categories that the participants had difficulty in estimating the face stimulus were age groups 70 – 80 and >80 with accuracies of 23.21% and 32.65% respectively. The specificity for these two age groups is 98.59% and 99.20% respectively indicating that participants are highly confident about a face image not belonging to these age groups. These measures provide valuable insights into age prediction judgments by humans.

In response to different visual stimuli, the participants need to make a decision on the correct age group. For each face stimulus shown, the participants have to be able to discriminate one among ten age groups which represent the perceptual judgment of each participant. The strongest response denotes the signal and represents the actual age group while the remaining nine alternatives denote noise or uncertainty distributed among other response categories. The distance between the means of the signal and the noise distributions are compared against the standard deviation of the noise distribution to compute the discriminability index (d') [78, 82]. The d' values calculated for each age group stimulus

is shown in Table 2.2. Higher values of d' signify that the participants are able to discriminate a particular age group category better. From Table 2.2, the results show that participants were able to discriminate the two age group categories 0 – 5 and 6 – 10 better than any other category and the d' values for these correspond to 2.7500 and 2.3971 respectively. It is also observed that the d' values for all age groups are positive, representing that the responses obtained are better than random guesses.

The process of choosing a specific age group based on the visual stimulus presented depends on the information perceived in the stimulus by the participants. The perceived information can be quantitatively represented by the information entropy [85]. The perceptual information may have some residual uncertainty due to noise in the actual stimulus leading to incorrect predictions by the participants. The uncertainty is also introduced when the number of response categories is more. From the stimulus-response confusion matrix (Table 2.1), face stimulus entropy $H(S)$ (Equation (1)) and noise or equivocation denoted by $H(S|r)$ (Equation (2)) are calculated for each age group where S denotes the stimulus, r denotes the response of the participants, and $p(\cdot)$ represents probability of respective terms. Information entropy $I(S|r)$ for each age group category is calculated by subtracting the noise, $H(S|r)$ from the signal, $H(S)$ (Equation (3)).

$$H(S) = - \sum_{j=1}^n p(S_j) \log(p(S_j)) \quad (2.1)$$

$$H(S|r) = - \sum_{j=1}^n \sum_{k=1}^n p(S_j, r_k) \log(p(S_j|r_k)) \quad (2.2)$$

$$I(S|r) = H(S) - H(S|r) \quad (2.3)$$

The values of the stimulus entropy, noise and information entropy for each age group are expressed in *bits* and are summarized in Table 2.2. The larger value of information entropy of an age group indicates that participants can accurately predict the stimulus belonging to that age group as the residual uncertainty is low. The results in Table 2.2 confirm that the two

age groups 0 – 5 and 6 – 10 have the highest information entropy of 0.3171 bits and 0.2718 bits respectively.

Low values of accuracy for older age groups such as 71 – 80 and >80 can be attributed to various factors which affect facial age progression of an individual. The factors including but not limited to gender, ethnicity, stress levels, dietary habits, facial aging patterns of kin combine to form a personalized facial age progression function for each person. Large variances in these factors may lead to an incorrect perception of facial age by humans.

We also compared the human performance with an independently trained automatic algorithm. The same images are evaluated using Face++ [50], a face recognition tool built using deep face representation. An overall age group prediction of 42% is obtained on the same set of images. Upon further analysis, it is observed that images belonging to age groups of 21 – 30, 41 – 50, 51 – 60, and 61 – 70 achieved only 20% accuracy which is lower than responses of participants. Images belonging to 71 – 80 and >80 yielded an accuracy of 0% (none of the images of these age groups were correctly estimated). This suggests that there is a large scope for further improvement in current automated age prediction algorithms, especially if we are able to emulate the way humans perceptually estimate facial age.

Effect of Facial Regions in Age Estimation

For understanding which facial region is most effective for estimating the age group of a given image, five facial regions are presented to the participants and the age is to be estimated on the basis of the given facial region. As shown in Figure 2.2, the five facial regions are T-region, T-region masked, binocular region, eyes portion masked, and chin-mouth region. The results for the same are presented in Table 2.3. It can be observed that the information contained in the chin and mouth regions is sufficient to yield an accuracy of **100%** for infants and toddlers (0 – 5 years age group). The reason for such a high accuracy is based on the fact that the lower jaw region of individuals in this age group is significantly different from other age groups. With the T-region obfuscated, maximum correct responses are obtained for 6 – 10 years age group, indicating that humans can show good performance if the features of

Table 2.3: Analyzing the effect of an individual facial region shown for predicting the age group of the stimuli. The results show that the chin region of children belonging to 0 to 5 years age is most prominent for age prediction.

Age Groups	Facial Region (Accuracy in %)				
	T Region	T-Region Masked	Binocular	Eyes Masked	Chin
0 – 5	50.00	50.00	95.92	76.00	100
6 – 10	76.00	85.42	26.00	77.08	68.75
11 – 20	56.09	32.00	59.09	63.26	23.68
21 – 30	51.02	22.22	54.00	35.42	42.00
31 – 40	32.00	40.81	45.83	50.00	46.81
41 – 50	39.13	55.55	22.22	39.13	39.02
51 – 60	48.84	24.49	24.49	33.33	41.67
61 – 70	46.81	56.01	36.36	57.14	54.00
71 – 80	12.50	53.48	38.09	26.83	14.58
> 80	11.90	19.04	20.83	20.93	20.93

the T-region for this age group are masked. A similar trend is also observed for age groups 41 – 50 and 71 – 80. These results indicate that if one source of information (i.e. facial region) is occluded, the performance of age estimation is not completely degraded [35, 36].

Table 2.4: Face recognition accuracy achieved with respect to stimuli age group and type of facial region shown. The values in the bold show which region is the most discriminating for recognizing the stimuli’s belonging to the given age groups. It can be observed that in general, the whole face yields the highest accuracy whereas, for children and elderly people, binocular and chin regions are the most prominent respectively.

Age Groups	Facial Region (Accuracy \pm Standard Deviation in %)				
	Full Face	Binocular	T Region	T-Region Masked	Chin
(0 – 5, 6 – 10)	60.41 \pm 1.13	67.02 \pm 0.27	59.37 \pm 2.10	33.33 \pm 0.14	50.55 \pm 0.13
(6 – 10, 11 – 20)	81.52 \pm 0.01	69.47 \pm 2.11	76.59 \pm 0.91	69.38 \pm 0.04	66.67 \pm 2.61
(11 – 20, 21 – 30)	87.00 \pm 0.38	68.89 \pm 3.00	67.34 \pm 1.07	65.21 \pm 0.02	43.75 \pm 0.3
(31 – 50, 51 – 70)	76.53 \pm 2.31	54.08 \pm 0.45	63.33 \pm 1.71	57.14 \pm 1.01	59.13 \pm 1.26
(51 – 70, > 70)	70.83 \pm 0.98	55.10 \pm 0.42	72.00 \pm 0.36	66.30 \pm 1.22	80.61 \pm 0.02

Face Recognition across Age Progression

After assessing the ability to estimate the age group, the next step is to understand how efficient humans are in recognizing age-separated images of an individual. As shown in Figure 2.3, the participants are presented with a pair of age-separated images and they are asked to determine if the two images belong to the same individual. The results are described in Table 2.4. The column ‘Stimuli Age Group’ represents the age group of the two presented images.

On analyzing the accuracies for various age group pairs, it can be observed that it is more challenging to identify individuals during the formative years of their lives. The row ‘0 – 5, 6 – 10’ of Table 2.4 shows that the accuracy obtained for these image pairs belonging to the two age groups is lower compared to any other age group. For this pair, the maximum accuracy of 67.02% is achieved for the binocular region. This is the least among the maximum accuracies obtained by all the age group pairs. The results indicate that during this time period, the face of an individual undergoes a significant amount of variations leading to difficulty in recognizing age-separated images. The best performance of 87% is attained when the pair of images belong to ‘11 – 20, 21 – 30’ years of age category.

It can be seen that for the majority of the cases, the maximum accuracy is obtained when the presented pair of images contains the full face of the individual, signifying that humans use the information present in the entire face for recognizing people. *Z*-test of proportions [32] at 95% confidence level, also supports this claim. It is also observed that the binocular region for ‘0 – 5, 6 – 10’ years age group contains invariant features which are required for recognition. In this scenario, the participants achieve an accuracy of 67.02%. Similar performance is observed when the participants are shown age-separated images of lower facial (chin) region belonging to ‘51 – 70 and >70’ years age groups. In order to compare the performance of human evaluation with an independently trained algorithm, the pairs of face image stimuli are evaluated using Face++ [50]. Using the same experimental setting, this tool yields the verification accuracy of 60% at Equal Error Rate (EER) of 40%. It is observed that when the gap between the images is too high or one of the images belongs

to the childhood of the subject, Face++ yields incorrect output. This result suggests that machine learning algorithms can incorporate cues from human perception and improve the accuracy of current face recognition systems.

2.2 Neural Correlates of Face Verification due to Facial Aging

In the previous section, it was showcased that humans are able to accurately verify age-separated images even after the stimuli face images have undergone these progressions. The task of face verification becomes challenging when the age difference between the two genuine face images is large. It is crucial to examine and understand this neural processing network because if this network fails to function, it can cause an inability to verify faces, leading to serious face processing related social impairment [56, 59]. To the best of our knowledge, this is the first research exploring the neural correlates of verification task of faces altered due to aging (gradual temporal changes).

2.2.1 Methods

Participants

Blood Oxygenation Level Dependent (BOLD) fMRI scans of 21 healthy and right-handed participants were collected for this study. The ethnicity of 9 participants was Indian (5 males, 4 females, age range = 18-28 years) and the ethnicity of 12 participants was Caucasian (6 males, 6 females, age range = 18-40 years). The study was approved by West Virginia University Institutional Review Board and IIIT-Delhi Ethics Board. All research was performed in accordance In the previous section, it was showcased that humans are able to accurately verify age-separated images even after the stimuli face images have undergone these progressions. The task of face verification becomes challenging when the age difference between the two genuine face images is large. It is crucial to examine and understand this neural

processing network because if this network fails to function, it can cause an inability to verify faces, leading to serious face processing related social impairment [56, 59]. To the best of our knowledge, this is the first research exploring the neural correlates of the verification task of faces altered due to aging (gradual temporal changes).with relevant guidelines and regulations. The participants were informed about the experimental task and the nature of stimuli. Written consent to participate in this study was obtained from each participant.

Experimental Design

For the purpose of this study, a *stimulus* consisted of a pair of facial images presented together. Two categories of visual stimuli were utilized: *genuine* face pairs and *imposter* face pairs. Genuine and imposter face stimuli pairs were developed using face pairs with age progression referred to as *age-separated face stimuli pairs* representing gradual temporal changes. 120 age-separated frontal images from FG-Net Facial Aging Database [61] and MORPH [108] database were selected to create the age-separated face pairs stimuli set. All the images were converted to grayscale and the background noise (non-face area) was suppressed by applying an elliptical mask on the face stimuli images.

The experiment was designed as a face verification task which involved detecting whether the two face images shown on the screen belong to the same person or not. The participants were asked to respond “Yes” or “No” using a two-button controller provided to them. Each participant was presented 60 face stimuli pairs while lying inside the fMRI scanner. The paradigm used in the fMRI experiment was an event-related face-matching task such that one verification task was performed in a segment of 4.5 seconds. One stimulus pair of face images was shown per segment. Figure 2.4, in each segment, the first face image of the pair was shown for 1 second, followed by both the images being displayed, for a maximum of 2.5 seconds. Participants were instructed to respond within these 2.5 seconds. As soon as a response was recorded, a blank screen was displayed to jitter the stimulus-response interval. The remaining time was treated as the Inter-Stimuli Interval (ISI) where no image was presented. The stimuli were displayed in this manner in order to minimize the head

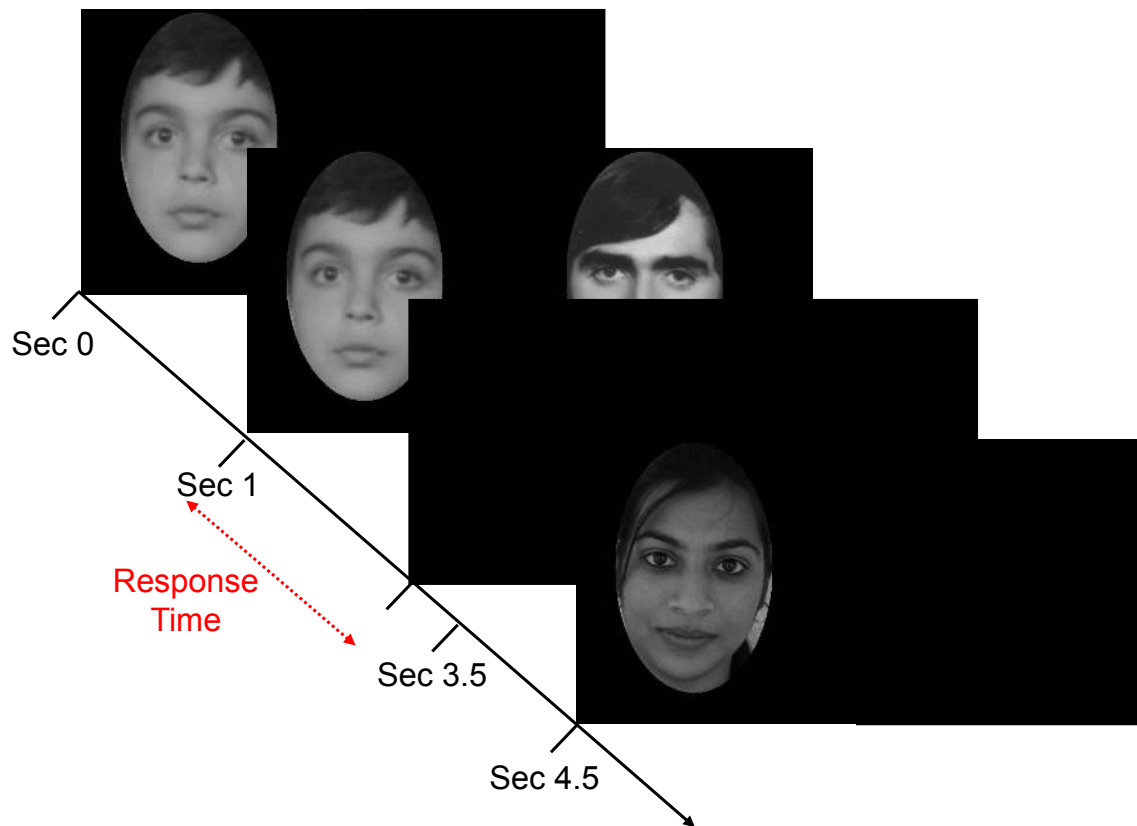


Figure 2.4: Protocol for face verification task. The response time is marked in red and the maximum limit on response time is 2.5 seconds. The screen turned blank after the participant provided response until 4.5 seconds.

movement of participants while matching, emulating real-world face verification scenarios without relying on prior memory like in *n-back* experimental protocol. The age-separated and disguised stimuli face pairs were presented alternatively in 4 runs comprising of such segments.

After fMRI data acquisition, each subject was instructed to participate in a behavioral study where the participant was required to indicate their prior familiarity/unfamiliarity with all the face stimuli pairs shown as some of the stimuli were celebrity face images. These responses were collected outside the scanner as this might lead to noise in the event-based experimental design.

Functional Localizer

One functional localizer [100] task run was designed to isolate face verification specific regions of interest (ROIs) in the human brain. The experimental setup for this localizer was similar to the experimental design described above. 60 full-frontal regular face images were collected from popular face databases [33, 84] to create the stimuli face pairs of this task. 60 face pairs were selected from unique (non-repeating) individuals out of which 39 face pairs were genuine and 21 face pairs were imposter. This experimental setup was also an event-related format where one verification task was performed in a segment of 4.5 seconds with 60 such face verification segments. The participants were instructed to press the “Yes” button if they determined that the displayed face pair stimulus was genuine or to press the “No” button, otherwise.

Data Acquisition

fMRI experiments were performed using a standard 32-channel head coil. The imaging for Caucasian participants was performed using a 3 Tesla (T) Siemens scanner while the imaging for Indian participants was performed using a 3T General Electric scanner. The imaging sequence was an interleaved T2*-weighted echo planar sequence (from negative to positive direction) with 35 axial slices (slice thickness = 3.5 mm, slice spacing = 0.0 mm, repetition time (TR) = 2 seconds, echo time (TE) = 30 ms, flip angle = 65°, field of view = 224 mm, and matrix = 64 × 64) and 128 volumes were captured per run (each volume was captured in 2 seconds with no gap of time between volumes).

High-resolution isometric (1 mm³) anatomical T1-weighted MRI volume scan of the entire head was also obtained for each subject (TR = 600 ms; flip angle = 10°, and field of view = 224 mm).

2.2.2 Preprocessing and Statistical Analysis

The fMRI signals were preprocessed and analyzed using SPM12 toolbox (Statistical Parametric Mapping Toolbox (<http://www.fil.ion.ucl.ac.uk/spm/>)). The functional volumes underwent slice timing correction for assigning all the slices within a single volume to the same time point to account for the acquisition timing differences. Realignment was performed to eliminate the motion artifacts within and between session motion artifacts in each subject. The first image volume was utilized as the reference volume and all the time series images within the subject were aligned with respect to the reference image using a least square minimization and a 6-parameter (rigid body) transformation. The preprocessed data generated during and analyzed during the current study are available from the corresponding author on reasonable request.

In order to compare data across all the subjects included in the study, all the functional images were normalized to the same 3D space. A standard EPI template was used as the reference image to which all the realigned images across all the sessions were matched and then resliced into Montreal Neurological Institute (MNI) space. A Gaussian kernel of FWHM (Full Width at Half Maximum) of 5mm was used for spatial smoothing to remove any noise present in the images. The quality of acquired fMRI data of each participant was measured using Artifact Repair Toolbox (ART) [81]. Next, fMRI data was analyzed using statistical models in SPM based on the General Linear Modeling (GLM) approach at both single subject and group level. The GLM model is expressed as $Y = A\beta + E$, where Y is the observed data i.e., the functional images from the scanner, β is the predicted data, E is the error, and A is the GLM design matrix where A models the experimental conditions responsible for the observed data Y and convolves it with the hemodynamic response function (HRF).

For each subject, a first level model was created and estimated using Restricted Maximum Likelihood approach. A contrast of *Age-Separated Face Pairs vs Baseline* was estimated for each subject. Patterns of significant activation associated with this contrast were identified by appropriately weighting the estimated model using simple t-contrasts and statistical parametric maps (t-maps). Group level analysis was performed to allow inference at the population

Table 2.5: Regional brain activity for age-separated stimuli face pairs. Coordinates are listed according to MNI space. Activations are reported at voxel-level threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ (FDR corrected).

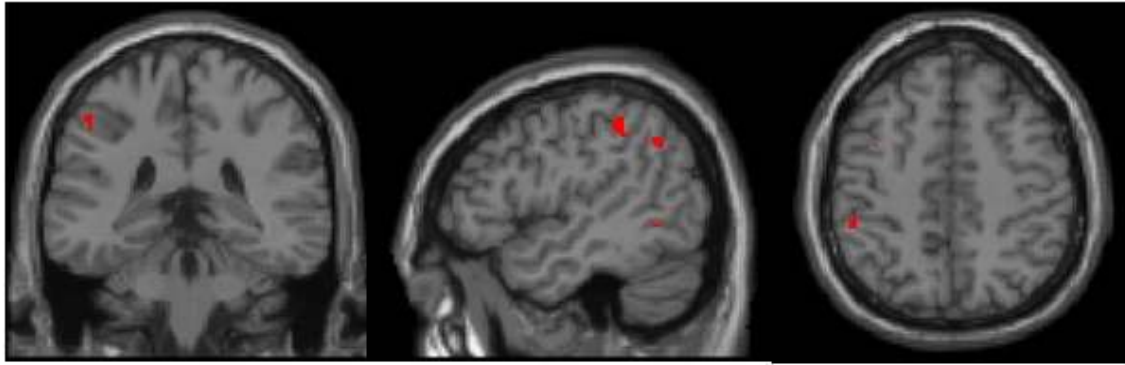
Peak MNI Coordinate (x,y,z)	Brain Regions	Peak Intensity	Num of Voxels (k)
-48, -36, 48	Parietal_Inf_L, Postcentral_L	5.68	23
-46, -56, 38	Parietal_Inf_L	4.72	18
-60, -30, 32	Inf_Parietal_L, Supramarginal_L	4.48	13
-40, -24, -6	Temporal_Sup_L	4.21	8
-46, -56, -6	Temporal_Mid_L, Temporal_Inf_L	4.70	6
-52, -56, 14	Temporal_Mid_L	4.14	6
-42, -62, 30	Angular_L	4.67	5
-40, 6, 52	Frontal_Mid_L, Precentral_L	3.87	5
-2, -58, 56	Precuneus_L	3.98	5

level by computing one-sample second level t-statistic (GLM random effects analysis) using the contrast images from the individual subject analysis. A statistical parametric map of the effect of verifying two face images was generated, thresholded at $p < 0.001$. The coordinates of the statistically significant neural activations are expressed in MNI152 space.

2.2.3 Results

The natural age separated faces represent gradual temporal variations while intentionally disguised faces represent abrupt temporal changes. We analyzed the neuroimaging results for the following contrasts to investigate the neural correlates of face verification of age-separated, i.e. *Age-Separated Face Pairs vs Baseline*. The contrast *Age-Separated Face Pairs > Baseline* was conducted with voxel-level threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ (False Discovery Rate [FDR] corrected). The results are shown in Table 2.5 and Figure 2.5.

Several key findings were observed. We observed that left middle temporal lobe, left inferior temporal lobe, left superior temporal lobe, left IPL, left supramarginal, left angular gyrus, left postcentral, left precentral, left middle frontal gyrus, and left precuneus were significantly active while verifying age-separated face pairs.



Age-Separated Face Pairs vs. Baseline condition.
 Cross-hair points at $x = -48$, $y = -36$, $z = 48$

Figure 2.5: Brain areas showing significant areas of activations in the group analysis for Age-Separated Face Pairs vs Baseline condition. Activations are reported at voxel-level threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ (FDR corrected).

It was seen that while performing age-separated face verification task, there were significant activations in the left hemisphere dominant regions including middle, inferior, and superior temporal lobe, IPL, postcentral lobe, angular gyrus, and precuneus. We observed that left inferior parietal lobe (IPL) played a crucial role in processing age-separated face stimuli. In the literature, the role of IPL has been linked with processing facial stimuli [76] and differentiating self-face from other faces [125]. Singh-Curry and Husain [117] performed a meta-analysis to ascertain the role of IPL in audio and visual vigilance tasks which required detecting small variations which occur at random time instances. As humans are naturally trained by social experiences in detecting and processing variations in facial features due to age progression, significant activation difference in IPL might be linked to visual vigilance task for verifying age-separated stimuli face pairs with temporal variations. Apart from IPL, activations in left angular gyrus may be linked with spatiovisual attention and memory retrieval [150] which may indicate a greater requirement of attention for processing age-separated face pairs. Brain areas such as postcentral gyrus and precentral gyrus may have been activated due to the pressing of the controller button to provide the response. The activations in precuneus may be attributed to its contribution to face perception [58, 97].

2.3 Summary

Faces undergo significant variations during the lifetime on an individual. This research attempts to analyze how humans perceive facial age and their ability to estimate age. The results indicate that age estimation for newborns and toddlers is easiest. The research presents the effect of facial regions such as binocular region, T-region, and mouth on the age prediction accuracy. As a global feature, full face achieves good performance in age-separated face recognition.

We presented the first fMRI study to examine and analyze the neural correlates of processing age-separated face pairs. The findings of this study suggest that verifying the identity of such face pairs may produce cognitive load in the brain. The inferior parietal lobe is found to play a key role in verifying age-separated stimuli faces. It is observed that humans are able to process and verify age-separated faces with high accuracy. The reason for this result may be attributed to social settings where humans gain experience and skill to match age-separated faces. The amalgamation of these findings suggests that regular faces are processed differently from age-separated faces by humans, highlighting the need for building specialized face recognition algorithms for processing such faces.

