

THE UNIVERSITY OF NEWCASTLE, AUSTRALIA

DOCTORAL THESIS

Affective Analysis of Visual Scenes using Face Pareidolia and
Scene-Context

Asad ABBAS
BSc (EE), MS (EE)

*A thesis submitted in partial fulfilment of the requirements
or the degree of Doctor of Philosophy in Computer Science*

in the

Interdisciplinary Machine Learning Research Group

School of Electrical Engineering and Computing

Faculty of Engineering and Built Environment

June 2021

Statement of Originality

I hereby certify that the work embodied in the thesis titled, 'Affective Analysis of Visual Scenes using Face Pareidolia and Scene-Context' is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Signed: Asad Abbas

Date: 22/10/2020

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of The University of Newcastle's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Acknowledgement of Authorship

I hereby certify that the work embodied in this thesis contains published paper/s/scholarly work of which I am a joint author. I have included as part of the thesis a written declaration endorsed in writing by my supervisor, attesting to my contribution to the joint publication/s/scholarly work.

By signing below I confirm that Asad ABBAS contributed to the paper/ publication entitled:

- **A. Abbas** and S.K. Chalup, “Affective Analysis of Visual Scenes using Face Pareidolia and Scene-Context”, published in Neurocomputing, Volume 437, 2021, Pages 72-83, ISSN 0925-2312, doi.org/10.1016/j.neucom.2021.01.016
- **A. Abbas** and S.K. Chalup, “From Face Recognition to Facial Pareidolia: Analysing Hidden Neuron Activations in CNNs for Cross-Depiction Recognition”, accepted in International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary on July 14-19, 2019. Publisher: IEEE, DOI: [10.1109/IJCNN.2019.8852013](https://doi.org/10.1109/IJCNN.2019.8852013)
- **A. Abbas** and S.K. Chalup, “The Impact of Image Resolution on Facial Expression Analysis with CNNs”, accepted in International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary on July 14-19, 2019. Publisher: IEEE, DOI: [10.1109/IJCNN.2019.8852264](https://doi.org/10.1109/IJCNN.2019.8852264)
- **A. Abbas**, S.K. Chalup, “Group Emotion Recognition in the Wild by Combining Deep Neural Networks for Facial Expression Classification and Scene Context Analysis”, ICMI’17, Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, Scotland (2017). Publisher: ACM, DOI: [10.1145/3136755.3143010](https://doi.org/10.1145/3136755.3143010)

by:

- Contributing to each study’s conception and design
- Developing analysis plans

- Developing research material and collecting data
- Performing both quantitative and qualitative analyses
- Writing code and execution of programs
- Interpreting data
- Leading the writing of the manuscripts

Signed: Stephan Chalup

Date: 02-11-2020

Acknowledgements

I would like to express my deepest appreciation to my supervisor A/Prof. Stephan K Chalup for supporting and motivating me throughout my research journey. Without his guidance and persistent help, the objectives of this thesis would not have been accomplished.

I would also like to extend my gratitude to members of IMLRG for their insightful suggestions and constructive criticism. Thanks should also go to UON HPC team especially Aaron Scott and Geoff Martin for providing technical support.

I am also grateful to my cricket team-mates for creating such wonderful memories and providing a way to reduce my stress levels when the going got tough during my thesis. Special thanks to Abdul Jabbar, Arif Hussain, Uzair Khan, Syed Qaisar Jalil, Usman Asghar, Ashfaq Ahmad, Bilal Ahmad, Malik Ahsan, Ajmal Nawaz and Mohsin Javed for their relentless support all the way from the cricket pitch to the lab computer.

I cannot leave the University of Newcastle without mentioning Gulam Dastagir Khan, Muhammad Javed and Prakash Mallick for lively discussions, constructive comments and practical suggestions. I'm also deeply indebted to my friend, Junaid Anwar, for his encouragement and helpful advice throughout the duration of this thesis.

Finally, I cannot begin to express my sincere gratitude to my family, my father Muhammad Ramzan; my mother, Kaneez Fatima; my brothers, Saqib Taimoor and Farhan Abbas; my sister-in-law, Mahjabeen and my loving wife, Mehwish Batool for their ongoing unparalleled support, guidance and profound belief in my abilities.