Chapter 3

Verifying Age-Separated Faces using Age Gap

Reducer GAN

The research in face recognition has witnessed a significant increase in the performance due to the development of different deep learning models. While building these algorithms, the emphasis is on developing an algorithm which is robust to variations such as pose, illumination, expression, and makeup. Another critical challenge of face recognition is matching face images with temporal variations, specifically age gaps. Building age-invariant face recognition algorithms can prove to be beneficial in many applications such as locating missing persons, homeland security, and passport services. In fact, for large-scale applications, adding invariance to aging is a very important requirement for face recognition algorithms.

There are two aspects of building an age-invariant face recognition system: (i) facial age estimation and (ii) age-separated face recognition. Kwon and Lobo [60] are among the first ones to formulate an age estimation approach based on the facial image. They used anthropometry of the face and facial wrinkle density to classify the input image into three broad categories: infants, young adults, and senior adults. Ramanathan and Chellappa [103] proposed an algorithm to estimate the age gap between a given pair of images. Fu and Huang [34] proposed the use of manifold learning to estimate the age. They applied various manifold learning techniques such as Locality Preserving Projections and Orthogonal Locality Preserving Projections to construct a low-dimensional manifold. Recently, several

deep learning based approaches have also been proposed for automatic facial age estimation. For instance, Hu et al. [48] utilized weakly labeled data along with convolutional neural networks to learn the age difference function. Similarly, Liu et al. [73] presented a label-sensitive deep metric learning technique for facial age estimation. Liu et al. [72] used facial features learned from a deep convolutional neural network to perform age estimation.

The other aspect of facial aging is face recognition across aging. Lanitis et al. [62, 63] proposed utilizing the training images for finding the relationship between the coded face representation and the facial age of the subject. This relationship is then utilized for estimating the age of a facial image and simulating the facial appearance at any given age. Park et al. [94] developed a 3D facial aging model to address the problem of age-invariant face recognition. Their approach is based on the fact that exact craniofacial aging can be developed only in 3D domain Li et al. [68] proposed a discriminative model for age-invariant recognition. They developed an approach involving the use of scale invariant feature transform (SIFT), multi-scale local binary pattern as local descriptors, and multi-feature discriminant analysis.

Recently, generative adversarial networks (GANs) [38] are being utilized to generate synthetic images using convolutional neural networks. Due to their popularity, different GAN based approaches have been proposed for facial age simulation. For instance, Wang et al. [126] generated faces across different ages by learning the intermediate transition states using a recurrent neural network. However, a major limitation of this approach is that it requires multiple face images at various ages per subject. Duong et al. [29] developed different approaches for the short-term age progression and long-term age progression. However, it did not consider any personality/identity information which resulted in varying identity in the generated faces. Yang et al. [140] proposed a GAN based approach to progress face images with < 30 age to older age groups. A major drawback of these GAN based approaches is that none of them demonstrate their efficacy in matching age-progressed probe and gallery face images which should be the main objective of such algorithms.

In this research, we propose a unified solution which incorporates facial age estimation and age-separated face verification using GANs. We propose Age Gap Reducer-Generative

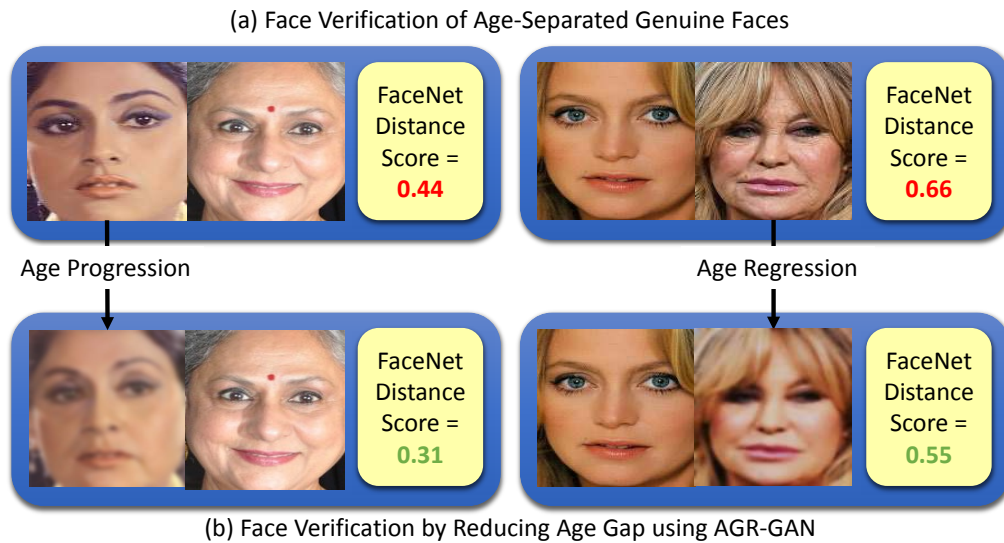


Figure 3.1: Illustrating the problem of age-separated face recognition and the solution proposed in this research using Age Gap Reducer (AGR)-GAN.

Adversarial Network (AGR-GAN) for this task. In the proposed AGR-GAN, the emphasis is on reducing the age gap between face images by using automatic age estimation and incorporating it into the overall loss function. The input image is conditioned on the individual's gender as well as the target age group to which the input face needs to be progressed. Additionally, the training critic simultaneously learns the age group of an input image while estimating how realistic the face appears. An additional constraint is placed on the loss function of AGR-GAN to account for keeping consistent identity across age progression by embedding the generated face closer towards the input image in a lower dimensional face subspace. The key advantage of this approach is that the age progression is bi-directional and given an input image, it can be regressed to an earlier age group or progressed to an older age group (as demonstrated in Figure 3.1).

The contributions of this research are as follows:

- A novel approach for matching age-separated face images is proposed. To accomplish this, a novel GAN architecture, AGR-GAN, is proposed which uses a multi-task discriminator that is able to progress/regress age of an input face to a target age group. Apart from the traditional GAN loss, the proposed AGR-GAN incorporates identity

preserving feature which ensures that the generated (regressed/progressed) face image has the same identity representation as the input face image.

- In the proposed AGR-GAN, joint learning of age group estimator module with the image generation is performed. This novel architecture eliminates the need of age-labeled data in the training phase.
- The efficacy of the proposed AGR-GAN is demonstrated on three different face aging databases for the problem of age-separated face recognition.

3.1 Proposed AGR-GAN

Generative Adversarial Network consists of a generator network (G) and a discriminator network (D) which competes in a two-player minimax game [38]. The aim of D is to distinguish real images from fake images generated by G and the aim of G is to synthesize *real-looking* images to fool D . Specifically, D and G play the minimax game with a value function V such that:

$$\min_G \max_D V = E_{x \sim P_d(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3.1)$$

where z is a noise sample from a prior probability distribution P_z and input face x follows a specific data distribution P_{data} . Upon convergence of the network, the distribution of generated faces P_g should become equivalent to P_{data} . Traditionally, G and D networks are trained alternatively. In the proposed AGR-GAN, the input and output face images are 128×128 RGB images. The input face image is encoded through a representor network to form a low dimension representation enc which learns the higher level features of the input face image. Using the learned enc and conditional information of gender (g) as well as target age group (a), the generator network generates the output face image (x'). Similar to [147], we apply an adversarial loss on enc (D_{enc}) to ensure it is uniformly distributed, thus,

leading to smooth age transformations. Additionally, an adversarial loss, D_{face} , is utilized to ensure that the generated images are realistic looking. Using the target age group a and the estimated age group of the generated face image, the age gap reduction loss is computed to minimize the age gap. Lastly, the feature representations of the generated face and input face image are computed using ϕ face feature mapping to encode identity representations and preserve the identity information after age progression/regression. Next, we provide in-depth details of the different components of the proposed AGR-GAN (as shown in Figure 3.2).

- **Representer:** An input face image is passed through the representer network R whose aim is to learn its low-dimensional representation enc . The input RGB face image x of size 128×128 is sent to the representer network R consisting of five blocks of convolutional layers with stride = 2 and 5×5 convolutional kernels followed by an exponential linear unit layer [20] to learn the facial features which are invariant to age progression/regression. Each convolutional layer is followed by spectral normalization which aids in stabilizing the training phase [83]. After the convolutional layers, a fully connected layer is applied to compute the low-dimensional encoding enc ($R(x) = enc$).
- **Generator:** The objective of generator network G is to utilize enc to synthesize a face image x' with the same gender g , and target age group a . enc is concatenated with the gender label (g) and expected age group label (a). This feature vector is now processed by the decoding part of the generator which consists of a fully connected layer followed by six blocks of transposed convolutional layers with stride = 2 and padding = 2. Each deconvolutional layer is followed by an Exponential Linear Unit (ELU) layer except the last layer which is succeeded by a $tanh$ layer. The output face from the generator is of size $128 \times 128 \times 3$.
- **Discriminator D_{face} :** Similar to traditional GANs, the objective of discriminator D_{face} is to distinguish synthetically generated images by G from real images. It consists of six blocks of convolutional layers with kernel size = 5, stride = 2, and padding = 2. Each convolutional layer is followed by ELU layer. Each convolutional layer is also followed by

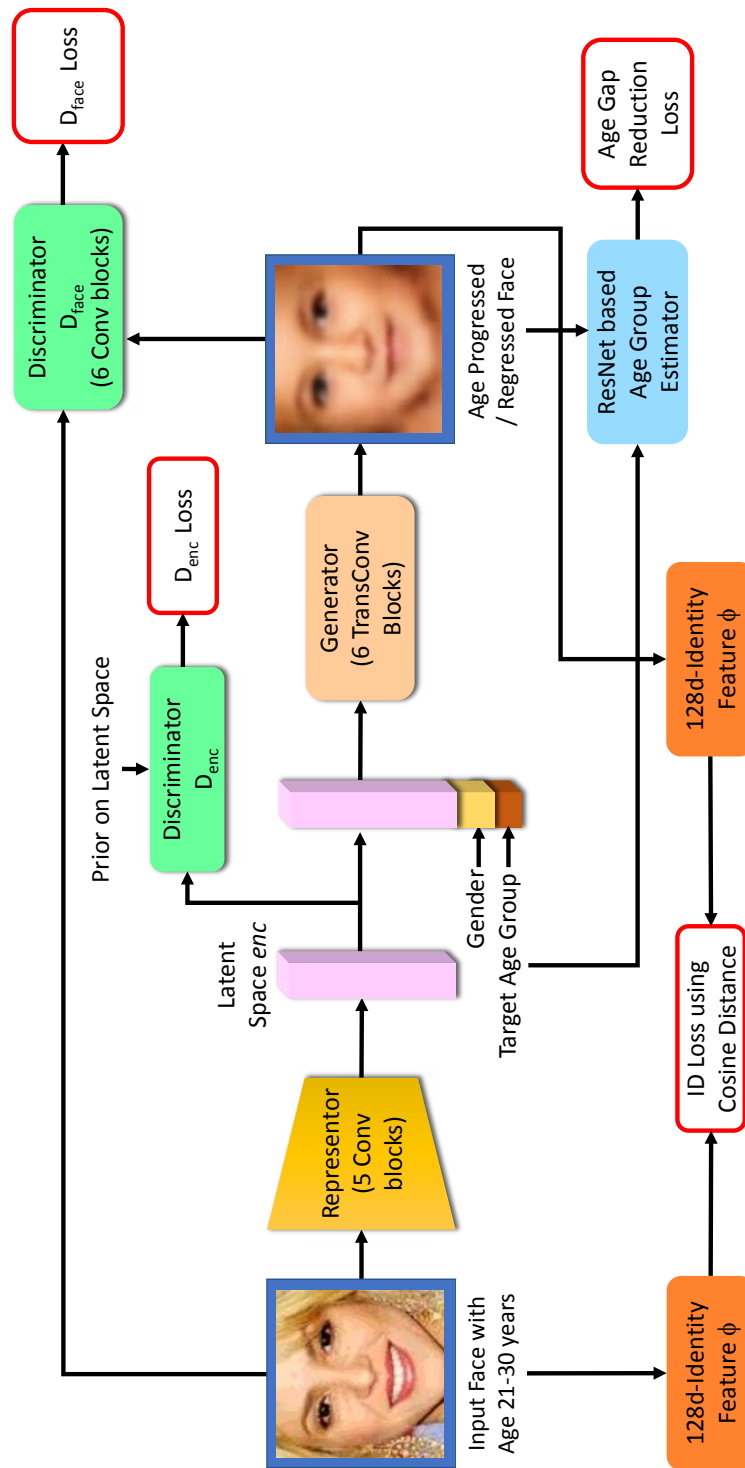


Figure 3.2: Proposed Age-Gap Reducer Generative Adversarial Network (AGR-GAN) architecture.

spectral normalization which aids in stabilizing the training phase [83]. Lastly, a sigmoid layer is utilized to classify the image as real/fake.

- **Discriminator D_{enc} :** The discriminator on enc (D_{enc}) ensures that the distribution of the encoded feature vector enc approaches the prior distribution. The goal of D_{enc} is to distinguish enc generated by R as compared to the prior distribution. On the other hand, R is forced to generate enc such that it can fool D_{enc} . This ensures that enc smoothly populates the low-dimensional latent space to remove unrealistic faces [147].
- **Age Group Estimator:** To reduce the age gap between input face (x) and the generated face image (x'), an age group estimator module is used in the proposed formulation. In this, given an input image, the age group estimator utilizes ResNet-18 model [46] as its backbone to predict the age group of the input image. For this purpose, a pre-trained ResNet-18 model is fine-tuned to predict the correct age group. An adaptive average (spatial) pooling layer is utilized for removing the limitation of the fixed size of the input image and a fully connected layer with output size 10 is employed to predict the final age group. The loss function utilized to train this network is the sum of cross-entropy loss between the correct and predicted age group and mean average group error.

3.1.1 Objective Function

Given an input face image x , the representor network R maps it to a latent space to learn the encoding enc . Given this learned enc , age group a , and gender g , the generator network G synthesizes a face image x' . To ensure that the identity of the generated face image (x') is same as the input face image, we compute the identity-specific feature mapping using function ϕ and minimize the cosine distance $CosDist$ between these two mappings. In this formulation, Light-CNN [134] is used as the identity-specific feature mapping (ϕ). Thus, the

identity loss for this component is:

$$\begin{aligned} & \min_{G,R} \text{CosDist}(\phi(x'), \phi(x)) \\ & = \min_{G,R} \text{CosDist}(\phi(G(R(x), a, g)), \phi(x)) \end{aligned} \quad (3.2)$$

Simultaneously, the age group estimator in G is trained by comparing its output with the ground truth age group label using L_1 loss (age gap loss).

$$\min L_1(a, a') \quad (3.3)$$

With respect to the face image based discriminator D_{face} which is conditioned on the age group a and gender label g , it is trained by the following loss function:

$$\begin{aligned} & \min_G \max_{D_{face}} E_{x,a,g \sim p_{train}(x,a,g)} [\log D_{face}(x, a, g)] + \\ & E_{x,a,g \sim p_{train}(x,a,g)} [\log(1 - D_{face}(G(R(x), a, g)))] \end{aligned} \quad (3.4)$$

Likewise, the discriminator on the encoded feature representation (enc) from the Representator network R is trained to ensure that enc follows the prior distribution using min-max objective function:

$$\begin{aligned} & \min_R \max_{D_{enc}} E_{x \sim p_{train}(x)} [\log(1 - D_{enc}(R(x)))] + \\ & E_{enc^* \sim p(enc)} [\log D_{enc}(enc^*)] \end{aligned} \quad (3.5)$$

To minimize perceptual loss and remove any ghosting artifacts, total variation loss is also computed as follows:

$$\min_{G,R} TV(G(R(x), a, g)) \quad (3.6)$$

By combining these different loss functions, the overall objective function becomes:

$$\begin{aligned}
& \min_{G,R} \max_{D_{enc}, D_{face}} \text{CosDist}(\phi(G(R(x), a, g)), \phi(x)) \\
& + L_1(a, a') + TV(G(R(x), a, g)) \\
& + E_{x,a,g \sim p_{train}(x,a,g)} [\log(1 - D_{face}(G(R(x), a, g)))] \\
& + E_{x,a,g \sim p_{train}(x,a,g)} [\log D_{face}(x, a, g)] \\
& + E_{x \sim p_{train}(x)} [\log(1 - D_{enc}(R(x)))] \\
& + E_{enc^* \sim p(enc)} [\log D_{enc}(enc^*)]
\end{aligned} \tag{3.7}$$

3.1.2 Implementation Details

Before providing the images as input to the network, the following preprocessing steps are performed. Face detection and alignment are performed using pre-trained MTCNN model [144] which produces an output face image of size $128 \times 128 \times 3$. Next, age labels are arranged in bins to create age groups. During the initial years, significant variations are observed in the facial appearance; hence the bins are of size 5. Starting from the age of 20 years, bins (20-30 years, 30-40 years, . . . , ≥ 60 years) are created of size 10. Next, the age group labels and the gender labels are transformed into one-hot encoding. Image normalization is performed to scale the intensity values to the range of $[-1, 1]$. Batch size of 128 is utilized. Adam optimizer with a learning rate = 0.0002 and momentum = 0.5 is used.

3.2 Experimental Evaluation

To demonstrate the efficacy of the proposed AGR-GAN, five experiments are conducted (i) visual fidelity, (ii) aging model evaluation, (iii) identity preservation across generated faces, (iv) ablation study, and (v) age-separated face recognition.

3.2.1 Databases

For evaluation, we utilized MORPH [108], CALFW [148], UTKFace [147], and CACD-VS [17] datasets. MORPH database comprises more than 55,000 age-separated face images of 13,000 subjects. The age range of face images in this database is 16-70 years with the average age = 33 years. Similar to the protocol described in [17], the training set contains face images from 10,000 subjects. The test set is formed using probe and gallery sets from the remaining 3000 subjects and each subject has two images with the largest age gap. CALFW database [148] is built on the concept of the popular LFW face database but with age gaps to add aging as another factor contributing to intra-class variance. It includes 3000 positive face pairs with the age gap and an equal number of negative pairs with 10-fold cross-validation sets. UTKFace dataset [147] is a large-scale database with more than 25,000 images with a long age span of 0 to 116 years. It does not contain labeled identity information and therefore, is utilized for training purposes only. The CACD database released by Chen et al. [17] consists of 160,000 images of 2000 celebrities. Even though the database is very challenging due to wild variations including age, it contains noisy labels. To counteract this, Chen et al. [17] released CACD-VS subset which contains 2000 positive pairs and 2000 negative pairs which are manually checked. For evaluation, CACD-VS database is split into 10 cross-validation folds with 200 positive pairs and 200 negative pairs in the testing.

For training the proposed AGR-GAN, the training folds of MORPH, CACD-VS, and CALFW are combined and the complete UTKFace database is also added to the training set as it does not contain multiple images of subjects. Testing is performed on the test sets of MORPH, CACD-VS, and CALFW databases.

3.2.2 Experimental Analysis

In this section, we describe the different experimental settings and quantitative analysis for evaluating the face images generated by the proposed AGR-GAN architecture.

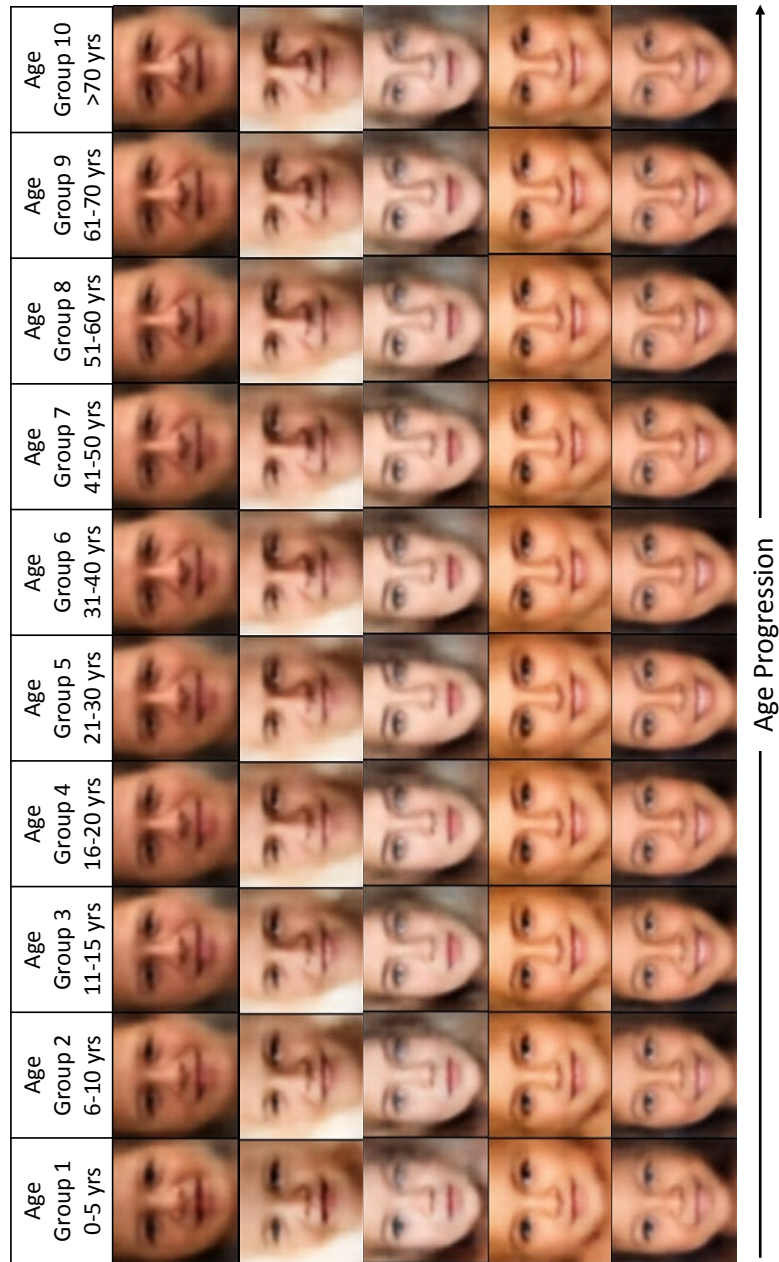


Figure 3.3: Sample generated outputs by the proposed Age Gap Reducer-GAN across different age groups. Each row contains outputs for different subjects from different databases.

Visual Fidelity of Aging Simulation

To evaluate the efficacy of AGR-GAN in synthesizing images across the different age groups, the testing sets of the databases are utilized. Figure 3.3 demonstrates the synthesis output of

Table 3.1: Age estimation (years) of faces generated by the proposed AGR-GAN on (a) MORPH, (b) CACD-VS, and (c) CALFW databases.

Age Group (Age Range)	MORPH	CACD-VS	CALFW
1 (0-5)	5.26	8.45	6.79
2 (6-10)	12.18	11.32	12.38
3 (11-15)	14.32	15.09	14.23
4 (15-20)	17.65	18.94	19.36
5 (21-30)	29.22	27.13	22.71
6 (31-40)	33.51	39.10	35.13
7 (41-50)	47.20	42.59	41.36
8 (51-60)	54.19	53.72	58.75
9 (61-70)	63.69	68.24	63.84
10 (>70)	69.85	74.32	78.38

multiple subjects from different databases across the 10 age groups. It is observed that the proposed GAN architecture is able to learn the aging patterns across different age groups as well as maintain the identity information across different synthesis outputs of the same subject. It is able to model the aging patterns even with varying facial hair, gender, and ethnicity. This illustrates that the proposed AGR-GAN is able to generate photorealistic face images across different age groups.

Aging Model Evaluation

Apart from analyzing the visual fidelity of the generated faces, it is critical to evaluate the ability of the proposed model to produce face images with the targeted age group. For this, we utilize Dex [109], an off-the-shelf age estimation algorithm and predict the age of every synthetically generated face image. The performance is evaluated by analyzing the mean estimated age for each age group and the results are shown in Table 3.1. Ideally, the mean age values for the 10 age group should be in the age range 0-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-50, 51-60, 61-70, and > 70.

For the MORPH database, face images from 3000 subjects are used for testing. The mean values of generated images for the 10 age groups are 5.26, 12.18, 14.32, 17.65, 29.22, 33.51,

Table 3.2: Equal error rate (%) for face verification performance of input test faces and faces from the 10 target age groups generated by the proposed AGR-GAN on (a) MORPH, (b) CACD-VS, and (c) CALFW databases.

Age Group (Age Range)	MORPH	CACD-VS	CALFW
1 (0-5)	12.97	8.23	4.72
2 (5-10)	9.82	8.94	3.37
3 (11-15)	8.65	5.25	3.18
4 (16-20)	6.94	1.27	2.94
5 (21-30)	1.63	0.23	1.83
6 (31-40)	0.48	0.12	1.36
7 (41-50)	1.04	0.03	0.34
8 (51-60)	1.07	0.41	0.81
9 (61-70)	2.38	0.83	0.74
10 (>70)	9.55	1.24	2.72

47.20, 54.19, 63.69, and 69.85 years respectively. Apart from age group 1 (age range 0-5 years), age group 2 (age range 6-10 years), and age group 10 (age range > 70 years), the mean age of GAN generated faces in all the other age groups follows the expected trend. The divergence in the values of age groups 1, 2 and 10 may be attributed to less number of training face images in the training set. Similar trends are also observed for CACD-VS and CALFW databases. These results demonstrate the efficacy of the proposed AGR-GAN in generating face images of the targeted age groups.

Identity Preservation across Generated Faces

Another key aspect of the proposed AGR-GAN is to ensure that the identity/personality information of the subject is preserved across the generated face images. To evaluate this, FaceNet [112], a popular convolutional neural network based face recognition framework is utilized. For each database, the input test image is matched with the corresponding generated image across the 10 age groups and the face verification score is computed. The face verification performance is showcased by calculating the equal error rate (EER) and the results are summarized in Table 3.2.

For all the three databases, the face verification performance of input images with the generated faces is high, as indicated by the low equal error rates in Table 3.2. This result substantiates that the proposed AGR-GAN is able to preserve the identity information in the generated face images and can be used for face matching purposes.

For MORPH database, the best face verification performance is observed when the input face image is matched with the face image generated for the age group 6 (31-40 years) with the equal error rate of 0.48%. This indicates that the proposed AGR-GAN model is able to accurately learn the facial age characteristics of this specific age group. On the other hand, the highest EER of 12.97% is observed for the age group 1 (0-5 years) which indicates the difficulty in matching very young faces. Similar results for this identity preservation experiment are also obtained on CACD-VS and CALFW databases.

Ablation Study

To understand the contribution of different components of the proposed AGR-GAN, an ablation study is performed. For this experiment, MORPH database is utilized and following three cases are constructed: (a) AGR-GAN without age gap loss, (b) AGR-GAN without identity loss, and (c) AGR-GAN without D_{enc} loss. The performance of the proposed GAN architecture is examined using the previously described aging model evaluation and identity preservation evaluation. The results for this experiment are shown in Figure 3.4 and Tables 3.3 and 3.4.

The objective of introducing the age gap loss is to force the model to generate face images with age group as close to the target age group as possible. Therefore, when this loss term is removed from the objective function, the proposed GAN may not produce faces with the target age group. This is evident in Table 3.3. Likewise, when the identity loss is removed from the proposed formulation, the identity preservation property of the network is removed. This may lead to lower face verification scores between the input face and generated faces. The result is shown in Table 3.4. The D_{enc} loss is a uniform prior on the latent space that ensures variations are occurring across the age groups. As is evident from

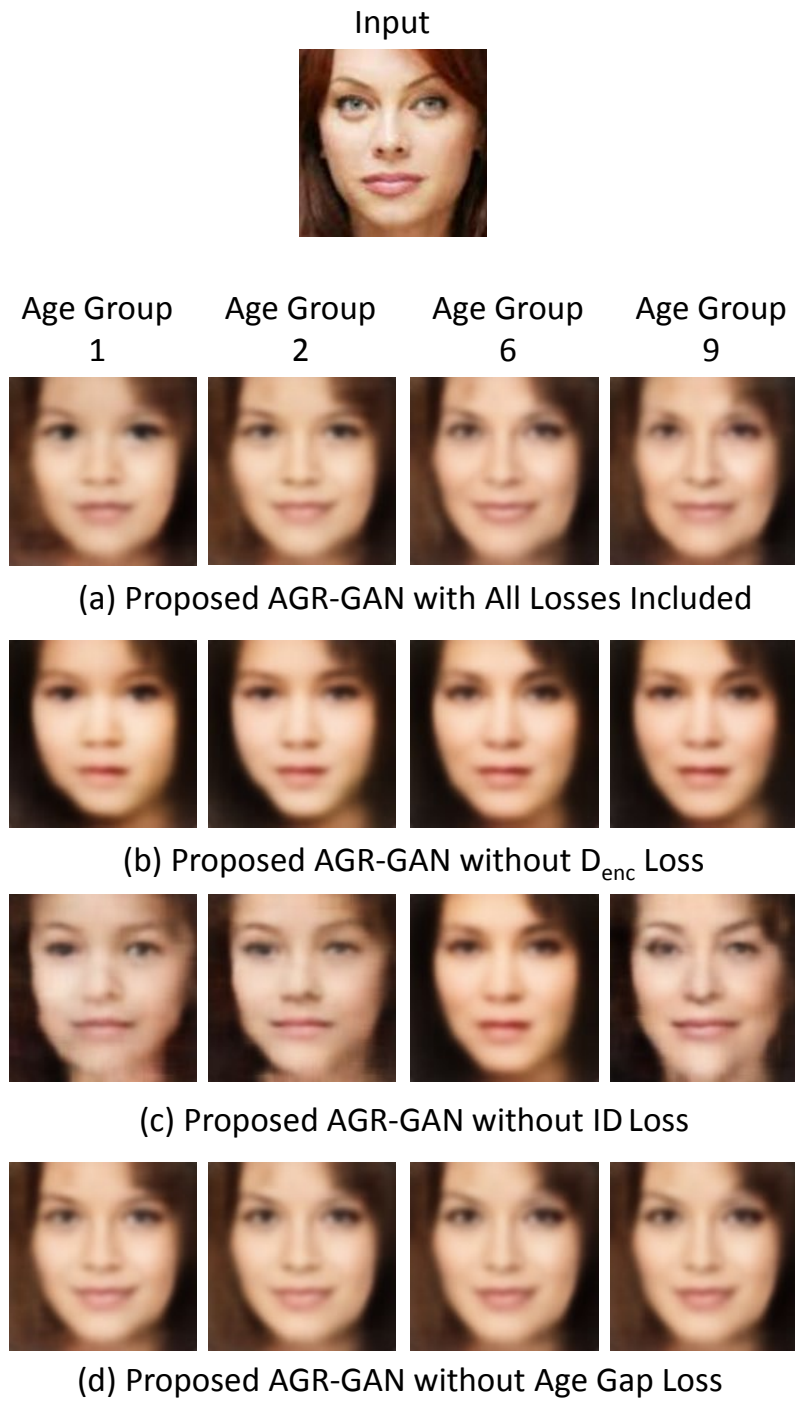


Figure 3.4: Illustrating the findings of ablation study to analyze the effect of removing the three loss functions in AGR-GAN. The age range of the input face image is 21-30 years) .

Figure 3.4, removing this loss leads to images having fewer variations across the different age groups.

Table 3.3: Ablation study using aging model evaluation experiment to analyze the contribution of the age gap loss in the proposed AGR-GAN. Mean estimated age (years) of generated faces is reported.

Age Group (Age Range)	Without Age Gap Loss	With Age Gap Loss
1 (0-5)	39.70	5.26
2 (5-10)	39.57	12.18
3 (11-15)	38.80	14.32
4 (16-20)	35.42	17.65
5 (21-30)	35.68	29.22
6 (31-40)	37.18	33.51
7 (41-50)	38.39	47.20
8 (51-60)	39.92	54.19
9 (61-70)	39.60	63.69
10 (>70)	39.18	69.85

Table 3.4: Ablation study using identity preservation evaluation experiment to evaluate the contribution of the identity loss in the proposed AGR-GAN. Equal error rate (%) of face verification performance is reported.

Age Group (Age Range)	Without Identity Loss	With Identity Loss
1 (0-5)	29.30	12.97
2 (5-10)	31.21	9.82
3 (11-15)	22.19	8.65
4 (16-20)	18.42	6.94
5 (21-30)	19.49	1.63
6 (31-40)	24.23	0.48
7 (41-50)	21.63	1.04
8 (51-60)	20.92	1.07
9 (61-70)	26.73	2.38
10 (>70)	23.03	9.55

Age-Separated Face Recognition Accuracy

The main objective of developing the age gap reducer GAN is to increase the performance of matching age-separated faces. To validate this, face recognition experiment is performed on the three databases. It is to be noted that such an experiment has not been performed in other GAN based age simulation papers.

Table 3.5: Illustrating the increase in face recognition performance of FaceNet by using faces generated from the proposed AGR-GAN on: (a) MORPH, (b) CACD-VS, and (c) CALFW databases.

Database	Metric	Only FaceNet	FaceNet with AGR-GAN
MORPH	Rank-1	94.03	94.15
CACD-VS	Accuracy @ FPR=0.1%	97.50	98.39
CALFW	Accuracy @ FPR=0.1%	57.50	87.15

To demonstrate the efficacy of the proposed AGR-GAN in matching age-separated faces, we utilize FaceNet [112] as the baseline face recognition algorithm. For the MORPH database, the test set containing the youngest image as the gallery and oldest image as the probe is used for performing the face identification experiment. For CACD-VS and CALFW databases, positive and negative face pairs are selected from each fold of cross-validation to perform the face verification experiment. Using the testing sets, we evaluate the performance of the FaceNet algorithm and the results are reported in *Only FaceNet* column of Table 3.5.

Next, AGR-GAN is applied to the input face pair and the query image is projected to the age group of the gallery face image. Using the age gap reduced face pair output from AGR-GAN, FaceNet face recognition algorithm is re-evaluated. Table 3.5 summarizes the result of this experiment. For all three databases, it is observed that utilizing AGR-GAN outputs with FaceNet increases the age-separated face matching performance. The highest increase is noticed for CALFW database (also shown in Figure 3.5) which contains real-world *wild* variations in the images apart from the age gap. The increase in face verification performance is pronounced at the false positive rate (FPR) = 0.1% where the accuracy increases by 29.65% as compared to using only FaceNet. This may be attributed to a significant age gap between the image pairs which is successfully mitigated by the proposed AGR-GAN. As mentioned in [148], VGG-face [95] yields 86.50% verification performance and Noisy Softmax [16] yields 82.52% verification performance at equal error rate on this database. On the other hand, using the outputs produced by the proposed AGR-GAN with FaceNet yields 92.62% face

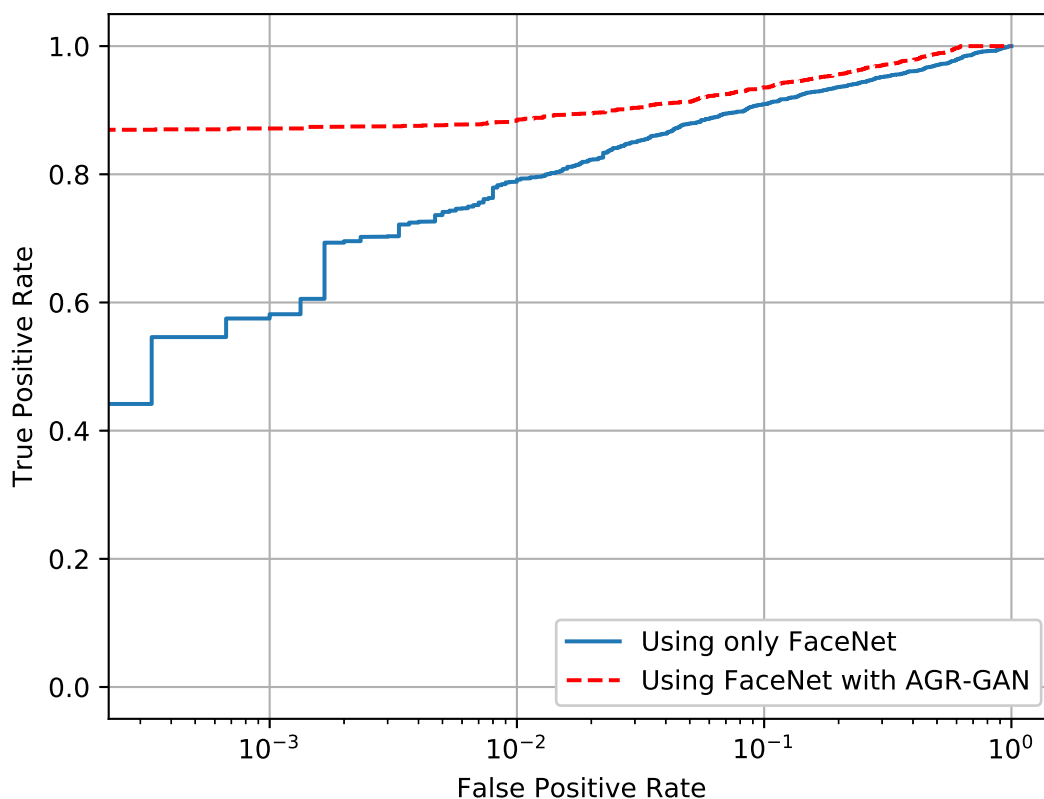


Figure 3.5: Receiver Operating Characteristic (ROC) curves demonstrating the increase in the performance of FaceNet by using AGR-GAN outputs on CALFW database.

verification accuracy at an equal error rate. These results highlight the challenging nature of the database as well as the increase in the performance of FaceNet after using AGR-GAN outputs.

A similar performance increase is also observed on the other databases. On CACD-VS database, the face verification performance at FPR = 0.1% increases by 0.89% after using the outputs from the proposed AGR-GAN. The face identification on the MORPH database shows that the proposed AGR-GAN is able to enhance the rank-1 identification accuracy by 0.12%. This improvement in the face recognition scores illustrates the efficacy of the proposed AGR-GAN in matching faces with age gaps.

3.3 Summary

In this research, we propose a novel solution for matching face images with age gap variation. The proposed architecture utilizes GAN based formulation which involves learning a low-dimensional representation of the face and using this representation conditioned with the target age and gender labels to generate a face image. Apart from the traditional GAN adversarial loss, the training critic also involves age gap loss and identity loss between the input and the generated face images. This novel critic ensures that apart from generating photo-realistic faces, the proposed GAN reduces the age gap and preserves the identity/personality information between the input and the generated face images. Extensive experimental evaluation is performed to validate the effectiveness of the proposed AGR-GAN in matching age-separated face images using different face aging databases.

Chapter 4

Recognizing Faces Altered by Plastic Surgery using Multiple Projective Dictionary

The previous chapter highlights the challenge of identifying individuals with temporal variations induced by facial aging. In this chapter, we focus on another source of temporal variations in faces, i.e. plastic surgery. The transformation of features in a face after plastic surgery can be considered as a covariate that decreases the ability of algorithms to accurately recognize or verify the identity of a person. Recently, the matching of an individual's current face with the photograph in their passport is becoming increasingly challenging, raising questions about the person's true identity, especially at the airports. Most of these incidents are associated with individuals having undergone some plastic surgery procedure [10]. The doctors have started issuing certificates to their patients as proof that they can present to immigration officials when any ambiguity regarding their identity arises. The increasing number of such incidents affirms an urgent need to develop a reliable automatic face verification system which can be seamlessly integrated with any existing face verification system to confront the growing problem of plastic surgery procedures.

In this research, a two-stage plastic surgery detection and face verification framework using Multiple Projective Dictionary Learning (MPDL) is proposed. The idea is to have a separate plastic surgery detection module that tests if a pair of images is *regular face* or

plastic surgery face followed by a region of interest (ROI) based face verification scheme. Specifically,

- A plastic surgery detection scheme is proposed such that it can be integrated with any face verification system. The requirements for such a scheme is that it should be memory efficient and less time-consuming. Both of these goals are obtained through (a) learning binary representations of the face and (b) learning overcomplete projective dictionaries that do not utilize the l_0 and l_1 norm. Due to the increasing number of people opting for plastic surgeries worldwide, current face recognition across plastic surgery algorithms may not handle such massive amount of gallery-probe comparisons. This plastic surgery detection stage will reliably separate possible pair of images as candidates of plastic surgery and perform subsequent operations on this smaller subset.
- A region of interest (ROI) based approach for face verification is presented. The aim of this stage is to identify if the pairs of images which have undergone plastic surgery belong to the same person or not. The facial regions that are likely to be modified due to specific plastic surgery are considered and are compared with the corresponding regions in the reference image. The assumption is that only specific facial features are altered in plastic surgery while the remaining face still contains sufficient discriminatory information to perform verification. This is the first study to mitigate the effect of plastic surgery in the face verification task.
- A key benefit of the proposed MPDL framework is that it can be used in conjunction with any existing face verification system and also improves the performance of verifying individuals who have undergone plastic surgery. The efficacy is demonstrated using two Commercial-Off-The-Shelf (COTS) system by studying the performance of COTS in regular face matching, face matching in a combined database of plastic surgery and regular faces, and seamlessly integrating the results of MPDL in COTS.

4.1 Multiple Projective Dictionary Learning Framework for Face Verification

The MPDL framework is shown in Figure 4.1 and consists of the steps below:

1. Preprocess the images so that they are geometrically aligned and normalized.
2. Detect if the face image pair has undergone plastic surgery.
3. Use the output from the previous stage to verify the identity of the individual.

Each step is described in detail in the subsequent sub-sections:

4.1.1 Preprocessing of Face Images

All the face images are converted to grayscale and landmark detection is performed using the two-stage cascaded deformable shape fitting method introduced by Yu *et al.* [143]. The algorithm uses procrustes analysis by introducing 3D face shape leading to pose-free landmark identification. The automatically provided landmarks are used for preprocessing using the CSU Face Identification Evaluation System [13] to generate normalized images. The system performs geometric normalization followed by masking so that only the face region is visible. It performs histogram equalization followed by pixel normalization such that the mean pixel value is zero and the standard deviation is one.

4.1.2 Proposed Dictionary based Detection of Plastic Surgery

Stage-1 in the proposed MPDL framework detects if plastic surgery has been performed between the given pair of images. The motivation of this stage is to reliably separate pairs of face images where no variation due to plastic surgery is detected. For such pairs, commercial face verification algorithm can be directly employed for identity verification.

In image processing and pattern recognition, dictionary learning has gained popularity in recent years. Dictionary learning based techniques have been shown to achieve superior

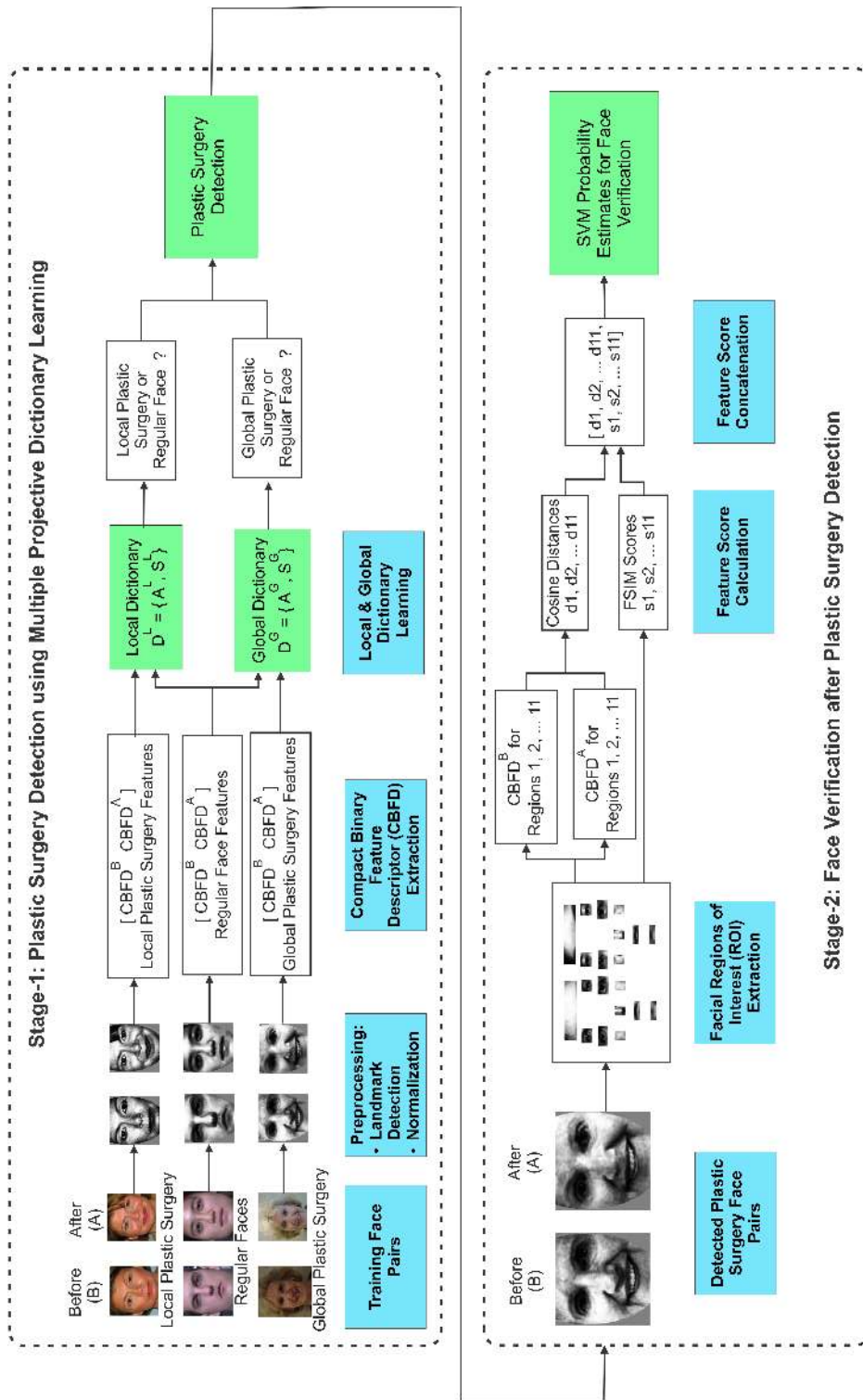


Figure 4.1: Proposed two-stage Multiple Projective Dictionary Learning (MPDL) framework to detect plastic surgery for face verification.

performance in areas such as pattern classification [41], image classification [142], and several covariates of face recognition [91, 146]. There are different approaches to dictionary learning such as sparsely represented dictionaries [6], sparse representation based synthesis dictionary that uses l_p norm with $p \leq 1$ [133], supervised dictionary learning methods that utilize the class labels [52, 79], and structured dictionaries that lead to discrimination between classes [129, 141]. However, most of these approaches utilize the l_0 or l_1 norm which is computationally intensive.

Gu *et al.* [41] proposed a projective dictionary pair learning (DPL) framework, where they propose joint learning of analysis and synthesis dictionaries to learn representations through linear projection without using the non-linear sparse encodings. The model can be described as:

$$\begin{aligned} \{A^*, S^*\} = \arg \min_{A, S} & \sum_{k=1}^K \|X_k - S_k A_k X_k\|_F^2 \\ & + \beta \|A_k \bar{X}_k\|_F^2, \quad s.t. \|d_i\|_2 \leq 1 \end{aligned} \quad (4.1)$$

where S represents the synthesis dictionary used to reconstruct X ; A represents the analysis dictionary used to code X ; A_k and S_k represent the sub-dictionary pair corresponding to class k ; \bar{X}_k represents the complementary data matrix of X_k in the training set; $\beta > 0$ is a scalar constant that denotes the regularization parameter to control the discriminative property of A , and d_i denotes the i th item of synthesis dictionary S . The role of the analysis dictionary A is to help in discrimination, where the sub-dictionary A_k can project the samples from class $i, i \neq k$ to a null space. The role of the synthesis dictionary S is to minimize the reconstruction error.

The advantage of the above framework is in its computation time since the framework does not contain any l_0 or l_1 norm. Utilizing the DPL model in Stage-1 is ideal because the first stage should be considered as a preprocessing step which should be able to efficiently and quickly determine if plastic surgery has been performed or not.