Contents

Statement of Originality	i
Acknowledgement of Authorship	ii
Acknowledgements	iv
Contents	v
List of Figures	viii
List of Tables	x
Abstract	xii
1 Introduction	1
1.1 Research Objectives	4
1.2 Research Publications	5
1.3 Structure of the Thesis	5
2 Background	8
2.1 Cognitive Science and Psychology	8
2.1.1 Phenomenon of Face Pareidolia	8
2.1.2 Face Pareidolia in History	9
2.1.3 Unconscious Processing of Emotions	13
2.1.4 Unconscious Modulation of Emotions	15
2.2 Affective Computing	16
2.2.1 Facial Action Coding System (FACS)	17
2.2.2 Categorical Model	18
2.2.3 Dimensional Model	20

2.2.4	Facial Expression Recognition (FER)	20
2.2.5	Facial Expression Datasets	22
2.3	Convolutional Neural Networks	28
2.3.1	Neuroscientific Inspirations for Convolutional Networks	29
2.3.2	The Convolution Operation	30
2.3.3	Structure of Convolutional Layer	34
2.3.4	Training of Convolutional Neural Networks	34
2.3.5	Visualisation Techniques	36
2.4	Face Pareidolia Simulation	42
2.4.1	Unavailability of Face Pareidolia Dataset	43
2.4.2	Abstract Phenomenon of Face Pareidolia	44
2.4.3	Limited Generalisation Ability of Fully Supervised Detectors	45
3	Group Emotion Recognition	46
3.1	Proposed Approach	47
3.1.1	Dataset	48
3.2	Experiments	49
3.2.1	Facial Emotion Prediction	49
3.2.2	Context Prediction	52
3.2.3	Overall Group-level Emotion Recognition	54
3.3	Results and Discussion	55
3.4	Chapter Summary	59
4	Impact of Image Resolution on Facial Expression Analysis with CNNs	60
4.1	Experiments	61
4.1.1	Datasets	61
4.1.2	CNN Models	64
4.1.3	Methodology	65
4.2	Results and Discussion	66
4.2.1	Performance on Meta-Dataset	66
4.2.2	Performance on Test Databases	67
4.2.3	Visualisation of CNNs	68
4.3	Chapter Summary	71
5	From Face Recognition to Facial Pareidolia	72
5.1	Our Approach	74
5.2	Experimental Protocol	77
5.2.1	CNN Model	77
5.2.2	Deep Feature Visualisation	78
5.2.3	Dimensionality Reduction Techniques and Scatter Plots	78
5.3	Results and Discussion	79
5.3.1	T1: Deep Feature Visualisations and Clustering of Neurons	79

5.3.2	T2: Simulating Facial Pareidolia using Hidden Neurons	84
5.4	Chapter Summary	84
6	Affective Analysis of Visual Scenes using Face Pareidolia and Scene-Context	86
6.1	Datasets	89
6.1.1	Face Pareidolia Simulation Datasets	89
6.1.2	Emotion Recognition Datasets	90
6.2	Proposed Approach	91
6.2.1	Face Pareidolia Simulation	91
6.2.2	Emotion Recognition	95
6.3	Experiments	97
6.3.1	Face Pareidolia Simulation	97
6.3.2	Face Pareidolia Analysis	98
6.3.3	Scene-Context Analysis	99
6.3.4	Ablation Study:	100
6.4	Results and Discussion	102
6.4.1	Quantitative Results	102
6.4.2	Qualitative Results	106
6.5	Chapter Summary	106
7	Discussion and Conclusion	108
7.1	Group Emotion Recognition (Chapter 3)	110
7.2	Impact of Image Resolution (Chapter 4)	110
7.3	Face Pareidolia Simulation (Chapter 5)	111
7.4	Affective Visual Perception (Chapter 6)	112
7.5	Further Work	112
	References	115

List of Figures

1.1	Emergence of face pareidolia in selected example images	3
2.1	Pebble of many faces	10
2.2	Our Lady of Guadalupe	12
2.3	Allah written in Arabic in a watermelon	13
2.4	Action Units (AUs) of the facial expressions of basic emotions	19
2.5	Two-dimensional valence arousal circumplex	20
2.6	Convolution operation with a kernel matrix	31
2.7	Comparison between regular fully connected layer and convolutional layer	31
2.8	Comparison between fully connected and sparsely connected network	32
2.9	Receptive field of neurons in shallow layers vs. deeper layers	33
2.10	Example of max pooling	34
2.11	Topology of convolutional layer	35
2.12	Overview of literature on deep visualisation techniques.	39
2.13	Deep feature visualisation techniques employed on selected neurons	41
2.14	Two-dimensional activation projections	42
2.15	Different depiction styles of faces	44
3.1	Proposed approach for group emotion recognition	48
3.2	Training and validation accuracies for CNN trained from scratch	51
3.3	Training and validation accuracies for scene-context analysis	54
3.4	Training and validation accuracy for overall group-level emotion recognition	55
3.5	Confusion Matrix for VGG16 network	56
3.6	Confusion Matrix for InceptionV3 network	57
3.7	Example image from Flickr correctly classified as Positive	58
3.8	Example image from Flickr correctly classified as Negative	58
3.9	Example image from Flickr correctly classified as Negative in absence of any face detection	59
4.1	Preprocessing output using OpenFace	65
4.2	Validation accuracy of CNN models on meta-dataset	67
4.3	Confusion matrices of VGG200 and VGG400 CNN models	69
4.4	Saliency map visualisations for high and low intensity emotions	70
4.5	Saliency map visualisations for complex emotions	71