As mentioned previously, plastic surgeries can be of two types: local and global. The effect of local plastic surgery is concentrated on a localized facial region, while the effect of global plastic surgery is more holistic in nature. We propose to learn two separate dictionaries \mathbf{D}^L and \mathbf{D}^G for local and global plastic surgeries since the effect of these surgeries is different from each other. Therefore, $\mathbf{D}^L = \{\mathbf{A}^L, \mathbf{S}^L\}$ simultaneously learns the changes between local plastic surgeries and regular faces; and $\mathbf{D}^G = \{\mathbf{A}^G, \mathbf{S}^G\}$ simultaneously learns the changes between the global plastic surgeries and regular faces. The proposed structure of the MPDL is equivalent to an overcomplete dictionary model since regular faces in training both the dictionaries \mathbf{D}^L and \mathbf{D}^G are the same. These non-disjoint shared training images make the proposed MPDL model more robust and discriminatory in nature.

In the literature, there has been great interest to learn binary representations of the face because they are efficient and computationally less intensive. In this research, Compact Binary Face Descriptor (CBFD) introduced by Lu *et al.* [75] is utilized. The advantage of using these features is that they are learned directly from the raw pixels, where all the redundant information is removed leaving a compact binary descriptor. In the case of plastic surgery, the textural changes occurring in a local facial patch due to a surgical process would be encoded into a binary vector, which should be distinct as compared to an unaltered patch.

The algorithm first calculates the pixel difference vectors (PDVs) in local patches by computing the differences in intensities between each pixel and its neighboring pixels. Then a feature mapping is learned to project these vectors into low dimensional binary vectors such that its variance is maximized, quantization loss between the encoded vectors and the original feature vector is minimized, and the feature bins are evenly distributed.

Let $X = [x_1, x_2, \dots, x_N]$ represent the N PDV samples of the training set calculated across all the training images and let x_i represent the i th PDV where $1 \leq i \leq N$. Then the goal of the algorithm is to learn K hash functions and convert each x_i into a binary vector $b_i = [b_{i1}, \dots, b_{iK}] \in \{0, 1\}^{1 \times K}$. The k th binary code b_{ik} of x_i can be computed as:

$$b_{ik} = \frac{1}{2} \times (\text{sgn}(w_k^T x_i) + 1) \quad (4.2)$$

where $\text{sgn}(v)$ equals to 1 if $v \geq 0$ and -1 otherwise; and $w_k \in R^d$ is the projection vector for k th hash function.

To learn the feature matrix $W = [w_1, w_2, \dots, w_K]$, the following objective function is used:

$$\begin{aligned} \min_{w_k} J(w_k) = & - \sum_{i=1}^N \| b_{ik} - \mu_k \|^2 \\ & + \lambda_1 \sum_{i=1}^{N=1} \| (b_{ik} - \frac{1}{2}) - w_k^T x_i \|^2 \\ & + \lambda_2 \| \sum_{i=1}^{N=1} (b_{ik} - \frac{1}{2}) \|^2 \end{aligned} \quad (4.3)$$

where N is the number of PDVs extracted, μ_k is the mean of the k th binary code, λ_1 is the parameter to balance the effect of quantization loss between the bins and the original samples, and λ_2 is the weight assigned for even distribution of the bins.

After learning the feature matrix W , the PDVs are projected into the low dimension space and are quantized using k -means. The codebook learned from the training data, is then used for projecting the testing vectors to the new space. Given a face image \mathbf{X}_i , non-overlapping patches are extracted from it and a histogram representation is learned for each patch. The final vector is a concatenation of all CBFDF feature vectors calculated across local patches.

Let b_i^B represent the compact binary face descriptor of a pre-surgery (before) face and b_i^A represent the compact binary face descriptor of post-surgery (after) face. In the experiments, the pre-surgery and post-surgery binary feature representation $F = [b_i^B \ b_i^A]$ are concatenated to learn the proposed dictionaries. Both of these dictionaries, \mathbf{D}^L and \mathbf{D}^G , can be learned using Equation 5.2 where the training data is provided for local plastic surgeries, global plastic surgeries, and regular faces where no surgery has been performed.

Let $F_y = [b_y^B \ b_y^A]$ represent the concatenated representation of a testing face identity where the b_y^B is the binary vector representation of the template and b_y^A represents the binary vector representation of the test face claiming to be identity y . Then the two learned dictionaries will evaluate whether local plastic surgery and global plastic surgery is performed or not using

Equations 4.4 and 4.5 where i represents the class specific information within the learned dictionaries.

$$Y_{Local} = \operatorname{argmin}_i || F_y - \mathbf{S}^L_i \mathbf{A}^L_i F_y ||_2 \quad (4.4)$$

$$Y_{Global} = \operatorname{argmin}_i || F_y - \mathbf{S}^G_i \mathbf{A}^G_i F_y ||_2 \quad (4.5)$$

The final decision of whether a plastic surgery has been performed (Y_{PS}) is denoted by an either-or operation on the values represented as $Y_{PS} = Y_{Local} \vee Y_{Global}$

4.1.3 Face Verification after Plastic Surgery Detection

After Stage-1 has identified that plastic surgery has been performed between the pairs, the next stage of the MPDL framework distinguishes between genuine and impostor pairs. Some previous studies [5, 24, 53] based on localized approaches for face recognition across plastic surgery have used specific face areas such as ocular regions, nose, mouth, and forehead. However, no study has focused on including all the facial regions affected by all the surgeries included in the database by Singh et al. [115].

In Stage-2 of the proposed MPDL, the specific regions of interest that will change due to a particular type of plastic surgery are targeted. The assumption is that specific facial features are modified in plastic surgery while the remaining face features still contain discriminatory information to perform verification. Based on related literature on various plastic surgeries, 11 ROIs are used that are likely to be modified because of a specific plastic surgery as described in Table 4.1 and shown in Figure 4.2.

For a given pair of face images, 11 facial ROIs are extracted and Cbfd features for each ROI are obtained. The *cosine* distance between the corresponding pre-surgery and post-surgery regions of interest is computed to obtain the distance among the 11 corresponding regions across the testing pair. The second set of features based on feature similarity

Table 4.1: Regions of interest (ROIs) affected by different plastic surgeries.

Targeted Facial Region of Interest (ROI)	Related Plastic Surgery
Nose	Rhinoplasty
Left Eye	Blepharoplasty
Right Eye	Blepharoplasty
Left Cheek	Cheek Surgery, Rhytidectomy
Right Cheek	Cheek Surgery, Rhytidectomy
Chin	Chin Surgery
Lips	Lip Augmentation
Left Eyebrow	Eyebrow lift
Right Eyebrow	Eyebrow lift
Forehead	Local Surgeries such as Botox
Full Face	Global Face Surgery

(FSIM) [145] descriptor is extracted across the same 11 ROIs to encode for structural and informational changes.

The feature similarity metric [145] (FSIM) is based on the fact that the human visual system focuses on low-level features. It utilizes the phase congruency (PC) and image gradient magnitude (G) for computing the final quality score. It is based on the assumption that the regions where the Fourier waves extracted at different frequencies have congruent phases are the most feature-oriented regions in an image. The FSIM between two images I_1 and I_2 can be written as:

$$FSIM(I_1, I_2) = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (4.6)$$

where x refers to a location in the image, PC_m represents the weighing parameter of the similarities across locations, $S_L(x)$ represents the product of the similarity between the phase congruency of the two images ($S_{PC(x)}$) and similarity between the gradient magnitude of the two images ($S_{G(x)}$), and Ω represents the image domain. Mathematically the terms are defined as,

$$PC_m = \max(PC_1(x), PC_2(x))$$

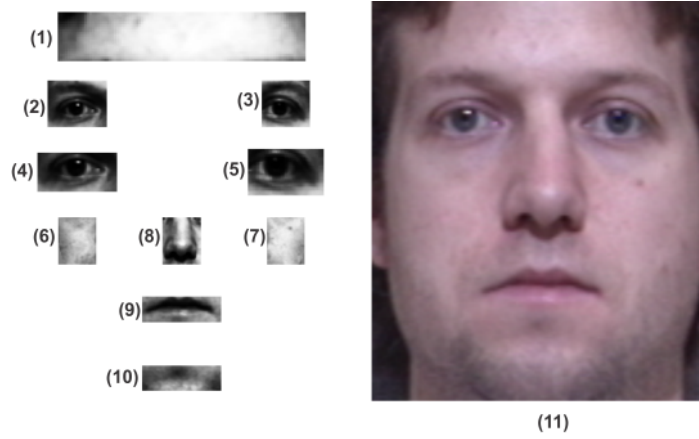


Figure 4.2: Facial regions extracted from pre-surgery and post-surgery images to perform face verification. The regions are (1) forehead, (2) right brow, (3) left brow, (4) left eye, (5) right eye, (6) left cheek, (7) right cheek, (8) nose, (9) lips, (10) chin, and (11) full face.

$$S_L(x) = S_{PC}(x) \cdot S_G(x)$$

where

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1}$$

and

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2}$$

where $G_1(x)$ and $G_2(x)$ are the computed gradient maps for images I_1 and I_2 , $PC_1(x)$ and $PC_2(x)$ are the phase congruency maps of the images I_1 and I_2 , and T_1 and T_2 are positive constants.

The ROI-based Cbfd *cosine* distance scores and FSIM scores are concatenated to form a feature vector and are used for training a Support Vector Machine (SVM) to obtain probability estimates for genuine and impostor classes.

4.2 Experimental Protocol

Datasets Used

To evaluate the performance of the MPDL framework, two types of face datasets are used: plastic surgery face database [116] and regular face database. Singh *et al.* [116] released the dataset containing pre-surgery images and post-surgery images of 900 individuals. In the database, there is one pre-surgery image and one post-surgery image for each individual. Images belonging to 19 subjects are discarded because of the presence of the same persons with different IDs in different categories of plastic surgery. The number of plastic surgery faces that belong to each of the 10 plastic surgery categories is summarized in Table 4.2. To simulate real-world conditions and perform a comparative analysis, an equal number of regular face images are added. The combined database, containing plastic surgery and regular faces, is created by merging several publicly available and widely used face datasets. Neutral expressions, frontal pose, and properly illuminated images are selected from visible spectrum images of CASIA NIR-VIS 2.0 [64] and Multi-PIE [39] face databases. For each subject, two images are selected for verification purposes to keep it consistent with the one before and one after plastic surgery face images.

Table 4.2: Number of image pairs used for different surgeries from plastic surgery face database [116].

Plastic Surgery	Number of Image Pairs
Rhinoplasty	174
Eyebrow Lift	56
Otoplasty	72
Blepharoplasty	105
Cheek Surgery	19
Rhytidectomy	305
Chin Surgery	6
Lip Augmentation	7
Local Surgeries such as Botox	78
Global Surgeries	59

Experiments Performed

Using images from plastic surgery and regular face databases, an equal number of genuine and impostor pairs are created. The impostor pairs are randomly created using the post-surgery images. For the experiments, unseen training and testing is performed with five times random cross-validation. In each run, 50% of the image pairs are used for training and the remaining 50% are used for testing. Also, the pairs belonging to different surgeries are equally divided among the training and testing pairs. The experiments are run on a desktop PC with 3.4 GHz Intel CPU and 16 GB memory.

- **Stage-1 Plastic Surgery Detection:** After performing preprocessing on training images from the combined database of plastic surgery and regular face images, CBFs are extracted for both pre-surgery and post-surgery face images of each pair. The CBFs of each pair are concatenated to form one feature vector. The overcomplete dictionaries D^L and D^G are learned and are used to predict whether the images have undergone plastic surgery or not.
- **Stage-2 Face Verification on Plastic Surgery Detected Pairs:** After a pair of reference and query images is detected as plastic surgery in Stage-1 of plastic surgery detection, the next task is to verify the identity of the individual in question. For this, a model is trained using genuine and impostor pairs from plastic surgery training image pairs. The 11 ROIs are extracted, as described in Table 4.1 and shown in Figure 4.2, from both the pre-surgery as well as post-surgery face images. For each ROI, CBF feature is calculated. The *cosine* distance between the two vectors is computed. Along with this, the feature similarity score between the same ROIs is calculated using FSIM [145]. Using the calculated measures as features, a SVM is trained with genuine and impostor class labels.

4.3 Results and Analysis

4.3.1 Parameter Setting

The parameters are tuned empirically for computing Cbfd and learning the projective dictionary.

- For Cbfd, a dictionary of size 1000 atoms with 10×10 non-overlapping patches is used to obtain the highest plastic surgery detection accuracy. It is observed that the accuracy increases as the number of patches for the images are increased. The size (K) of the binary vector $b_i = [b_{i1}, \dots, b_{iK}]$ is adjusted as well and the best performance is observed for $K = 25$. While computing Cbfd, whitened PCA is used for dimensionality reduction. By experimental analysis, top 600 coefficients are chosen.
- For projective dictionary learning, a dictionary of size 600 atoms is found to give the highest plastic surgery detection performance. The parameter β given in Equation 1 is a regularization parameter to control the discriminative property of the analysis dictionary A . By parameter testing, this value is set to 1×10^{-4} to obtain the best results for plastic surgery detection.

4.3.2 Plastic Surgery Detection

Average classification accuracy of detecting if a face image has undergone plastic surgery or not is reported in Table 4.3. Also, a comparison of currently used techniques such as Local Binary Patterns (LBP) [86], Histogram of Oriented Gradients (HOG) [22], and Uniform Circular Local Binary Patterns (UCLBP) [87] with the Stage-1 of the proposed MPDL is shown.

From Table 4.3, it is observed that the Stage-1 of MPDL yields classification accuracy of 97.96%. It outperforms existing approaches shown in Table 4.3 where conventional feature descriptors are used for the same task. When UCLBP [87] is used as a feature descriptor with

Table 4.3: Accuracy (%) for plastic surgery detection using different classification algorithms.

Classification Algorithm	Accuracy (%)
UCLBP [87] + SVM	82.15
LBP [86] + SVM	84.65
HOG [22] + SVM	86.32
TPLBP [131] + SVM	87.12
Label Consistent K-SVD [52]	92.25
Proposed MPDL Plastic Surgery Detection (Stage-1)	97.96

SVM as the classifier, an accuracy of 82.15% is observed. Similarly, LBP [86], HOG [22], and TPLBP [131] give classification accuracy of 84.65%, 86.32%, and 87.12% respectively.

To compare the performance of Stage-1 of MPDL with other dictionary learning techniques, label consistent K-SVD [52] is also implemented. Plastic surgery detection accuracy of 92.25% is observed. The time taken by this technique to learn the dictionary is 1249.50 seconds as compared to 20.92 seconds for the proposed multiple projective dictionary learning.

Upon analysis of the output of Stage-1 of the MPDL framework, it is observed that the proposed plastic surgery detection algorithm correctly detected 97.50% of plastic surgery pairs (true positive rate). On the other hand, 98.41% of regular face pairs are also correctly identified as ‘no plastic surgery’ (true negative rate) by the algorithm.

When further analysis of these classifications is done based on the type of plastic surgery that is performed on the testing pair, we observe the results shown in Table 4.4. 100% of testing image pairs belonging to cheek surgery and chin surgery are classified correctly by the proposed plastic surgery detection algorithm. It is to be noted that the plastic surgery database contains a comparatively fewer number of images from cheek and chin surgeries, as shown in Table 4.2. Face pairs from Rhytidectomy are detected with 99.67% accuracy and eyebrow lift surgery are detected correctly with 97.14% accuracy. The lip augmentation surgery has the lowest detection accuracy of 80.00%. A possible explanation of lip augmentation showing the lowest performance could be high variability in the shape of lips due to smiling and other expressions during image capture. Also, in the local surgeries category, there is a wide variety of procedures such as Botox, Dermabrasion and scar removal which are included.

Table 4.4: Plastic surgery detection accuracy based on different plastic surgeries using Stage-1 of the proposed MPDL framework.

Types of Plastic Surgery	Accuracy (%)
Cheek Surgery	100.00
Chin Surgery	100.00
Rhytidectomy	99.67
Rhinoplasty	97.93
Global Face Surgery	97.62
Eyebrow lift	97.14
Blepharoplasty	96.57
Local Surgeries such as Botox	89.74
Lip Augmentation	80.00

Due to large variations in the targeted area and their effect, we observe that it is difficult to model all these variations leading to a lower accuracy of 89.74% for these local surgeries.

4.3.3 Face Verification Performance on Plastic Surgery Pairs

To evaluate the effect of Stage-2 of the proposed MPDL framework, face verification accuracy is computed only on plastic surgery images. Figure 4.3 shows the effect of Stage-2 of the algorithm computed independently from Stage-1 of MPDL. For this evaluation, it is assumed that the input pairs are correctly detected as plastic surgery pairs. We compare the Stage-2 of the proposed MPDL with other face descriptors and algorithms. The corresponding ROC is shown in Figure 4.3. The lowest Equal Error Rate (EER) of 6.74% is achieved by Stage-2 of the proposed MPDL while TPLBP [131], HOG [22], LBP [86], and UCLBP [87] yield EERs of 17.75%, 20.45%, 22.25%, and 28.31% respectively. For comparison, recent face recognition across plastic surgery algorithms proposed by De Marsico *et al.* [25] and Sun *et al.* [119] are implemented in verification mode to verify plastic surgery faces only. EERs of 19.55% and 37.30 % are observed using approaches by De Marsico *et al.* [25] and Sun *et al.* [119] respectively. These results demonstrate the superior ability of Stage-2 of the proposed MPDL framework to verify the identity when considering only plastic surgery query pairs.

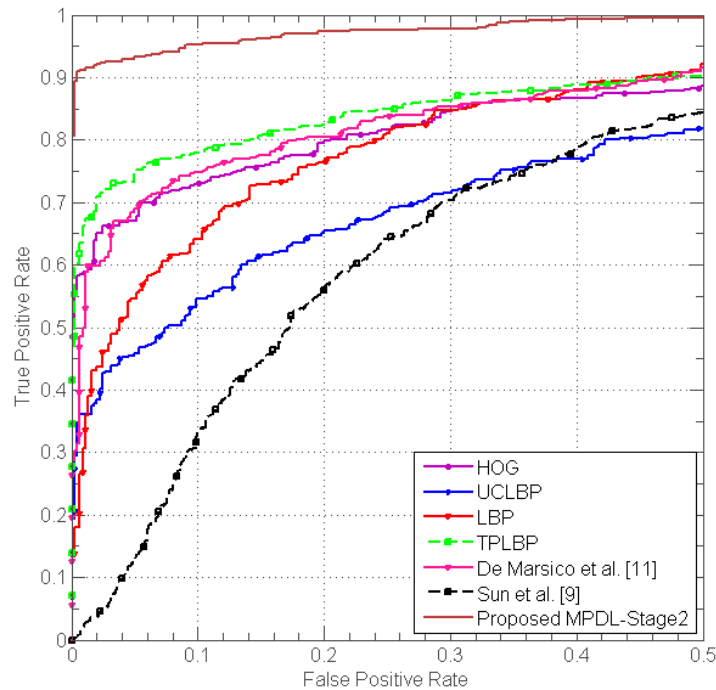


Figure 4.3: ROC for face verification on plastic surgery images only using HOG, UCLBP, LBP, and TPLBP face descriptors, De Marsico *et al.* [25] and Sun *et al.* [119], and Stage-2 of the proposed MPDL framework.

4.3.4 Performance of integrating the Proposed MPDL Framework with COTS

The performance of the proposed MPDL framework is further studied in the context of two commercial face recognition systems: FaceVacs [21] referred to as COTS-1 and FaceSDK [77] referred to as COTS-2. Three scenarios are considered:

- Scenario-1: Performance of regular faces without any plastic surgery face images using COTS-1 and COTS-2 (Stage-2)
- Scenario-2: Performance of regular faces and plastic surgery faces using COTS-1 and COTS-2
- Scenario-3: Performance of regular faces and plastic surgery faces when the proposed MPDL framework is integrated with COTS-1 and COTS-2 systems.

Table 4.5 shows that commercial systems COTS-1 and COTS-2 for regular faces (Scenario-1) have an EER of 3.85% and 0.72% respectively. When the database includes plastic surgery faces along with regular faces (Scenario-2), current COTS systems are not able to efficiently verify plastic surgery faces. The verification performance of the combined database decreases and this is reflected in the increased EER values of 9.65% for COTS-1 and 5.26% for COTS-2.

The major contribution of proposed MPDL is the ability to detect any plastic surgery face from the combined database of plastic surgery and regular face images in Stage-1 and form two separate groups containing plastic surgery faces and regular faces. Contextual switching (Scenario-3) is performed where images that are detected as regular faces are verified using COTS-1 and COTS-2 (or any other face verification algorithm); while the images which are detected as having undergone plastic surgery are verified using Stage-2 of the proposed MPDL framework. These two scores are combined to improve the verification performance by decreasing the EER value to 6.24% (Stage-2 of MPDL and COTS-1) and an EER value of 3.63% (Stage-2 of MPDL and COTS-2).

In comparison, the approach by De Marsico *et al.* [25] applied on the combined database of plastic surgery and regular face images implemented in verification mode gives an EER of 15.37% and SSIM index-based approach by Sun *et al.* [119] gives an EER of 29.72%, highlighting the superior performance of the proposed MPDL framework for face verification.

Table 4.5: Performance of COTS-1 and COTS-2 with regular faces, combined database of plastic surgery and regular face images, and the proposed MPDL framework integrated with commercial systems.

Experimental Scenarios	COTS-1 EER %	COTS-2 EER %
Regular Faces (Scenario-1)	3.85	0.72
Regular Faces & Plastic Surgery (Scenario-2)	9.63	5.26
Proposed MPDL framework integrated with COTS systems (Scenario-3)	6.24	3.63

4.4 Summary

We propose a novel multiple projective dictionary learning framework to detect plastic surgery for face verification. We train one dictionary to learn features of local plastic surgery face and regular face images simultaneously; while a second dictionary is trained to learn features of global plastic surgery face with the same set of regular face images for improved discrimination and detection of plastic surgery faces.

Traditional dictionary learning systems learn a sparse representation using l_0 and l_1 norms which is computationally intensive. We propose using projective dictionary learning since it does not involve the time-consuming computation of l_0 and l_1 norms. We further learn the face representation of plastic surgery images using a compact binary face descriptor. Experimental results using the plastic surgery database show the efficacy of our proposed MPDL framework. Detection accuracy of 97.96% is observed for plastic surgery faces as compared to commonly used features and classification algorithms.

After a pair of reference and query images is detected as plastic surgery, the Stage-2 of the proposed MPDL framework verifies the identity of the individual. For this task, the CBFs of 11 ROIs are computed from a pair of plastic surgery face images. The *cosine* distance scores between the two plastic surgery faces are used as the feature. Along with this, feature similarity scores between the same ROIs are calculated. These calculated measures are combined into a single feature vector to train a SVM model to verify if the plastic surgery faces belong to the same person.

The proposed MPDL framework for face verification is seamlessly integrated with two commercial systems. When a combined database of plastic surgery and regular face images is used, we observe an improvement in verification performance by a decrease in the equal error rate in the two COTS systems of 3.41% and 1.63%.

Chapter 5

Introducing the Effect of Illicit Drug Abuse based Temporal Variations on Face Recognition

Illicit drug abuse has become one of the primary health and social concerns in today's world. According to the World Drug Report [4], an estimated 183,000 drug-related deaths were reported in 2013. It is estimated that a total of 246 million people aged 15-64 have used illicit drugs, mainly substance belonging to cannabis, cocaine or amphetamine-type stimulants. There has been a lot of research in understanding the deteriorating effect of drugs on physical and mental health [121]. Certain drugs, when taken continuously in large quantities, can cause temporal and physiological changes in the skin. For instance, the long-term effects of methamphetamine (meth) and heroin can cause severe weight loss and skin sores. Reece [106] noted the evidence of accelerated aging due to the addiction to opiates.

While illicit drug abuse and its detection have applications in health and medical areas where several research directions are being explored [7], the effect of illicit drug abuse based temporal variations on automated face recognition systems has not been explored. There are several large-scale national ID projects and biometric systems that utilize face as a modality. In these applications, such temporal facial variations caused due to illicit drug abuse are not considered. Therefore, these systems may not be able to match before-and-after images affected due to illicit drug abuse.



Figure 5.1: Sample images that demonstrate the significant effect of illicit drug abuse on faces. Noticeable variations can be seen in the facial features of the *after* images of these subjects.

To the best of our knowledge, there is a lack of experimental evaluation or formal study to understand and analytically demonstrate the effect of illicit drug abuse based temporal variations on the current face recognition algorithms. This research attempts to bridge this gap and showcases the impact of temporal facial variations caused due to illicit drug abuse. We also present a dictionary learning approach to classify faces into categories: *drug abuse faces* and *regular faces*. The contributions of this research are summarized below:

1. Creating the first Illicit Drug Abuse Face (IDAF) dataset containing before and after images of 105 subjects, collected from the internet.
2. Demonstrating the impact of facial temporal variations caused due to illicit drug abuse on face recognition. The low performance of two commercial face recognition systems and two face descriptors on faces that have considerably changed due to consistent use of drugs.

3. Proposing a non-invasive classification algorithm using dictionary learning to detect face images affected due to illicit drug abuse such that it can be used in conjunction with current face recognition systems. The aim of such a framework is to reliably separate possible drug abuse face images where current systems may not identify the person correctly.

5.1 Effect of Illicit Drug Abuse on Face Recognition Algorithms

Due to the novelty of the research problem, there is no publicly available database. Therefore, we first collected a database from online resources. As part of the research efforts, we will release the database to the research community. To understand the performance of face recognition algorithms on the database, a set of state-of-the-art face recognition algorithms including two commercial systems are used.

5.1.1 Creation of Illicit Drug Abuse Face (IDAF) Database

Due to the sensitive nature of the process and privacy issues, it is extremely difficult to find images where people admit prolonged illicit drug abuse. In 2004, a deputy at Multnomah County Sheriff's Office put together mug shots of persons booked into the detention center of Multnomah County [1]. The face images were released as "Faces of Meth" in order to make people realize the detrimental effect of substance abuse. Later, some faces belonging to heroin, crack and cocaine were added to the collection. A sample of such images are shown in Figure 5.1 where the accelerated aging and formations of scars are very evident. Using these images and other images collected from the internet [2], Illicit Drug Abuse Face (IDAF) dataset¹ is created. This dataset contains two frontal face images: first when the subject was not taking any kind of drug (before image) and second when there has been a

¹The database can be accessed at <http://iab-rubric.org/resources.html#face>.

substantial amount of illicit drug abuse (after image). The database contains 210 images pertaining to 105 subjects. The database comprises of face images of methamphetamine, cocaine, heroin, and crack addicts.

5.1.2 Face Recognition Algorithms for Evaluation

Two Commercial-Off-The-Shelf (COTS) systems, FaceVACS [21] and Luxand [77] are utilized to study the effect of illicit drug abuse on face recognition algorithms. Local Binary Patterns (LBP) [86] and Histogram of Oriented Gradients (HOG) [22] along with *chi – square* distance measure are also used to study the effect of their performance on these images. These algorithms take into account the texture and the oriented gradients and are popular for the task of face recognition.

5.1.3 Experimental Scenarios

To evaluate the effect of illicit drug abuse on face recognition, three different experimental scenarios are constructed.

Scenario 1: The first experiment (called *Drug Faces*) is conducted using the Illicit Drug Abuse Face (IDAF) dataset that contains *before* and *after* face images of 105 subjects who are drug users.

Scenario 2: The second experiment (called *Regular Faces*) is used to obtain the baseline performance of the face recognition systems. For this purpose, two frontal images of 300 subjects from CMU-MultiPIE dataset [39] and 700 subjects from the visible spectrum of CASIA-VIS-NIR dataset [64] are chosen to form a database of regular faces.

Scenario 3: The third experiment (called *Combined Faces*) uses two frontal images of 268 random subjects from the 300 subjects of CMU-MultiPIE used for the previous experiment, and similarly 627 subjects from 700 subjects of CASIA-VIS-NIR used

in the previous experiment. In addition, 105 subjects are added from the Illicit Drug Abuse Face (IDAF) dataset to make it a complete set of 1000 subjects.

To evaluate the performance of the commercial systems and existing descriptor-based face recognition algorithms, the *before* image in all the three scenarios are used as a gallery while the *after* image is used as a probe.

5.1.4 Experimental Results

Figures 5.2, 5.3 and 5.4 show the Cumulative Matching Characteristic (CMC) plots pertaining to the above experiments. The key findings are reported below:

The CMCs in Figure 5.2, 5.3 and 5.4 show Rank-1 accuracy for *Regular Faces*, *Combined Faces* and *Drug Faces* scenarios. Both the commercial systems perform very well on the images from the *Regular Faces* database and the results are used as baseline for comparison purposes. However, there is a noticeable drop in performance for both the systems, when images of Illicit Drug Abuse Face (IDAF) dataset are introduced in the *Combined Faces* scenario. As can be seen, the Rank-1 accuracy is 99.4% and 99% respectively for FaceVACS and Luxand for the *Regular Faces* scenario. In the *Combined Faces* scenario, this drops down to 96.6% and 94.9%.

A similar decrease in performance is observed for both LBP and HOG descriptors as shown in Figures 5.3 and 5.4. The decrease in performance can be attributed to the images from the Illicit Drug Abuse Face (IDAF) dataset. This can be seen from the *Drug Faces* scenario where both the commercial systems and facial descriptors perform badly in the recognition task. Figure 5.4 shows the performance of the commercial systems is 0.07% and 0.05% for FaceVACS and Luxand in the *Drug Faces* Scenario respectively. HOG performs the best with an identification accuracy of 37% Rank-1 while LBP gives 28% Rank-1 accuracy.

These experimental results clearly highlight the challenges of illicit drug abuse face images using current face recognition algorithms. With the widespread use of drugs globally

and their effect on the skin texture and alteration of facial features, there is a major need for face recognition algorithms to address this challenge.

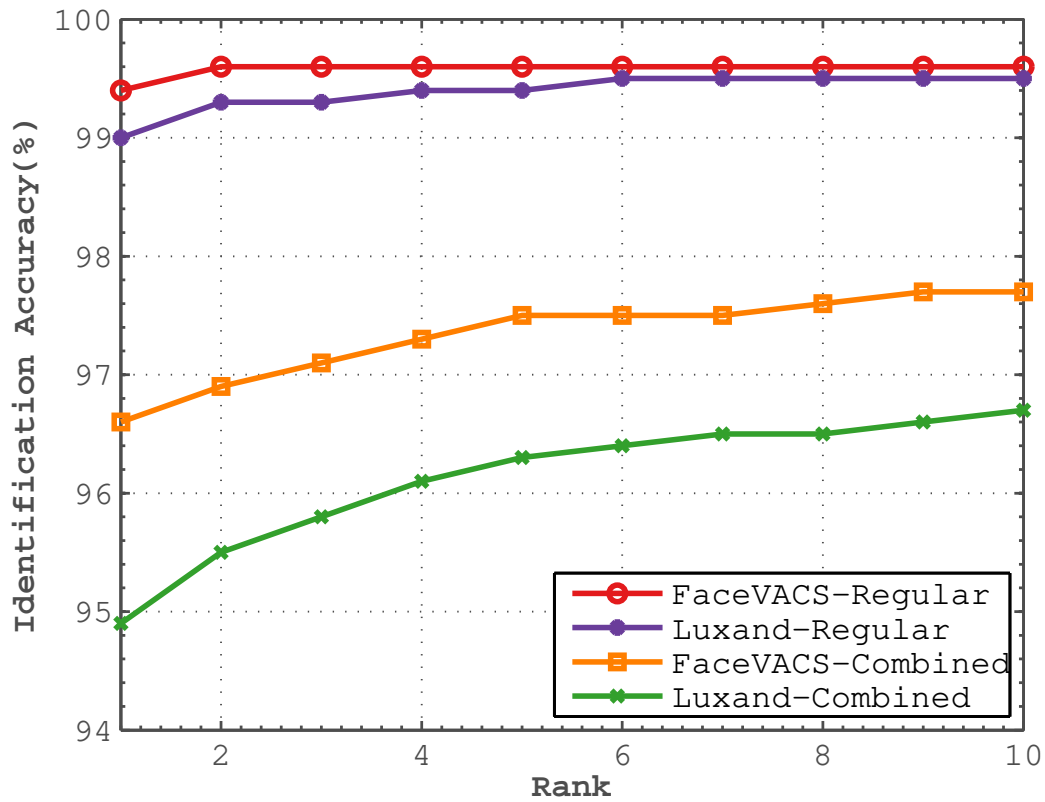


Figure 5.2: CMC curves for *Regular Faces* and *Combined Faces* Scenario when COTS systems FaceVACS [21] and Luxand [77] are used. It is seen that introduction of illicit drug abuse images lowers the performance of COTS.

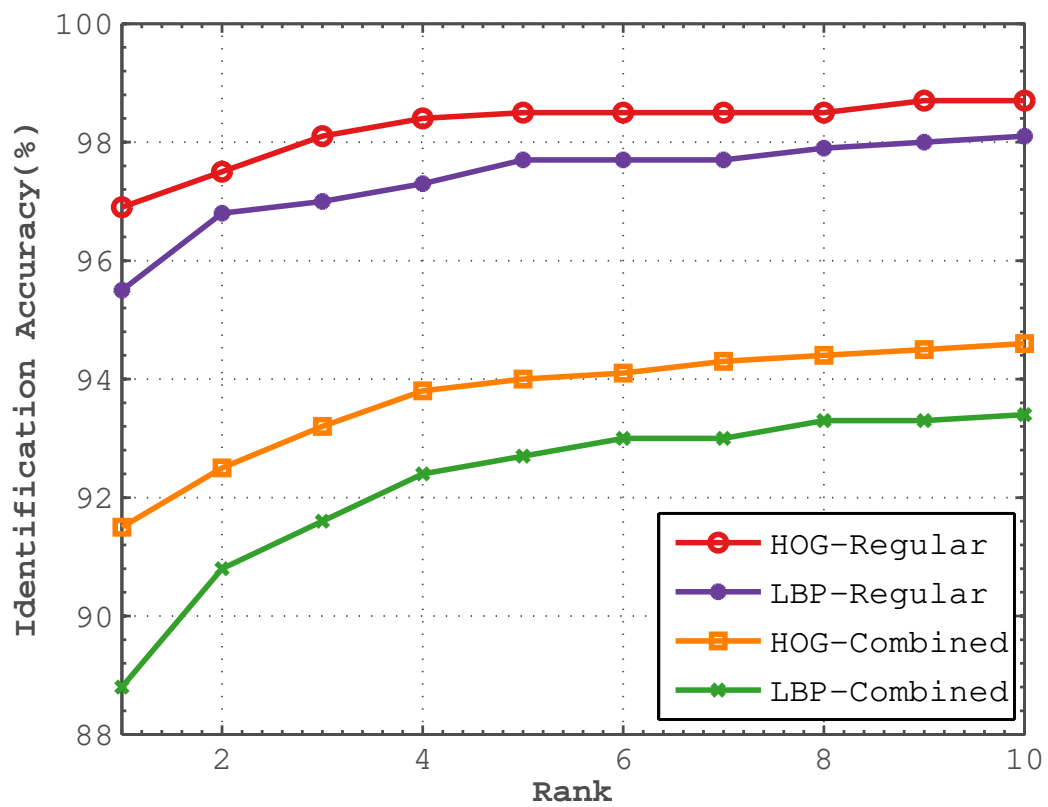


Figure 5.3: CMC curves for *Regular Faces* and *Combined Faces* Scenario when two face descriptors, LBP [86] and HOG [22] are used. It is seen that introduction of illicit drug abuse images lowers the performance of the face descriptors as well.

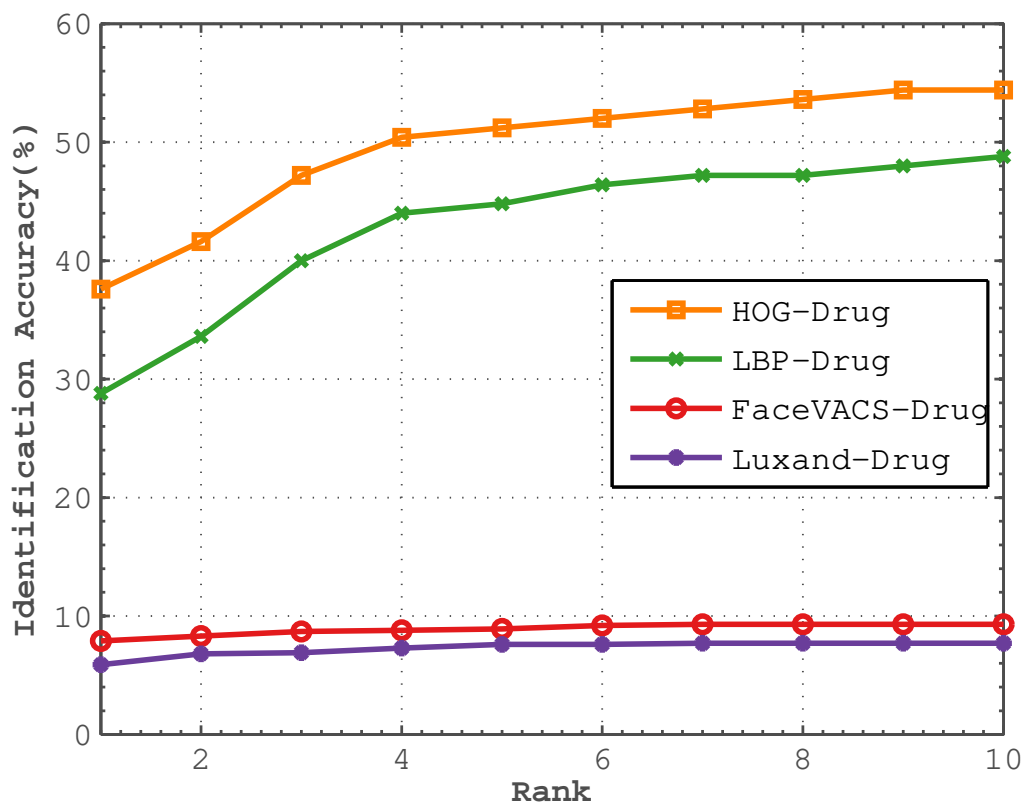


Figure 5.4: CMC curves for *Drug Faces* Scenario when two COTS, FaceVACS [21] and Luxand [77], and two face descriptors, LBP [86] and HOG [22] are used. It is seen that very few subjects have been correctly identified.

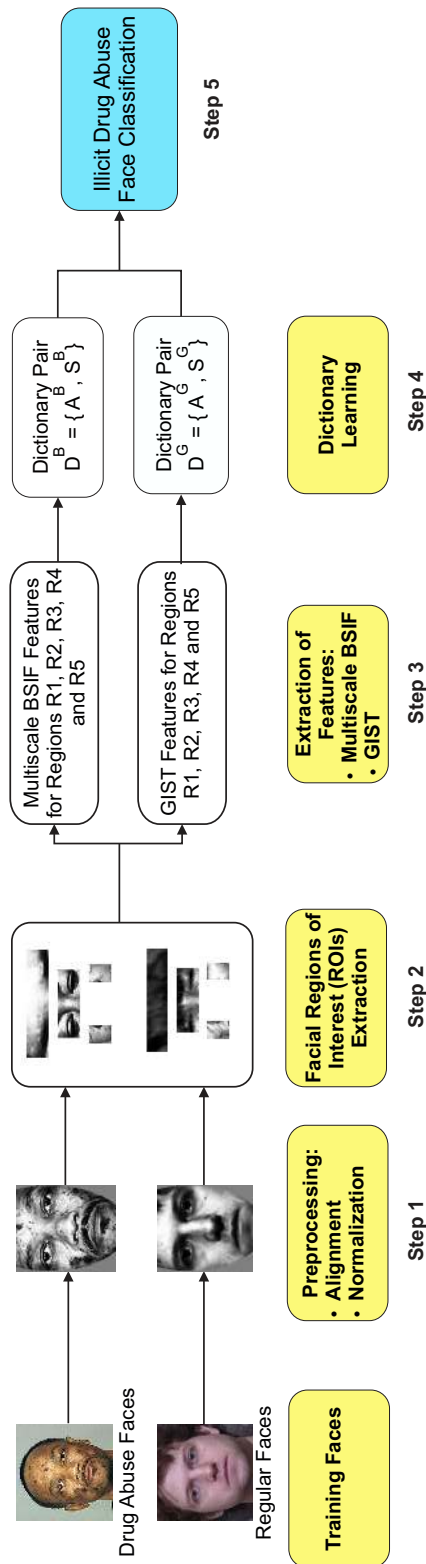


Figure 5.5: Proposed Dictionary learning based illicit Drug Abuse face Classification (DDAC) framework to classify faces affected by drug abuse.

5.2 Proposed Dictionary Learning based Illicit Drug Abuse Face Classification Framework

The previous section demonstrates the effect of facial feature variations caused due to illicit drug abuse on current facial recognition systems. In this section, we propose a Dictionary learning based illicit Drug Abuse face Classification (DDAC) framework where the goal is to separate the face images into two categories, namely *drug-abuse* and *regular*. This can be considered as a filtering process where these images can be matched differently using individually tailored adaptive algorithms for improved face recognition performance. The DDAC framework, shown in Figure 5.5, comprises of the following steps:

Step 1: Face pre-processing by performing alignment and normalization.

Step 2: Facial regions of interest (ROIs) extraction.

Step 3: ROI based multi-scale Binarized Statistical Image Features (BSIF) [55] and GIST [88] feature computation.

Step 4: Learning multi-scale Binarized Statistical Image Features (BSIF) and GIST specific paired dictionaries to detect possible instances of illicit drug abuse faces from a combined face database.

Step 5: Combining decisions from the learned paired dictionaries.

5.2.1 Pre-processing and Extraction of Facial Regions of Interest

Faces are pre-processed by converting them to grayscale. All the faces are geometrically aligned by performing landmark detection [143]. This is followed by normalization and background masking using CSU Face Identification Evaluation System [13].

Prolonged illicit drug abuse may lead to significant temporal changes in facial features such as pronounced wrinkles, blisters, and scarring. The facial regions that are more likely to be altered due to illicit drug abuse are chosen as regions of interest (ROIs). The five selected

ROIs showing most prominent variations are the full face (R1), binocular region (R2), right cheek (R3), left cheek (R4), and forehead (R5). For each face image, these five local and global ROIs are extracted for feature computations and classification.

5.2.2 Computation of Multiple Discriminatory Features

In the proposed approach, two features are extracted and used for classification: Binarized Statistical Image Features (BSIF) [55] and GIST descriptor [88]. The effect of extensive abuse of drugs may cause accelerated aging, open sores, blisters, blemishes, and scarring on the faces. To model the textural changes and deformations occurring due to these variations, multi-scale BSIF and GIST are computed on the five local and global ROIs.

Binarized Statistical Image Features (BSIF) [55] are gaining popularity as an efficient texture feature in the field of computer vision. In this approach, each element of the binary code is calculated by binarizing the output of a linear filter with thresholding. Each bit of the code denotes a different filter by projecting the image patches to a subspace. Statistical properties of natural images decide the binary code because the set of filters are learned from natural images by maximizing the statistical independence of the responses. Due to this property, statistically meaningful texture information can be learned from the data. The number of filters is an important parameter in BSIF. Using a single filter in BSIF may not encode sufficient discriminatory textural information. Hence, BSIF can be computed at multiple scales (multi-scale BSIF) to enhance the representation of the textural model.

On the other hand, the idea behind GIST descriptor [88] is to learn a low dimensional representation of an image. It also encodes the shape and structure of the image. The descriptor combines statistical information of the responses of filters. It is used to obtain a coarse vector encoding of distributions of different filter orientations and scales in the scene.

A filtering scheme described below is first applied to remove illumination variations.

$$I'(x,y) = \frac{I(x,y) \times f(x,y)}{\varepsilon + \sqrt{[I(x,y) \times f(x,y)]^2 \times g(x,y)}} \quad (5.1)$$

where, $I(x, y)$ is the input image, $g(x, y)$ is a low pass Gaussian filter, and $f(x, y) = 1 - g(x, y)$ is the corresponding high pass filter. The image is divided into 16 blocks. A set of 32 Gabor filters with 4 different scales and 8 different orientations are used to convolve with each of the 16 blocks and the mean moment is concatenated to form the resultant feature vector.

5.2.3 Feature Specific Paired Dictionary Learning

The proposed DDAC framework performs a filtering step to classify faces in a large dataset under two categories, namely *drug-abuse* faces and *regular* faces. This important step needs to be reliable and fast in terms of computation.

Two separate paired dictionaries D^B and D^G are learned using multi-scale BSIF and GIST features, respectively. These dictionaries are then utilized for classification and their individual decisions are combined using decision level fusion. For training the paired dictionaries, joint learning of analysis and synthesis dictionaries is performed to learn representations through linear projection without using the non-linear sparse encodings [41]. The model can be described as:

$$\begin{aligned} \{A^*, S^*\} = \arg \min_{A, S} \sum_{k=1}^K \|X_k - S_k A_k X_k\|_F^2 \\ + \beta \|A_k \bar{X}_k\|_F^2, \quad s.t. \|d_i\|_2 \leq 1 \end{aligned} \quad (5.2)$$

where, S represents the synthesis dictionary used to reconstruct X ; A represents the analysis dictionary used to code X ; A_k and S_k represent the sub-dictionary pair corresponding to class k ; \bar{X}_k represents the complementary data matrix of X_k in the training set; $\beta > 0$ is a scalar constant that denotes the regularization parameter to control the discriminative property of A , and d_i denotes the i^{th} item of synthesis dictionary S .





		Predicted Labels	
		Drug Abuse Face	Regular Face
Actual Labels	Drug Abuse Face		
	Regular Face		

Figure 5.6: Sample images from classes *drug-abuse faces* and *regular faces* which are correctly and incorrectly classified by the proposed DDAC framework.

5.2.4 Illicit Drug Abuse Face Classification and Decision-Level Fusion

Reliable illicit drug abuse face classification is performed using dictionaries D^B and D^G separately. Let y be the testing image. The detection of illicit drug abuse faces can be calculated using the two individual dictionaries as defined below:

$$Y_{BSIF} = \operatorname{argmin} \|y - \mathbf{S}^B \mathbf{A}^B y\|_2 \quad (5.3)$$

$$Y_{GIST} = \operatorname{argmin} \|y - \mathbf{S}^G \mathbf{A}^G y\|_2 \quad (5.4)$$

While there can be multiple ways of combining the output of dictionaries, here, we combine them using decision-level fusion [51]. Logical OR is applied on the two decisions to obtain a final decision of whether the given face image belongs to *drug-abuse* face category or not

$$Y_{DRUG} = Y_{GIST} \vee Y_{BSIF}$$

5.3 Experimental Results

5.3.1 Experimental Setup

To evaluate the detection performance of the proposed Dictionary Learning based Illicit Drug Abuse Face Classification (DDAC) framework and simulate the real world scenario, two types of face images are used: IDAF database and regular face database. One *before* and one *after* images of 105 subjects from the IDAF database and the first 105 subjects from CMU Multi-PIE [39] face database are utilized to form the combined face database. Two images with a neutral expression, frontal pose, and proper illumination are selected from CMU Multi-PIE [39] face database.

For the purpose of this experiment, unseen training and testing is performed with five-fold cross-validation. Images from the combined database of illicit drug abuse faces and regular faces are pre-processed. Five ROIs (full face, binocular region, right cheek, left cheek, and forehead) are extracted from each face image. For each ROI, BSIF is computed at multiple scales of 3, 5, 7, and 11 and the feature vectors are concatenated. Using the concatenated feature vector, two separate dictionaries are learned where the *after* images from the IDAF dataset are considered as one class (positive class) while the remaining data points are combined to form the negative class. Also, GIST features for each ROIs are calculated. Similarly, GIST based paired dictionary is learned to classify possible drug abuse face images. The classification results from the two learned dictionaries are combined using an OR operator to yield a final classification decision.