5.1	Emergence of face pareidolia in selected example images	73
5.2	Proposed approach for face recognition to face pareidolia simulation	76
5.3	Deep feature visualisations of selected neurons in layer conv5_3	79
5.4	Projection of neuron activations of convolutional layer conv5_3 using dimensionality reduction.	80
5.5	Clustering of neuron activations of convolutional layer conv5_3 based on similarity measure.	81
5.6	Clustering using MDS	82
5.7	Deep visualisations of selected neurons in conv5_2	82
5.8	Selected outliers from layer conv5_3 with associated deep visualisations.	83
5.9	Results of face pareidolia detection	85
6.1	Architecture of proposed deep learning framework	88
6.2	Dataset annotations for source and target domains	92
6.3	Techniques for synthetic sample generation	93
6.4	Proposed approach for three step weakly supervised progressive domain adaptation	94
6.5	Prediction results of deep learning framework	104
6.6	Face pareidolia occurrences in architectural facades and automobiles.	105

List of Tables

2.1	FACS - Facial Action Coding System	18
2.2	EMFACS - Emotional Facial Action Coding System	19
2.3	Facial Expression Datasets	28
3.1	Dataset distribution including additional training samples for Group Emotion Recognition	49
3.2	Configuration of CNN trained from scratch to predict facial emotions	51
3.3	Comparison of validation accuracies of the proposed models	56
4.1	Composition of the training/validation meta-dataset	63
4.2	Composition of the test dataset	63
4.3	Configurations of the CNN models	64
4.4	Accuracies of CNN models on Test Databases	68
5.1	Configuration of the VGG16 model trained for face recognition	77
6.1	Distribution of datasets used in our experiments	90
6.2	Relative contribution of each domain adaptation step across detectors on the FIP dataset	101
6.3	Comparison of different object detectors on the FIP dataset.	102
6.4	Valence and Arousal RMSE values on Validation and Test sets	103

Dedicated to my Parents

Muhammad Ramzan and Kaneez Fatima

For their endless love, support and encouragement

Abstract

With recent advancements in affective computing, computers have become better at describing human emotional responses using facial expression and scene-context analysis. However, the topic of face pareidolia, which is the imagination of non-existent faces in random patterns, its computational modelling and application in product design is recently gaining increased attention in the research community. This thesis aims to answer an interdisciplinary research question, “Can we simulate the face pareidolia ability of humans and predict associated emotional responses using deep learning?” This interdisciplinary thesis question was based on research literature published in the disciplines of Cognitive Science and Psychology. Findings from these research disciplines suggest that abstract and minimal face-like patterns can subconsciously activate face-selective regions in the brain producing face pareidolia. Moreover, face pareidolia, along with various other contextual variables present in the visual scene, can trigger an emotional response in our neurobiological systems without conscious awareness. The first part of this thesis investigates a group emotion recognition task by using an ensemble of convolutional neural networks. The findings from this part are consistent with the literature in cognitive science and psychology that an individual’s face and associated facial expressions are always perceived and evaluated within a surrounding context. The second part of this thesis presents two research studies addressing the poor generalisation capacity of convolutional neural networks, as poor generalisation capacity limits the direct application of deep learning techniques to simulate face pareidolia. In the first study, the impact of image resolution on the generalisation capacity of convolutional neural networks trained for facial emotion recognition was analysed. The second study was inspired by face perception in the brain and improves the generalisation capacity of a convolutional neural network, trained for face recognition, to simulate the facial pareidolia capability of the human visual system. The third part of this thesis consolidates previous findings and presents a novel deep learning-based cross-domain weakly supervised three-step progressive domain adaptation framework. The proposed framework can simulate face pareidolia and predict associated emotional responses in two-dimensional valence and arousal space for various product designs. The proposed framework was evaluated on a new image dataset. Both quantitative and qualitative experimental results show that our approach can outperform other state-of-the-art methods.