Results and Analysis

Average classification accuracy, across five folds, of whether a given face image is affected by drug abuse or not is reported in Table 5.1. For comparison purposes, classification accuracy obtained using commonly used texture and face descriptors such as multi-scale BSIF [55], HOG [22], Self-Similarity [128], GIST [88], and LBP [86] are also shown. From Table 5.1, it is observed that the proposed DDAC framework yields the highest accuracy of 88.81% and

outperforms the commonly used image descriptors. Multi-scale BSIF [55] with SVM yields an accuracy of 75.00%. Similarly, HOG [22], Self-Similarity [128], GIST [88], and LBP [86] yield detection accuracy of 76.90%, 78.01%, 81.09%, and 81.41% respectively.

The proposed DDAC framework is also compared with the performance of paired dictionary learned on multi-scale BSIF, GIST, Self-Similarity, HOG, and LBP separately. Classification accuracy of 86.38% is observed with the learned paired dictionary with multi-scale BSIF and 85.12% is observed with the learned paired dictionary with GIST.

Further analysis is performed on the output of the proposed framework. The proposed DDAC framework correctly classified 80.95% of illicit drug abuse face images (true positive rate) whereas, 92.06% of regular faces are correctly detected as the negative class (true negative rate). Figure 5.6 shows sample images from classes *drug-abuse faces* and *regular faces* which are correctly and incorrectly classified.

The proposed illicit drug abuse face classification DDAC framework uses decision-level fusion to combine the outputs from the multi-scale BSIF learned dictionary and GIST learned dictionary. For comparison purposes, sum rule fusion is also applied to the two classification outputs. From Table 5.1, it is observed that the accuracy using sum rule fusion is 85.84%.

Computationally, the proposed algorithm requires 0.59 seconds for classifying one image on a desktop PC with 3.4 GHz Intel CPU and 16 GB memory. This includes extraction of multi-scale BSIF and GIST features for the five local and global ROIs.

5.4 Summary

In this chapter, we introduce the Illicit Drug Abuse Face (IDAF) dataset and present the effect of illicit drug abuse based temporal variations as another challenge of face recognition. In the current scenario, illicit drug abuse has become one of the major health and social concerns in the world. As seen in the before and after face images, abuse of drugs drastically alters the temporal facial features and hence, it is a challenging research issue. Experiments have been performed to show the deterioration in the performance of commercial face recognition

Table 5.1: Average detection accuracy (%) for illicit drug abuse face classification using different classification algorithms.

Classification Algorithm	Classification Accuracy (%)
Multi-scale BSIF [55] + SVM	75.00
HOG [22] + SVM	76.90
Self-Similarity [128] + SVM	78.01
GIST [88] + SVM	81.09
LBP [86] + SVM	81.41
Dictionary Learning using LBP	80.48
Dictionary Learning using HOG	83.32
Dictionary Learning using Self-Similarity	84.18
Dictionary Learning using GIST	85.12
Dictionary Learning using Multi-Scale BSIF	86.38
Sum Rule Fusion of Multi-Scale BSIF and GIST Dictionaries	85.84
Proposed DDAC Framework	88.81

algorithms as well as commonly used face descriptors when illicit drug abuse face images are added to the database of regular faces. The results clearly demonstrate the need to further study and mitigate the effect of illicit drug abuse on face recognition algorithms. We also propose a detection framework to seamlessly classify in real-time using multiple dictionaries if a given face image is a *regular* face or a *drug-abuse* face and improve the face recognition performance. This framework can act as a crucial pre-processing step in mitigating the effect of such images. The proposed DDAC framework gives a classification accuracy of 88.81% when applied on a combined database of illicit drug abuse faces as well as regular faces.

Chapter 6

Detecting Plastic Surgery and Illicit Drug Abuse Faces using Adaptive Deep Multi-One Shot Similarity based Multiple Instance Learning

In the previous chapters, dictionary learning based approaches are proposed to detect faces impacted by temporal variations. In this chapter, we describe a novel single-face image-based algorithm to detect plastic surgery and illicit drug abuse.

As observed in Figure 6.1, plastic surgical procedures drastically alter the facial features and increase the intra-class distance between the before and after face images of the subjects which contribute to a decrease in the face matching performance. Similar to plastic surgery, illicit drug abuse also impacts facial features. Prolonged intake of a certain drug may cause drug dependency, which may, in turn, lead to health concerns such as weight loss, accelerated aging, sores, blisters, and skin irritation [31, 74] as shown in Figure 6.2. Similar to facial plastic surgery, these temporal variations increase the intra-class distance between different subjects, hence, declining the recognition accuracy of such faces.

The contributions of this research are summarized as follows:

1. We propose a novel Deep Multi-One Shot Similarity-based Multiple Instance Learning (DMOSMIL) algorithm for face image classification where a given object is considered as an image bag with a known label comprising of its constituent unlabeled instances.



Figure 6.1: Illustrating the variations in facial features due to facial plastic surgery.



Figure 6.2: Illustrating the variations in facial features due to prolonged illicit drug abuse.

DenseNet feature based representative instances of the positive bags are selected and a multi-one shot similarity score is applied with these representatives to learn the diverse representation of the class.

2. A variant of DMOSMIL algorithm termed as Adaptive DMOSMIL (or ADMOSMIL) is also proposed where the best DenseNet based feature descriptor is adaptively selected for a given local instance. The two proposed algorithms are able to precisely represent complex real-world objects and encode inter and intra-class variations.

3. Quantitative evaluations on two case studies: facial plastic surgery detection and illicit drug abuse classification using face images; demonstrate that the superior performance of the proposed algorithms.
4. We incorporate the face classification score by the proposed ADMOSMIL algorithm as a soft biometric trait in an existing face recognition algorithm. This proposed formulation leads to a boost in the face identification accuracy on three different databases.

6.1 Proposed Algorithms

As described earlier, detecting faces affected by facial plastic surgery and illicit drug abuse is crucial to mitigate their impact on face recognition accuracy. These variations may affect different facial regions differently. For instance, sores may appear in the forehead region and wrinkles may be more prominent in the chin area. Likewise, various plastic surgeries target different local regions to enhance their appearance. Therefore, it is imperative to examine the contribution of different local regions for detecting these face images. In this research, we present two novel algorithms: Deep Multi-One Shot similarity based Multiple Instance Learning (DMOSMIL) and Adaptive DMOSMIL (ADMOSMIL) for detecting such faces. The proposed algorithms incorporate deep features using DenseNet in a multi-instance learning framework to capture the variations across different local facial regions. To encode the diversity in these faces, we incorporate Laplacian Score to determine k representatives from these faces and multi-one shot similarity metric learning approach is employed to determine the similarity of an input face with the k representatives, which increases the discriminability power of the classifier. Figure 6.3 provides an overview of the proposed DMOSMIL algorithm.

Next, we will describe the ingredients of the proposed algorithms in the Preliminaries followed by the in-depth description of the proposed algorithms.

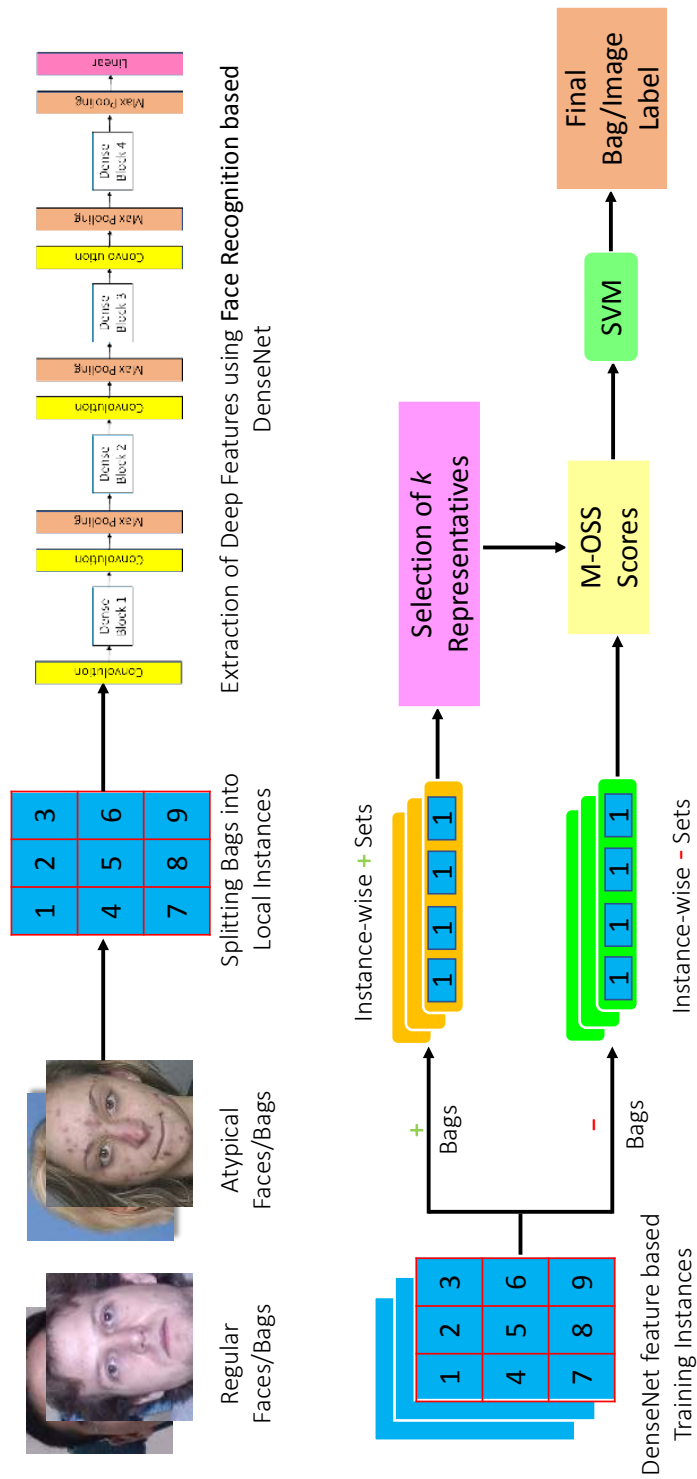


Figure 6.3: Proposed Deep Multi-One Shot Similarity-based Multiple Instance Learning (DMOSMIL) algorithm for atypical face image classification.

6.2 Preliminaries

6.2.1 One-Shot Similarity

One-shot similarity (OSS) measure [131, 132] compares two vectors by utilizing a set of negative examples to encode which representations should be considered *different*. Given two vectors v_i and v_j and a set S of negative samples, the OSS score is calculated by training one model for each vector, distinguishing it from the negative set. Then, these two learned models are used to determine if a vector has the same label as the other or if it belongs to the set of negative examples. To expand the model ($Model_i$) with respect to vector v_i is computed as: $Model_i = TrainClassifier(S, v_i)$

Similarly, $Model_j$ is trained by considering the negative set S and the positive instance, v_j . Next, testing is performed and the similarity score of v_j as compared to negative set S and v_i is computed by: $Score_j = TestClassifier(Model_i, v_j)$

Likewise, $Score_i$ is computed which measures the similarity of v_i to the negative set as well as the other vector, v_j . The resultant OSS scores between the two vectors v_i and v_j is:

$$FinalScore = \frac{Score_i + Score_j}{2} \quad (6.1)$$

One-sided OSS score is computed by training the classifier for only one of the vectors and testing on the other. Wolf et al. [132] demonstrated the advantages of using OSS with LDA as the classifier as it is a conditionally positive definite kernel. Therefore, OSS between when two vectors v_i and v_j using LDA as the base classifier is:

$$\frac{(v_i - \mu_S)^T M_W^+ (v_j - \frac{v_i - \mu_S}{2})}{\| M_W^+ (v_i - \mu_S) \|} + \frac{(v_j - \mu_S)^T M_W^+ (v_i - \frac{v_j - \mu_S}{2})}{\| M_W^+ (v_j - \mu_S) \|} \quad (6.2)$$

where μ_S is the average of the negative set S and M_W is the sum of the covariance matrix of both the classes.

As OSS metric does not incorporate the class label information, its foundations were utilized to propose multi-one shot similarity (M-OSS) by [122]. They hypothesized that

computing multiple OSS scores with different negative sets or different single positive instances can lead to improved performance of the vector pair-matching classifier. These multiple OSS scores can accurately encode the variations in the same class set, thus, showing improved performance as compared to OSS metric.

6.2.2 Multi-instance Learning

Multi-instance learning (MIL) [27] gained attraction due to its unique capability to represent real-world objects in a more natural manner as compared to traditional supervised learning genre. In MIL, every object is considered a bag which contains instances. Given a labeled bag data set $\{(X_1, Y_1), \dots, (X_N, Y_N)\}$ where $X_i = \{x_{i,1}, \dots, x_{i,n}\}$ is composed of n number of instances, $Y_i = \{-1, +1\}$ is the label of the bag, N is the number of training bags, and the task is to classify unseen bags or instances. The traditional assumption of MIL states that if there is at least one value of $b \in \{1, 2, \dots, n\}$ where instance $x_{i,b}$ is a positive instance, then the label (Y_i) of the bag $X_i = +1$. Otherwise, Y_i is assigned the label -1 . Different modifications have been proposed to this traditional assumption of multi-instance learning based on the inherent structure of the problem.

6.2.3 Laplacian Score

Laplacian score feature selection [47] involves identifying the features that have higher locality preservation characteristics. While computing the Laplacian score, the features that respect the overall graph structure are given a higher score where the graph structure is determined through the nearest neighbors.

Let the training data representation be \mathbf{X} , where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ and each sample has K dimensions i.e $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iK}]$. The objective is to select the best t features where $1 \leq t \leq K$ that are which preserve the locality structure. The Laplacian score for each feature is computed as follows:

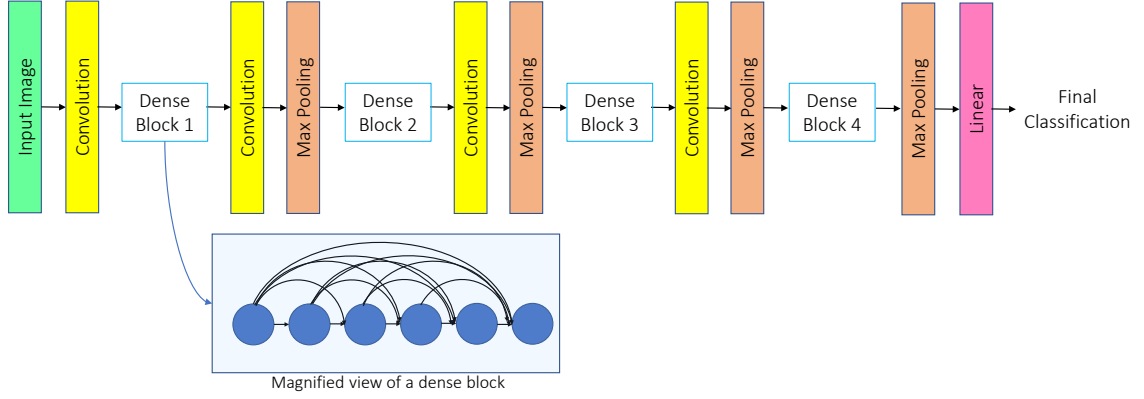


Figure 6.4: Architecture of DenseNet based convolutional neural network.

- Create a graph G of N nodes where two nodes a and b have an edge between them, if \mathbf{x}_a is among the p nearest neighbors of \mathbf{x}_b or vice versa.
- Let S = weight matrix of the graph G . If nodes a and b are connected $S_{ab} = e^{-\frac{\|\mathbf{x}_a - \mathbf{x}_b\|^2}{m}}$; where m is a constant, else $S_{ab} = 0$.
- Let $\mathbf{f}_r = [f_{r1}, f_{r2}, \dots, f_{rN}]^T$ be defined for r^{th} feature, $D = \text{diag}(S\mathbf{1})$, $\mathbf{1} = [1, \dots, 1]^T$ and $L = D - S$.
- Let $\tilde{\mathbf{f}}_r = \mathbf{f}_r - \frac{\mathbf{f}_r^T D \mathbf{1}}{\mathbf{1}^T D \mathbf{1}} \mathbf{1}$.
- Laplacian score of r^{th} feature is calculated as:

$$L_r = \frac{\tilde{\mathbf{f}}_r^T L \tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^T D \tilde{\mathbf{f}}_r} \quad (6.3)$$

Laplacian scores L_1, \dots, L_K are sorted in descending order to get the higher ranked important features.

6.2.4 DenseNet Architecture

DenseNet based convolutional neural network architectures have recently shown remarkable performance in several image classification tasks [49]. As shown in Figure 6.4, in DenseNet,

each layer is connected to every other layer in a feed-forward fashion as compared to traditional convolutional networks where there is a single connection between each layer and its subsequent layers. For an input image x , the i^{th} layer obtains the concatenation of the feature representations from all the previous layer as input. Thus, the output x_i from layer i is computed as:

$$x_i = L_i([x_0, x_1, \dots, x_{i-1}]) \quad (6.4)$$

where L_i refers to the non-linear transformations such as batch normalization and pooling. The unique architecture of the DenseNet network strengthens feature propagation, encourages feature reuse, and substantially reduces the number of parameters in the trained model.

In this architecture, every dense block is followed by a transition block that consists of a convolution layer and pooling layer to reduce the size of the output. As seen in Figure 6.4, each dense block in consists of six convolution blocks where every convolution block is densely connected. A convolution block itself consists of a convolution layer, batch normalization layer, convolution layer, and a dropout layer. The output feature vector from dense block number db is denoted as F_{db} where $db = 1, 2, 3, 4$. Thus, the features extracted from each dense block can be utilized to accurately encode the representation of an input image.

6.3 Proposed DMOSMIL Algorithm for Face Image Classification

As described earlier, each local region in a given face image may contain information representing different classes. Hence, to evaluate the influence of the deep feature representation of each local region individually, the input images are treated as a bag comprising non-overlapping local regions or instances. Thus, the objective of the task is to classify an image bag as positive or negative based on the properties of its constituent instances.

Figure 6.3 provides an overview of the proposed DMOSMIL algorithm. In the first step of the proposed Deep Multi-One Shot Similarity-based Multiple Instance Learning (DMOSMIL) algorithm, given a pre-processed image (bag) X , it is split into $Z \times Z$ non-overlapping local patches (constituent instances). Each instance P_j in bag X is provided as an input to a DenseNet model. In this research, the DenseNet model has been trained for the task of face recognition to learn relevant information to encode the facial features. Next, deeply learned feature descriptor F_{db} is computed using the trained DenseNet model to encode its underlying information. After computation of F_{db} for every instance in the negative training bags, instance-specific negative sets (S_j) are created.

$$S_j = \{F_{db}(P_{1,j}), \dots, F_{db}(P_{i,j}), \dots, F_{db}(P_{N,j})\} \quad (6.5)$$

$$\forall j = \{1, \dots, n\} \text{ where } Y_i = -1$$

Additionally, using positive training bags, j^{th} instances from each bag are grouped together to form instance-specific positive sets, T_j as follows:

$$T_j = \{F_{db}(P_{1,j}), \dots, F_{db}(P_{i,j}), \dots, F_{db}(P_{N,j})\} \quad (6.6)$$

$$\forall j = \{1, \dots, n\} \text{ where } Y_i = +1$$

The next step involves selecting the best k representatives from the instance-specific positive sets T_j to encode the variations in the positive bags. For this purpose, Laplacian score [47] of each instance is computed. The Laplacian score indicates the locality preserving power and seeks representatives which best indicate the underlying manifold of each local patch. The computed Laplacian scores are sorted in descending order and k instances with the highest Laplacian score (termed as *Rep*) are chosen to represent the positive bags.

After this, for each instance in a bag, there exists an instance-specific negative set (S) and k representatives of the positive bags in *Rep*. Using this, k one-sided multi-class one-shot similarity scores are calculated for every training instance as described in Algorithm 1.

Algorithm 1: Multi-Class One-Shot Similarity Measure

```
1 function trainMOSS  $inst, S_j, Rep_j$ 
   Input :  $i^{th}$  Training Instance  $inst$ , Negative Set:  $S_j$ , and  $k$  Positive Representatives
           Set:  $Rep_j$ 
   Output :  $k$  M-OSS Scores
2 for  $a \leftarrow 1$  to  $k$  do
3   | Model $_a$  = TrainClassifier( $S_j, Rep_j[a]$ )
4   | Score $_a$  = TestClassifier(Model $_a, inst$ )
5 end
```

The positive representative set Rep contains the most diverse instance variations from the positive bags. These OSS scores signify the probability of a given instance is closer to the negative set or the k representatives of the positive bags. Thus, the multiple OSS scores measure the likelihood of the given instance belonging to a specific representative of the positive or negative class.

The score vector obtained by concatenating the k multiple OSS scores is used as an input to a Support Vector Machine (SVM) along with the label of the input bag. The predicted probability values of different instances of a bag are combined using sum-rule fusion to compute the final label of the bag. The proposed DMOSMIL is summarized in Algorithm 2.

6.4 Proposed Adaptive DMOSMIL Algorithm for Face Image Classification

It is to be noted that the proposed DMOSMIL requires selecting a feature descriptor F_{db} from a specific dense block number db to encode each instance in the bag. This feature descriptor may encode the textural or structural variations in the given instance. However, different local regions in an image may not be effectively encoded by a single feature descriptor due to varying information content. Thus, DMOSMIL is extended by adaptively selecting feature descriptors learned from different dense blocks for different instances of the bag.

Algorithm 2: Proposed Deep Multi-One Shot Similarity based Multiple Instance Learning (DMOSMIL) Algorithm

1 **function** trainDMOSMIL *TrSet*

Input : *TrSet* = $(X_i, Y_i), i = 1 \dots N$ where X denotes the set of training bags (images). $Y_i \in \{-1, +1\}$ is the class label. n is the number of instances in each bag.

Output: Trained Model *ModelSVM*

2 **for** *bag* $\in X$ **do**

3 Compute deep feature descriptor F_{db} for the *bag*

4 Divide *bag* into $Z \times Z$ non-overlapping instances I

5 **for** $j \leftarrow 1$ **to** n **do**

6 $inst = j^{th}$ instance I

7 **if** $Y(bag) = -1$ **then**

8 add *inst* to Negative Set S_j

9 **else**

10 add *inst* to Positive Set T_j

11 **end**

12 **end**

13 **end**

14 Compute Laplacian scores of T_j

15 Select *inst* with top- k Laplacian scores to form representative set *Rep*

16 $Scores = \text{trainMOSS}(inst, S_j, T_j)$

17 $ModelSVM = \text{trainSVM}(Scores, Y)$

The proposed DMOSMIL algorithm is applied with deep features extracted from different dense blocks and based on the validation set, the k M-OSS scores from the best performing feature for a particular instance are combined to train the SVM classifier. Similar to the proposed DMOSMIL algorithm, sum rule fusion is applied to the instance-specific distance scores obtained from the SVM classifier to obtain the final bag classification. Therefore, in the proposed ADMOSMIL algorithm, different feature descriptors $F_1 \dots F_4$ may be utilized to effectively encode the bags.

6.5 Experiments

6.5.1 Implementation Details

Every image is pre-processed by performing face detection and alignment to obtain an image of the size of 224×224 . For training the DenseNet architecture for the task of face recognition, Multi-PIE face database [39] is used which consists of face images from 337 subjects. In this implementation, DenseNet-121 architecture is utilized with Stochastic Gradient Descent optimization algorithm, a momentum of 0.9, and the negative log-likelihood loss criterion.

Next, the input image is split into $Z \times Z$ non-overlapping patches. For each patch, DenseNet based features are extracted from the four dense blocks where $db = 1$, $db = 2$, $db = 3$, and $db = 4$ yield feature vectors of size 3136, 784, 196, and 49 respectively. These feature vectors are then used for multi-OSS score computation where Linear Discriminant Analysis (LDA) is utilized as the base classifier. The k multi-OSS scores are then provided to radial basis function-kernel based SVM for classification purposes. After obtaining each instance label, the final bag label is computed.

6.5.2 Databases and Experimental Protocol

The performance of the proposed DMOSMIL and ADMOSMIL algorithms in detecting faces which have undergone plastic surgery using a single image is demonstrated. To the best of our knowledge, this is the first research on a single image based facial plastic surgery detection. For experimental evaluation, the plastic surgery (PS) database by Singh et al. [115] is utilized which contains *before* and *after* plastic surgery face images of 900 individuals. Similar to the problem of illicit drug abuse detection, before and after face images of regular 900 individuals are added to the PS database to simulate the real-world scenario. The average plastic surgery detection accuracy across five cross-validation folds is computed.

In this research, IDAF database is extended with before and after face images of an additional 65 subjects. The images have been collected from Faces of Meth initiative by the Multnomah Sheriff county¹ as well as the Internet. The extended database, termed as Extended-Illicit Drug Abuse Face (E-IDAF) database contains 340 face images belonging to 170 subjects with variations such as methamphetamine, cocaine, and heroin abuse. The database contains two frontal face images of each subject: the *before* image is acquired when the subject is not consuming any kind of drug and *after* image is captured when the subject has undergone significant illicit drug abuse (after image). For evaluating the performance of the proposed algorithm in real-world scenarios, regular (without illicit drug abuse) face images are also utilized. For this, an equal number of before and after frontal face images of 170 subjects are selected from publicly available face databases. The regular faces and E-IDAF database are combined together to form the *combined* database. The average illicit drug abuse classification accuracy across the five cross-validation folds is reported.

6.5.3 Feature Extractors for Comparative Analysis

Comparative analysis of the proposed DMOSMIL and ADMOSMIL algorithms is performed with the following existing feature descriptors:

¹<http://www.mcso.us/facesofmeth/>

1. Local Binary Pattern (LBP) [87]: We consider 3×3 neighborhood with respect to the center pixel while computing the LBP descriptor. This texture descriptor assigns a value to each pixel by thresholding and concatenating the results to form a decimal number.
2. Multiscale-Binarized Statistical Image Features (M-BSIF) [55]: BSIF is computed by learning a subspace from an independent component analysis of natural images in a patch-wise fashion. This is followed by binarization of the coordinates on this new basis by thresholding to obtain the final binary codes for each pixel. In this implementation, BSIF computed with scale = 5, 7, and 11 are combined together to form M-BSIF.
3. Histogram of Oriented Gradients (HOG) [22]: This feature descriptor encodes the input by counting the occurrences of gradient orientations in various localized portions.
4. Three Patch Local Binary Patterns (TPLBP) [131]: In TPLBP, for each pixel of the input, patches of size $w \times w$ are selected for comparison. Two such patches at α distance from the center pixel are compared to compute the bit value for the pixel.
5. VGG features [114]: VGG-16 convolutional net trained on ImageNet Large Scale Visual Recognition Competition dataset is utilized to encode deeply learned features of the images.

6.5.4 Existing Metric Learning Algorithms for Comparative Analysis

The experimental evaluation is also performed with the following existing metric learning approaches:

1. Sparse Determinant Metric Learning (SDML) [102]: It exploits sparsity nature in high dimensional space and is specifically designed for smaller sample size problems.
2. Least Squares Metric Learning (LSML) [71]: It utilizes relative comparisons between data samples to learn a Mahalanobis distance metric.

3. Multiple-Instance Learning via Embedded instance (MILES) [18]: It involves mapping of each input bag into a feature space defined by the instances using an instance similarity measure.
4. Multiple-Instance Learning via Disambiguation (MILD) [65]: In this paper, the authors proposed a disambiguation method to identify true positive instances in the training positive bags.
5. Multiple Clustered Instance Learning (MCIL)-Boost [136]: In this approach, image-level classification, pixel-level segmentation, and patch-level clustering are performed.
6. D^{RS} [19]: It is based on a classifier which is inspired by a random subspace ensemble and assumes subspaces of the dissimilarity space, defined by subsets of instances, as the prototypes.

6.5.5 Results

The average classification accuracy across the five cross-validation folds for detecting faces affected by illicit drug abuse and plastic surgery is reported in Tables 6.1, 6.2, 6.3, and 6.4. Receiver Operating Characteristic (ROC) curves depicting the performance of the proposed algorithms curves are shown in Figure 6.5 and 6.7.

The results showcase the superior performance of the proposed DMOSMIL algorithms across the two case studies. It is observed that for every feature descriptor (DenseNet based as well as existing features), it outperforms its constituent algorithms: LDA, M-OSS, and MIL.

It is also observed that the proposed DMOSMIL algorithm with DenseNet based features (F_{db}) consistently demonstrates better face classification accuracy as compared to existing features (both handcrafted and deep learning based). This highlights the efficacy of densely trained CNN architecture in encoding these variations. A similar trend is observed for the proposed ADMOSMIL algorithm where different feature representations are adaptively selected for different facial regions. The performance of the proposed ADMOSMIL algorithm

using DenseNet features is higher than the proposed ADMOSMIL algorithm with existing features by 1.47% and 6.11% for detecting faces affected by illicit drug abuse and plastic surgery respectively.

Table 6.1: Classification accuracy of the proposed DMOSMIL, proposed ADMOSMIL, LDA, M-OSS, and MIL for plastic surgery detection.

Feature	Algorithm	Accuracy (%)
DenseNet based Features		
$F_{db=1}$	LDA	54.17
	M-OSS	59.44
	MIL	67.92
	Proposed DMOSMIL	72.92
$F_{db=2}$	LDA	64.44
	M-OSS	61.53
	MIL	68.19
	Proposed DMOSMIL	72.50
$F_{db=3}$	LDA	53.61
	M-OSS	52.92
	MIL	53.33
	Proposed DMOSMIL	67.08
$F_{db=4}$	LDA	48.19
	M-OSS	62.64
	MIL	64.44
	Proposed DMOSMIL	72.92
Existing Features		
LBP	LDA	52.78
	M-OSS	58.06
	MIL	66.53
	Proposed DMOSMIL	71.53
M-BSIF	LDA	61.67
	M-OSS	58.75
	MIL	65.42
	Proposed DMOSMIL	69.72
HOG	LDA	50.28
	M-OSS	49.58
	MIL	50.00
	Proposed DMOSMIL	63.75
TPLBP	LDA	44.03
	M-OSS	58.47
	MIL	60.28
	Proposed DMOSMIL	68.75
VGG	LDA	65.42
	M-OSS	68.06
	MIL	69.58
	Proposed DMOSMIL	71.94
Proposed ADMOSMIL (Existing Features)		72.50
Proposed ADMOSMIL (DenseNet Features)		78.61

Table 6.2: Classification accuracy of the proposed ADMOSMIL and comparative algorithms for plastic surgery detection.

Algorithm	Accuracy (%)
SDML [102]	51.81
LSML [71]	60.97
MILES [18]	61.94
MILD [65]	62.92
MCILBoost [136]	60.69
D^{RS} [19]	64.58
Yadav et al. [137]	60.14
Proposed ADMOSMIL (DenseNet Features)	78.61

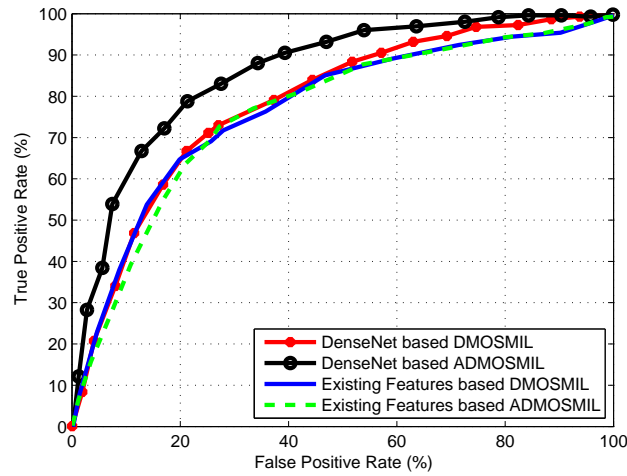


Figure 6.5: ROC curves of the proposed DMOSMIL and ADMOSMIL algorithms on the plastic surgery face database.

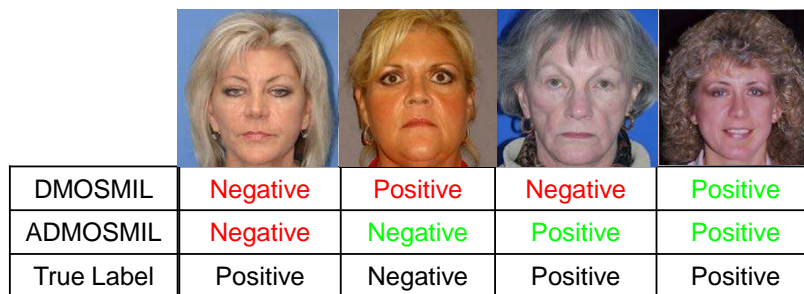


Figure 6.6: Sample classification outputs by DenseNet feature $F_{db=2}$ based DMOSMIL and ADMOSMIL algorithms on the plastic surgery face database. The label Positive is plastic surgery face and Negative is regular face.

Table 6.3: Classification accuracy of the proposed DMOSMIL, proposed ADMOSMIL, LDA, M-OSS, and MIL on the E-IDAF database for illicit drug abuse detection.

Feature	Algorithm	Accuracy (%)
DenseNet based Features		
$F_{db=1}$	LDA	71.32
	M-OSS	72.06
	MIL	70.59
	Proposed DMOSMIL	86.03
$F_{db=2}$	LDA	69.12
	M-OSS	70.59
	MIL	71.32
	Proposed DMOSMIL	90.44
$F_{db=3}$	LDA	67.65
	M-OSS	71.32
	MIL	72.06
	Proposed DMOSMIL	87.50
$F_{db=4}$	LDA	70.59
	M-OSS	72.06
	MIL	72.79
	Proposed DMOSMIL	86.03
Existing Features		
LBP	LDA	69.85
	M-OSS	70.59
	MIL	69.12
	Proposed DMOSMIL	84.56
M-BSIF	LDA	66.91
	M-OSS	69.12
	MIL	69.85
	Proposed DMOSMIL	88.97
HOG	LDA	66.18
	M-OSS	67.65
	MIL	69.85
	Proposed DMOSMIL	86.03
TPLBP	LDA	69.12
	M-OSS	70.59
	MIL	71.32
	Proposed DMOSMIL	84.56
VGG	LDA	88.24
	M-OSS	87.50
	MIL	88.97
	Proposed DMOSMIL	91.91
Proposed ADMOSMIL (Existing Features)		93.38
Proposed ADMOSMIL (DenseNet Features)		94.85

Table 6.4: Classification accuracy of the proposed ADMOSMIL and comparative algorithms for illicit drug abuse detection.

Algorithm	Accuracy (%)
SDML [102]	84.56
LSML [71]	82.35
MILES [18]	85.29
MILD [65]	81.62
MCILBoost [136]	83.09
D^{RS} [19]	86.03
Yadav et al. [137]	84.41
Proposed ADMOSMIL (DenseNet Features)	94.85

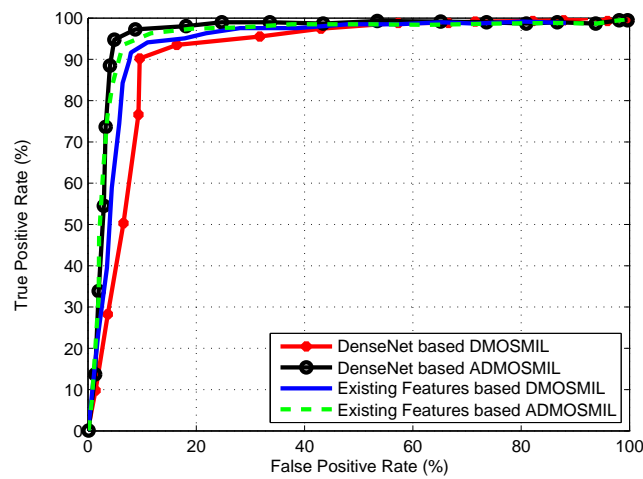


Figure 6.7: ROC curves of the proposed DMOSMIL and ADMOSMIL algorithms on the proposed E-IDAF database.

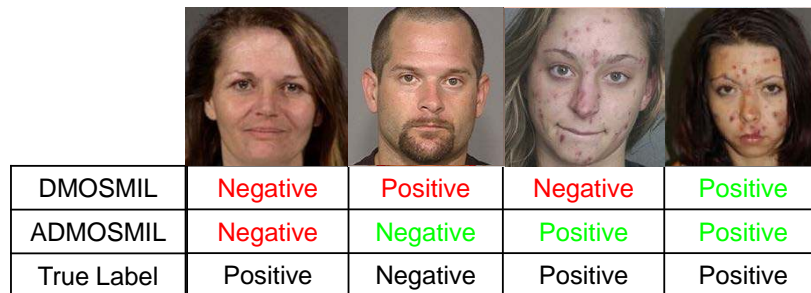


Figure 6.8: Sample classification outputs by the DenseNet feature $F_{db=2}$ based DMOSMIL and ADMOSMIL algorithms on the proposed E-IDAF database. The label Positive is illicit drug abuse and Negative is regular face.

The comparison of the proposed adaptive DMOSMIL (ADMOSMIL) algorithm using DenseNet features with different variants of the DMOSMIL algorithm illustrates the significance of the adaptive nature of the proposed ADMOSMIL algorithm where it adaptively selects feature descriptors learned from different dense blocks for different instances of the bag. Across the two case studies, it is seen that the ADMOSMIL algorithm achieves the highest face image classification accuracy as compared to the proposed DMOSMIL and various comparative algorithms.

For faces altered by plastic surgeries, Table 6.1 shows that DMOSMIL algorithm with features from different dense block yields classification accuracy ranging from 67.08% to 72.92% with $F_{db=1}$ and $F_{db=4}$ based DMOSMIL achieving the best performance. Figure 6.6 illustrates some sample classification outputs. Among the various handcrafted feature based versions of DMOSMIL, LBP based MOSMIL demonstrates the highest accuracy. This highlights that the LBP feature is able to encode subtle textural and structural variations associated with plastic surgery procedures. The proposed MOSMIL algorithms outperform its constituent algorithms with the different feature descriptors. Likewise, DenseNet based ADMOSMIL which is based on a mixture of different dense feature vectors achieves the highest plastic surgery detection accuracy of 78.61%. Comparative analysis is also performed with [137] to detect plastic surgery faces which demonstrate 18.47% lower accuracy as compared to the proposed ADMOSMIL. These results illustrate that the proposed ADMOSMIL algorithm produces a state-of-the-art performance on single face image based plastic surgery detection.

It is observed that using DenseNet features from different dense blocks with the proposed DMOSMIL algorithm, highest illicit drug abuse face classification accuracy of 90.44% is obtained. This highest accuracy is obtained using features from the second dense block ($F_{db=2}$) which is 4.41% higher than $F_{db=1}$, 2.94% higher than $F_{db=3}$, and 4.41% higher than $F_{db=4}$. The proposed DMOSMIL algorithm demonstrates superior performance as compared to traditional LDA, M-OSS, and MIL algorithms for features from all the four dense blocks.

Sample classification outputs by the best performing DenseNet feature based DMOSMIL ($F_{db=2}$) and ADMOSMIL algorithms are shown in Figure 6.8.

For comparative analysis, the proposed DMOSMIL algorithm is evaluated using different handcrafted features. It outperforms other approaches by at least 13.97% for LBP, 19.12% for M-BSIF, 16.18% for HOG, 13.24% for TPLBP, and 2.94% for VGG. It is also seen that the proposed DMOSMIL algorithm with M-BSIF as F yields the best classification accuracy as compared to other handcrafted feature descriptors. The justification for this superior performance can be attributed to the ability of M-BSIF to effectively encode discriminatory textural differences across varying scales between regular faces and illicit drug abuse faces.

6.6 Utilizing Face Classification as Soft Biometric for Boosting Face Recognition

Traditionally, soft biometric traits in a face such as the presence of facial hair and freckles provide ancillary information [93, 37]. These traits may not be fully distinctive or unique by themselves in face recognition (FR). Such soft biometrics traits are utilized to enhance the identification accuracy and provide additional information about the similarity/dissimilarity between two given face images [23].

In this research, we propose to utilize the face classification score as a soft biometric trait to improve the performance of face identification systems. The threshold decision value of face recognition should be determined by whether the subject has undergone plastic surgery. To account for this scenario, the decision boundary of the face recognition classifier should be modified accordingly based on whether the probe face image is a regular face or not. Thus, we propose a formulation which incorporates the face classification score by the proposed ADMOSMIL algorithm to boost the performance of existing face identification algorithm.

In the experimental evaluation, we utilize the popular pre-trained VGG-face [114] feature based face matching for the identification task which has demonstrated remarkable face recognition performance in various scenarios. For a given probe image, face identification

score, as well as face classification score, are computed with respect to the gallery face images which are used as input in the proposed formulation. Product of likelihood ratio (PLR) based approach [28] is employed to boost the performance in the following manner:

Let s be the face recognition (FR) score computed by matching a probe image with a gallery image. c represent the face classification score computed from the probe and gallery face images indicating if the face pair is altered due to plastic surgery or illicit drug abuse. Using these values, the product of the likelihood ratio score is calculated as:

$$PLRscore = \frac{P(s|gen)}{P(s|imp)} \times \frac{P(c|pos)}{P(c|neg)} \quad (6.7)$$

where pos is the positive face class, neg represents the negative/regular face class, gen is the genuine identification scores class, and imp is the imposter identification scores class. The probability terms $P(s|gen)$ and $P(s|imp)$ refer to the class conditional probability values. The four random variables are modeled as Gaussian mixture models.

6.6.1 Experimental Protocol and Results

In the experimental evaluation, for a given probe image, VGG face identification score, as well as ADMOSMIL classification score, are computed with respect to the gallery face images which are used as input in the proposed formulation. Product of likelihood ratio (PLR) based approach is employed to boost the performance as per Eq 6.7.

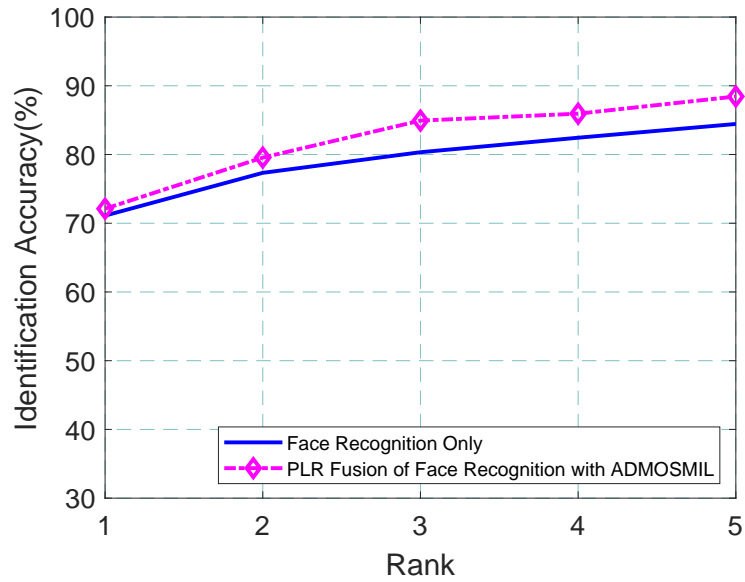
As per the experimental protocol, PS face and E-IDAF databases are processed separately. Across five cross-validation folds, using the E-IDAF database, VGG-face features of the before and after illicit drug abuse face images of the testing, set are computed. This is followed by computing the face matching score s using the *cosine* distance between the two feature vectors. Next, the ADMOSMIL algorithm is employed to compute face classification score c indicating if the gallery-probe face pair has been altered due to illicit drug abuse. Likewise, plastic surgery face pairs from the PS face database are processed.

Figure 6.9 showcases the Cumulative Match Characteristic (CMC) curves of the proposed formulation for boosting the face identification using the ADMOSMIL algorithm. For the E-IDAF database for illicit drug abuse, VGG-based face recognition algorithm yields 67.65% rank-1 face identification accuracy. However, by utilizing the proposed formulation, the rank-1 accuracy increases by 5%. Similarly, for matching face images altered by plastic surgery, the rank-1 FR performance of VGG-face algorithm is increased by 1.2% using the proposed boosting formulation.

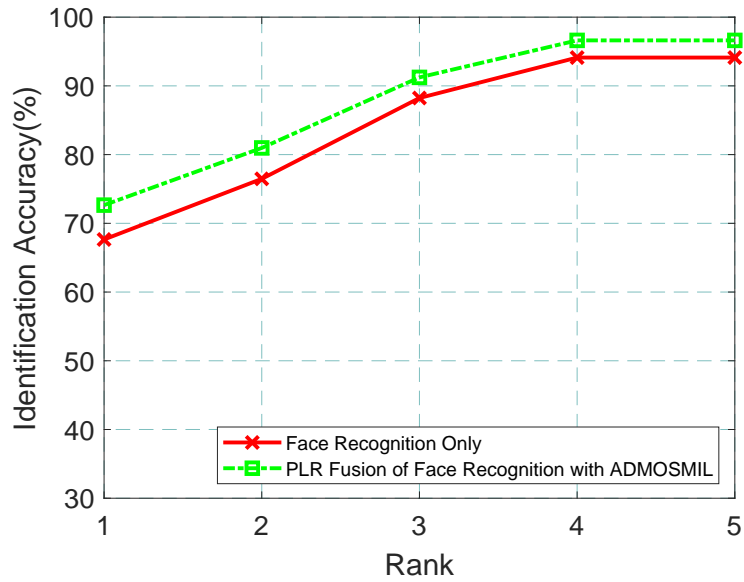
It should be noted that such an experiment can be conducted by using any existing face identification algorithm and the experimental evaluation reveals that by employing the face classification as illicit drug or regular face/plastic surgery or regular face as a soft biometric trait leads to improvement in the face identification performance.

6.7 Summary

This chapter presents DMOSMIL and ADMOSMIL algorithms for face affected by temporal variation. The proposed algorithms are based on a novel formulation to take advantage of both the multi-one shot similarity and multi-instance learning techniques. They can be utilized in the scenarios where the training data may be limited or where the objects are complex to be modeled using a single label. The experimental results on different real-world face covariate databases demonstrate that ADMOSMIL achieves superior face classification performance as compared to its constituents and existing metric learning approaches. Additionally, using the face classification scores from the proposed ADMOSIL algorithm as a soft biometric trait, we demonstrate a boost in the performance of existing face recognition algorithm.



(a) Faces Altered by Plastic Surgery



(b) Faces Affected by Illicit Drug Abuse

Figure 6.9: CMC curves demonstrating the increase in face recognition performance by utilizing ADMOSMIL scores as soft biometric.

Chapter 7

Conclusions and Future Work

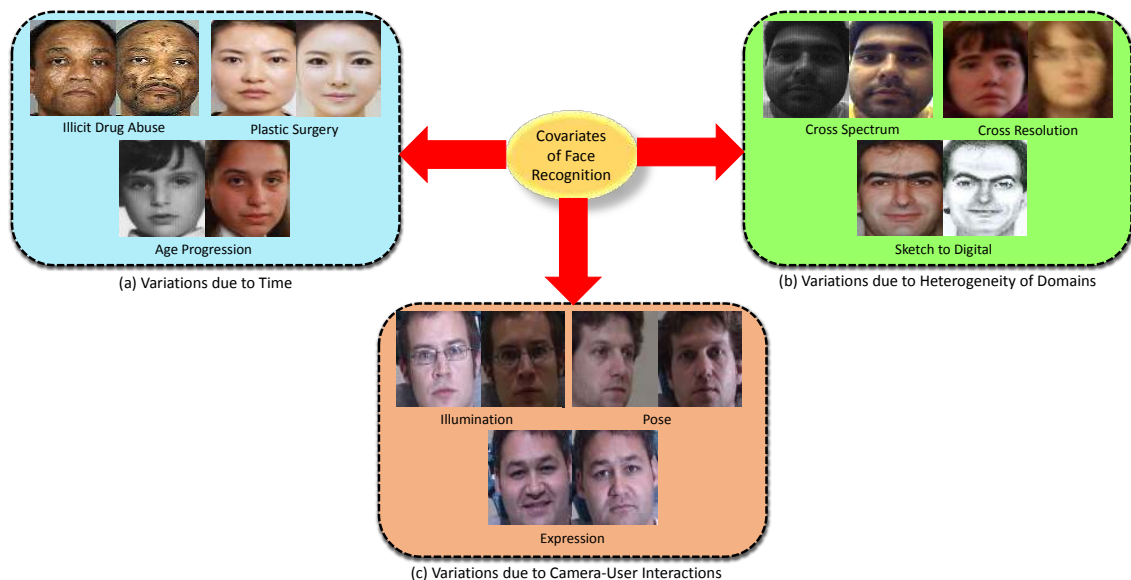


Figure 7.1: Types of covariates of face recognition: (a) covariates which impact the facial features due to temporal variations, (b) covariates due to heterogeneity of the face image domains, and (c) variations arising due to the interaction of the user with the image acquisition sensor.

Over recent years, many automatic algorithms have been developed which mitigate the effect of different covariates of face recognition. In our opinion, the covariates of face recognition can be classified into 3 categories:

- Covariates which impact the facial features due to temporal variations such as age progression, plastic surgery, and illicit drug abuse,
- Covariates which arise due to the heterogeneity of the face image domains such as cross spectral, cross resolution, and sketch-digital face matching, and
- Covariates which occur due to variability in user's interaction with the camera such as the change in pose, illumination, and expressions.

In this dissertation, we present a novel view of face recognition by considering faces affected by age progression, plastic surgery, and illicit drug abuse as covariates due to temporal variations. This research focuses on mitigating the influence of variations in facial features occurring over time due to facial aging, facial plastic surgeries, and prolonged illicit drug abuse. Developing temporal variation-invariant face recognition algorithms can prove to be highly beneficial in various applications such as locating missing people, homeland security, and passport services.

First, we designed two human studies to understand the human perception of facial aging. In the first study, behavioral analysis of more than 400 subjects is performed to examine the human ability of facial age estimation and age-separated face matching. Moreover, we also conducted the first study to utilize fMRI for analyzing the neural correlates of age-separated face verification task by humans. When controlled for task difficulty, the findings of this research revealed that the inferior parietal lobe was significantly active for age-separated stimuli face pairs. These findings suggested that separate cognitive units of the brain are associated with processing gradual temporal variations in age-separated face pairs, illustrating the need for building dedicated face recognition algorithms for processing such faces.

Inspired by the above findings, we proposed a novel deep learning based algorithm for matching faces with temporal variations caused by age progression. In the proposed algorithm, a unified approach combining facial age estimation and age-separated face verification with generative adversarial networks is utilized. The proposed GAN architecture learns the age variations across time by conditioning the input image on the subject's gender and the

target age group to which the input face needs to be progressed. To showcase the performance of the proposed algorithm, age-separated face recognition experiments were conducted on different face aging databases.

Next, we described plastic surgery as another component of temporal variations in face images. For handling such faces, we proposed a novel multiple projective dictionary learning (MPDL) framework to detect plastic surgery for face verification. Experimental results using the plastic surgery database showed the efficacy of our proposed MPDL framework. Detection accuracy of 97.96% is observed for plastic surgery faces as compared to commonly used features and classification algorithms. After a pair of reference and query images was detected as plastic surgery, the Stage-2 of the proposed MPDL framework verifies the identity of the individual. The proposed MPDL framework for face verification was seamlessly integrated with two commercial systems.

As part of this dissertation, we also introduced illicit drug abuse as a temporal variation based face recognition covariate and we presented the Illicit Drug Abuse Face (IDAF) dataset. Experimental evaluation revealed the decline in the performance of commercial face recognition algorithms as well as commonly used face descriptors when illicit drug abuse face images were to the database of regular faces. To detect such faces, we also proposed a detection framework to seamlessly classify in real-time using multiple dictionaries if a given face image is a regular face or a drug-abuse face. The proposed detection framework yields a classification accuracy of 88.81% when applied on a combined database of illicit drug abuse faces as well as regular faces.

Finally, we presented novel single-image based face classification algorithms for detecting faces with temporal variations, specifically, plastic surgery and illicit drug abuse. We presented DMOSMIL and ADMOSMIL algorithms for face classification which are based on DenseNet based multi-instance learning and multi-one shot similarity score. The experimental results demonstrated that ADMOSMIL achieves superior face classification performance as compared to its constituents and existing metric learning approaches. Additionally, using

the face classification scores from the proposed ADMOSIL algorithm as a soft biometric trait, we demonstrated an increase in the performance of existing face recognition algorithm.

We conclude this dissertation with future research directions which can be explored for enhancing face recognition performance under temporal variations.

- As described in the text, every individual has a personalized aging function which makes synthesizing *realistic* images, very challenging. Factors such as current age, gender, and genetics play a crucial in the facial age progression. The current architecture of age-gap reducer GAN can be enhanced by including other critical demographic information about the subjects, enabling the model to generate more personalized images.
- The problem of face recognition under plastic surgery and illicit drug abuse is impeded by the lack of any large-scale database which can be utilized for training a deep learning based model. To mitigate this, transfer learning or domain adaptation approaches can be employed to build an accurate face recognition model which is able to handle such variations.
- Developing video-based face recognition algorithms is another challenging research problem. The algorithms proposed in this dissertation for handling temporal variations in face images can be extended to the video domain. Such a temporal variations-invariant video face recognition system can be highly beneficial in surveillance based settings.
- Facial disguise can also be considered as a covariate with temporal variations which was not included in this thesis. Therefore, this research can be extended by augmenting the current experimental protocol to include these types of faces.

Bibliography

- [1] Faces of meth. <http://www.facesofmeth.us/>. [Online; accessed December-11-2018].
- [2] Rehabs.com. <http://www.rehabs.com>. [Online; accessed December-13-2018].
- [3] American Society of Plastic Surgeons. 2017. [Online; accessed December-9-2018].
- [4] United Nations Office on Drugs and Crime World Drug Report. 2017. [Online; accessed August-11-2018].
- [5] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer. A sparse representation approach to face matching across plastic surgery. In *IEEE Winter Conference on Applications of Computer Vision*, pages 113–119, 2012.
- [6] M. Aharon, M. Elad, and A. Bruckstein. KSVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
- [7] S. Ali, C. P. Mouton, S. Jabeen, E. K. Ofoemezie, R. K. Bailey, M. Shahid, and Q. Zeng. Early detection of illicit drug use in teenagers. *Innovations in Clinical Neuroscience*, 8(12):24, 2011.
- [8] D. G. Altman and J. M. Bland. Diagnostic tests. 1: Sensitivity and specificity. *BMJ*, 308(6943):1552, 1994.
- [9] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In *IEEE International Conference on Image Processing*, pages 2089–2093, 2017.
- [10] B. Ashcraft. How South Korean plastic surgeons make passport photos worthless. <http://kotaku.com/how-south-korean-plastic-surgeons-make-passport-photos-1563323919>, 2015.

- [11] H. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, and A. Noore. Evolutionary granular approach for recognizing faces altered due to plastic surgery. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, pages 720–725, 2011.
- [12] S. Biswas, G. Aggarwal, N. Ramanathan, and R. Chellappa. A non-generative approach for face recognition across aging. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2008.
- [13] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper. The CSU face identification evaluation system: its purpose, features, and structure. In *Computer Vision Systems*, pages 304–313. 2003.
- [14] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6 (1)(1):3–5, 2011.
- [15] C. Carbon, M. Grüter, and T. Grüter. Age-dependent face detection and face categorization performance. *PLoS ONE*, 8 (10)(10):e79164, 10 2013.
- [16] B. Chen, W. Deng, and J. Du. Noisy softmax: Improving the generalization ability of dcnn via postponing the early softmax saturation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4021–4030, 2017.
- [17] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6):804–815, 2015.
- [18] Y. Chen, J. Bi, and J. Z. Wang. MILES: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1931–1947, 2006.
- [19] V. Cheplygina, D. M. Tax, and M. Loog. Dissimilarity-based ensembles for multiple instance learning. *IEEE Transactions on Neural Networks and Learning Systems*, 27(6):1379–1391, 2016.
- [20] D. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units(ELUs). *CoRR*, abs/1511.07289, 2015.

- [21] Cognitec. Facevacs. <http://www.cognitec.com/technology.html>, 2018.
- [22] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.
- [23] A. Dantcheva, C. Velardo, A. D’angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, 2011.
- [24] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Robust face recognition after plastic surgery using local region analysis. In *Image Analysis and Recognition*, pages 191–200. 2011.
- [25] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Robust face recognition after plastic surgery using region-based approaches. *Pattern Recognition*, 48(4):1261–1276, 2015.
- [26] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLOS ONE*, 9(7):1–16, 07 2014.
- [27] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1):31–71, 1997.
- [28] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2012.
- [29] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, and M. Savvides. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In *IEEE International Conference on Computer Vision*, pages 3755–3763, 2017.
- [30] L. G. Farkas. *Anthropometry of the head and face*. New York: Raven Press, 2 edition, 1994.
- [31] B. Fink, M. Landthaler, and C. Hafner. Skin alterations due to illegal drug abuse. *Journal der Deutschen Dermatologischen Gesellschaft*, 9(8):633–639, 2011.
- [32] J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.
- [33] P. J. Flynn, K. W. Bowyer, and P. J. Phillips. Assessment of time dependency in face recognition: An initial study. In *International Conference on Audio-and Video-Based*

Biometric Person Authentication, pages 44–51, 2003.

- [34] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10 (4)(4):578–584, June 2008.
- [35] P. A. George and G. J. Hole. Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception*, 24 (9):1059–1073, 1995.
- [36] P. A. George and G. J. Hole. The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, 7 (4)(4):485–509, 2000.
- [37] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez. Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation. *IEEE Transactions on Information Forensics and Security*, 13(8):2001–2014, 2018.
- [38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [39] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.
- [40] P. J. Grother and M. L. Ngan. Face recognition vendor test (FRVT) performance of face identification algorithms NIST IR 8009. Technical report, 2014.
- [41] S. Gu, L. Zhang, W. Zuo, and X. Feng. Projective dictionary pair learning for pattern classification. In *Advances in Neural Information Processing Systems*, pages 793–801, 2014.
- [42] J. S. Guntupalli, K. G. Wheeler, and M. I. Gobbini. Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex*, 27(1):46–53, 2017.
- [43] G. Guo, G. Mu, and K. Ricanek. Cross-age face recognition on a very large database: The performance versus age intervals and improvement using soft biometric traits. In *International Conference on Pattern Recognition*, pages 3392–3395, 2010.
- [44] I. Gupta, I. Bhalla, R. Singh, and M. Vatsa. Scattering transform for matching surgically altered face images. In *IEEE International Conference on Pattern Recognition*, 2018.

- [45] C. Hawco, N. Kovacevic, A. K. Malhotra, R. W. Buchanan, J. D. Viviano, M. Iacoboni, A. R. McIntosh, and A. N. Voineskos. Neural activity while imitating emotional faces is related to both lower and higher-level social cognitive performance. *Scientific Reports*, 7(1):1244, 2017.
- [46] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [47] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In *Advances in Neural Information Processing Systems*, pages 507–514, 2006.
- [48] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan. Facial age estimation with age difference. *IEEE Transactions on Image Processing*, 26(7):3087–3097, 2017.
- [49] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 3, 2017.
- [50] M. Inc. Face++ research toolkit. www.faceplusplus.com, Dec. 2013.
- [51] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270 – 2285, 2005.
- [52] Z. Jiang, Z. Lin, and L. S. Davis. Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2651–2664, 2013.
- [53] R. Jillela and A. Ross. Mitigating effects of plastic surgery: Fusing face and ocular biometrics. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 402–411, 2012.
- [54] I. A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S. K. Shah, and T. Theoharis. 3D-2D face recognition with pose and illumination normalization. *Computer Vision and Image Understanding*, 154:137–151, 2017.
- [55] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *IEEE International Conference on Pattern Recognition*, pages 1363–1366, 2012.
- [56] N. M. Kleinmans, T. Richards, L. Sterling, K. C. Stegbauer, R. Mahurin, L. C. Johnson, J. Greenson, G. Dawson, and E. Aylward. Abnormal functional connectivity in autism

- spectrum disorders during face processing. *Brain*, 131(4):1000–1012, 2008.
- [57] N. Kohli, D. Yadav, and A. Noore. Multiple projective dictionary learning to detect plastic surgery for face verification. *IEEE Access*, 3:2572–2580, 2015.
- [58] H. Kosaka, M. Omori, T. Iidaka, T. Murata, T. Shimoyama, T. Okada, N. Sadato, Y. Yonekura, and Y. Wada. Neural substrates participating in acquisition of facial familiarity: an fMRI study. *NeuroImage*, 20(3):1734–1742, 2003.
- [59] H. Koshino, R. K. Kana, T. A. Keller, V. L. Cherkassky, N. J. Minshew, and M. A. Just. fMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. *Cerebral Cortex*, 18(2):289–300, 2008.
- [60] Y. H. Kwon and N. da Vitoria Lobo. Age classification from facial images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–767, 1999.
- [61] A. Lanitis. Comparative evaluation of automatic age-progression methodologies. *EURASIP Journal on Advances in Signal Processing*, 2008:239480, 2008.
- [62] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34 (1)(1):621–628, Feb. 2004.
- [63] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (4)(4):442–455, 2002.
- [64] S. Z. Li, D. Yi, Z. Lei, and S. Liao. The CASIA NIR-VIS 2.0 face database. In *Computer Vision and Pattern Recognition Workshops*, pages 348–353, 2013.
- [65] W. J. Li and D. Y. Yeung. MILD: Multiple-instance learning via disambiguation. *IEEE Transactions on Knowledge and Data Engineering*, 22(1):76–89, 2010.
- [66] Y. Li, G. Wang, L. Nie, Q. Wang, and W. Tan. Distance metric optimization driven convolutional neural network for age invariant face recognition. *Pattern Recognition*, 75:51–62, 2018.
- [67] Z. Li, D. Gong, X. Li, and D. Tao. Aging face recognition: a hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing*, 25(5):2146–2154, 2016.

- [68] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE Transactions on Information Forensics and Security*, 6 (3)(3):1028–1037, 2011.
- [69] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. A study of face recognition as people age. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.
- [70] C. H. Liu, M. A.-A. Bhuiyan, J. Ward, and J. Sui. Transfer between pose and illumination training in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4):939, 2009.
- [71] E. Y. Liu, Z. Guo, X. Zhang, V. Jovic, and W. Wang. Metric learning from relative comparisons by minimizing squared residual. In *IEEE International Conference on Data Mining*, pages 978–983, 2012.
- [72] H. Liu, J. Lu, J. Feng, and J. Zhou. Ordinal deep feature learning for facial age estimation. In *IEEE International Conference on Automatic Face Gesture Recognition*, pages 157–164, 2017.
- [73] H. Liu, J. Lu, J. Feng, and J. Zhou. Label-sensitive deep metric learning for facial age estimation. *IEEE Transactions on Information Forensics and Security*, 13(2):292–305, 2018.
- [74] S. W. Liu, M. H. Lien, and N. A. Fenske. The effects of alcohol and drug abuse on the skin. *Clinics in Dermatology*, 28(4):391 – 399, 2010. Nutrition and the Skin: Part I.
- [75] J. Lu, V. Liong, X. Zhou, and J. Zhou. Learning compact binary face descriptor for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(10):2041–2056, 2015.
- [76] S. Lu, M. Hämäläinen, R. Hari, R. Ilmoniemi, O. Lounasmaa, M. Sams, and V. Vilkmän. Seeing faces activates three separate areas outside the occipital visual cortex in man. *Neuroscience*, 43(2–3):287 – 290, 1991.
- [77] Luxand. Facesdk. <https://www.luxand.com/facesdk/>, 2015.
- [78] N. A. Macmillan and C. D. Creelman. *Detection theory: A user’s guide*. Psychology press, 2004.

- [79] J. Mairal, F. Bach, and J. Ponce. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):791–804, 2012.
- [80] A. E. Mayes, P. G. Murray, D. A. Gunn, C. C. Tomlin, S. D. Catt, Y. B. Wen, L. P. Zhou, H. Q. Wang, M. Catt, and S. P. Granger. Environmental and lifestyle factors associated with perceived facial age in Chinese women. *PLOS ONE*, 5 (12):e15270, 2010.
- [81] P. Mazaika, S. Whitfield-Gabrieli, A. Reiss, and G. Glover. Artifact repair for fMRI data from high motion clinical subjects. In *Human Brain Mapping*, 2007.
- [82] D. McNicol. *A primer of signal detection theory*. Psychology Press, 2005.
- [83] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
- [84] S. Nagpal, M. Singh, R. Singh, and M. Vatsa. Regularized deep learning for face recognition with weight variations. *IEEE Access*, 3:3010–3018, 2015.
- [85] K. H. Norwich. *Information, sensation, and perception*. Academic Press San Diego, 1993.
- [86] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.
- [87] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [88] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.
- [89] A. J. O’Toole, F. Jiang, D. Roark, and H. Abdi. Predicting human performance for face recognition. *Face Processing: Advanced Methods and Models*. Elsevier, 2006.
- [90] A. J. O’Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi. Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9), 2007.

- [91] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, and Z. Zhu. Robust face recognition via occlusion dictionary learning. *Pattern Recognition*, 47(4):1559 – 1572, 2014.
- [92] P. Pandey, R. Singh, and M. Vatsa. Face recognition using scattering wavelet under illicit drug abuse variations. In *International Conference on Biometrics*, pages 1–6, 2016.
- [93] U. Park and A. K. Jain. Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.
- [94] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (5)(5):947–954, 2010.
- [95] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.
- [96] O. Pascalis and D. J. Kelly. The origins of face processing in humans: Phylogeny and ontogeny. *Perspectives on Psychological Science*, 4(2):200–209, 2009.
- [97] A. M. Passarotti, B. M. Paul, J. R. Bussiere, R. B. Buxton, E. C. Wong, and J. Stiles. The development of face and location processing: An fMRI study. *Developmental Science*, 6(1):100–117, 2003.
- [98] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King. Aspects of age variation in facial morphology affecting biometrics. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2007.
- [99] Y. Peng and H. Yin. Facial expression analysis and expression-invariant face recognition by manifold-based synthesis. *Machine Vision and Applications*, 29(2):263–284, 2018.
- [100] R. A. Poldrack. Region of interest analysis for fmri. *Social Cognitive and Affective Neuroscience*, 2(1):67, 2007.
- [101] G. Pourtois, S. Schwartz, M. L. Seghier, F. Lazeyras, and P. Vuilleumier. View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study. *NeuroImage*, 24(4):1214–1224, 2005.
- [102] G.-J. Qi, J. Tang, Z.-J. Zha, T.-S. Chua, and H.-J. Zhang. An efficient sparse metric learning in high-dimensional space via l_1 -penalized log-determinant regularization.

- In *ACM International Conference on Machine Learning*, pages 841–848, 2009.
- [103] N. Ramanathan and R. Chellappa. Face verification across age progression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 462–469, 2005.
- [104] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 387–394, 2006.
- [105] N. Ramanathan and R. Chellappa. Modeling shape and textural variations in aging faces. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8, 2008.
- [106] A. S. Reece. Evidence of accelerated ageing in clinical drug addiction from immune, hepatic and metabolic biomarkers. *Immunity & Ageing*, 4(1):1–10, 2007.
- [107] M. G. Rhodes. Age estimation of faces: A review. *Applied Cognitive Psychology*, 23(1):1–12, 2009.
- [108] K. Ricanek Jr and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 341–345, 2006.
- [109] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops*, pages 10–15, 2015.
- [110] C. Sagonas, Y. Panagakis, S. Arunkumar, N. Ratha, and S. Zafeiriou. Back to the future: A fully automatic method for robust age progression. In *IEEE International Conference on Pattern Recognition*, pages 4226–4231, 2016.
- [111] C. Sagonas, Y. Panagakis, A. Leiding, S. Zafeiriou, et al. Robust joint and individual variance explained. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, page 6, 2017.
- [112] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.
- [113] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan. Personalized age progression with aging dictionary. In *IEEE International Conference on Computer Vision*, pages 3970–3978,

2015.

- [114] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [115] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyzedan. Plastic surgery: A new dimension to face recognition. *IEEE Transactions on Information Forensics and Security*, 5(3):441–448, 2010.
- [116] R. Singh, M. Vatsa, and A. Noore. Effect of plastic surgery on face recognition: A preliminary study. In *Computer Vision and Pattern Recognition Workshops*, pages 72–77, 2009.
- [117] V. Singh-Curry and M. Husain. The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, 47(6):1434 – 1448, 2009. Perception and Action.
- [118] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.
- [119] Y. Sun, M. Tistarelli, and D. Maltoni. Structural similarity based image quality map for face recognition across plastic surgery. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, 2013.
- [120] S. Suri, A. Sankaran, M. Vatsa, and R. Singh. On matching faces with alterations due to plastic surgery and disguise. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2018.
- [121] M. Swartz, J. Swanson, V. Hiday, R. Borum, H. Wagner, and B. Burns. Violence and severe mental illness: the effects of substance abuse and nonadherence to medication. *The American Journal of Psychiatry*, 155(2):226–231, 1998.
- [122] Y. Taigman, L. Wolf, T. Hassner, et al. Multiple one-shots for utilizing class label information. In *British Machine Vision Conference*, volume 2, pages 1–12, 2009.
- [123] M. Tistarelli, D. Yadav, M. Vatsa, and R. Singh. Short-and long-time ageing effects in face recognition. In M. Fairhurst, editor, *Age Factors in Biometric Processing*, chapter 14, pages 253–275. IET, 2013.

- [124] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 3, page 7, 2017.
- [125] L. Q. Uddin, I. Molnar-Szakacs, E. Zaidel, and M. Iacoboni. rTMS to the right inferior parietal lobule disrupts self–other discrimination. *Social Cognitive and Affective Neuroscience*, 1(1):65, 2006.
- [126] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2378–2386, 2016.
- [127] Y. Wang, Z. Zhang, W. Li, and F. Jiang. Combining tensor space analysis and active appearance models for aging effect simulation on face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):1107–1118, 2012.
- [128] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [129] Z. Wang, J. Yang, N. Nasrabadi, and T. Huang. A max-margin perspective on sparse representation-based classification. In *IEEE International Conference on Computer Vision*, pages 1217–1224, 2013.
- [130] Y. Wen, Z. Li, and Y. Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4901, 2016.
- [131] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Workshop on faces in ‘real-life’ images: Detection, alignment, and recognition*, 2008.
- [132] L. Wolf, T. Hassner, and Y. Taigman. The one-shot similarity kernel. In *IEEE International Conference on Computer Vision*, pages 897–902. IEEE, 2009.
- [133] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

- [134] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018.
- [135] C. Xu, Q. Liu, and M. Ye. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing*, 222:62–71, 2017.
- [136] Y. Xu, J.-Y. Zhu, E. Chang, and Z. Tu. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 964–971, 2012.
- [137] D. Yadav, N. Kohli, P. Pandey, R. Singh, M. Vatsa, and A. Noore. Effect of illicit drug abuse on face recognition. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1–7, 2016.
- [138] D. Yadav, R. Singh, M. Vatsa, and A. Noore. Recognizing age-separated face images: Humans and machines. *PLOS ONE*, 9(12):1–22, 12 2014.
- [139] D. Yadav, M. Vatsa, R. Singh, and M. Tistarelli. Bacteria foraging fusion for face recognition across age progression. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 173–179, 2013.
- [140] H. Yang, D. Huang, Y. Wang, and A. K. Jain. Learning face age progression: A pyramid architecture of GANs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [141] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *IEEE International Conference on Computer Vision*, pages 543–550, 2011.
- [142] M. Yang, L. Zhang, X. Feng, and D. Zhang. Sparse representation based Fisher discrimination dictionary learning for image classification. *International Journal of Computer Vision*, 109(3):209–232, 2014.
- [143] X. Yu, J. Huang, S. Zhang, W. Yan, and D. Metaxas. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In *International Conference on Computer Vision*, pages 1944–1951, 2013.

- [144] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.
- [145] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.
- [146] Q. Zhang and B. Li. Discriminative K-SVD for dictionary learning in face recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 2691–2698, 2010.
- [147] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *The IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, 2017.
- [148] T. Zheng, W. Deng, and J. Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *CoRR*, abs/1708.08197, 2017.
- [149] J.-Y. Zhu, W.-S. Zheng, F. Lu, and J.-H. Lai. Illumination invariant single face image recognition under heterogeneous lighting condition. *Pattern Recognition*, 66:313–327, 2017.
- [150] K. Zwosta, H. Ruge, and U. Wolfensteller. Neural mechanisms of goal-directed behavior: outcome-based response selection is associated with increased functional coupling of the angular gyrus. *Frontiers in Human Neuroscience*, 9, 2015.