Spring 5-2020

Connecting the Dots for People with Autism: A Data-driven Approach to Designing and Evaluating a Global Filter

Viseth Sean
Chapman University, sean103@mail.chapman.edu

Connecting the Dots for People with Autism: A Data-driven Approach to Designing and Evaluating a Global Filter

A Dissertation by

Viseth Sean

Chapman University

Orange, California

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Data Sciences

May 2020

Committee in charge:

LouAnne Boyd, Ph.D., Committee Chair

Deanna Hughes, Ph.D., Committee Member

Vincent Berardi, Ph.D., Committee Member

CHAPMAN UNIVERSITY

SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY

Computational and Data Sciences

The dissertation of Viseth Sean is approved.

LouAnne Boyd, Ph.D., Committee Chair

Deanna Hughes, Ph.D., Committee Member

Vincent Berardi, Ph.D., Committee Member

May 2020

Connecting the Dots for People with Autism: A Data-driven Approach to Designing and Evaluating a Global Filter

by Viseth Sean

DEDICATION

To my dear late grandfathers:

Pisean Than, thank you for your unconditional love and protection.

Chhivky Tiv, thank you for always believing in me and your advice to thrive as a beloved son, student, and person in our community. Your last advice that I will always emulate is "Never freak out even if a mountain falls on you. Think and you will be able to solve any problem."

ACKNOWLEDGMENTS

I am sincerely grateful to the people who made this dissertation possible.

I would like to thank my advisor, **Dr. LouAnne Boyd**, for all the inspiration and guidance. You introduced me to this interesting and meaningful topic, and provided me sound directions and the prompt feedback that propelled my research. Thanks for your effort and humor, as well as the late-night meetings and dinners we convened in order to meet conference deadlines. Your enthusiasm in research and attitude of not taking no as an answer speak volumes about your hard work and eagerness to continue learning. I express my sincere thanks for your support, advice, patience, and encouragement throughout this important last year of my graduate studies. I have been fortunate to have you as my advisor and friend.

I am also thankful to **Dr. Deanna Hughes** and **Dr. Vincent Berardi** for agreeing to be on my committee with such short notice. The first time we met in person was at my topic proposal presentation, and you have been kind, cheerful, and supportive. Thank you, Deanna, for your clinical expertise and constructive feedback in polishing this dissertation. Thank you, Vincent, for your comments and guidance on the data analysis.

I would like to acknowledge **Dr. Hesham El-Askary** for the Computational and Data Sciences (CADS) Fellowship, and **Dr. Erik Linstead** for the Experian Scholarship. Thanks for providing me the funding throughout my Ph.D. journey.

I would like to thank my collaborators: **Dr. Franceli Cibrian**, **Jazette Johnson**, **Brandon Makin**, **Hollis Pass**, **Dr. Eliza DelPizzo-Cheng**, **Sara Jones**, and **Karen Lotich**. I have enjoyed working together with you all throughout this project.

Thanks to the CADS faculty and staff for their instruction, knowledge, and assistance.

Particularly, I am very thankful to **Robin Pendergraft**, the previous program coordinator for her tireless and genuine assistance.

I am very grateful to my friends who stand beside me through thick and thin throughout my time at Chapman University. **Natalie Best** and **Jordan Ott**, thank you for being my besties and introducing me to your families. **Sam Ford**, thank you for always being supportive and sharing the room for holding our TA office hours. **Chelsea Parlett-Pelleriti** and **Chris Watkins**, thank you for answering my statistics questions. **Ismael Paiva**, thank you for guiding me through my first year at Chapman. **Daniel Choi**, thank you for hanging out a bunch with me, from studying in the library to taking on Harry Potter marathon. **Esther Shin**, thank you for being the best neighbor and taking me to buy groceries. **Abdulah Haikal**, you are my first friend at Chapman even though you were not a Chapman student. Thanks for lending me your car for my driving test. **Dacoda Strack**, thanks for your technical assistance and introducing me to camping.

A huge thank to my grandmothers: **Chhay Ya Ho** and **Tho Phuong**, and all of my aunts, uncles, and cousins. I am very proud to have you as my family. Thank you for your support and being my biggest fans.

I am extremely grateful to my dear brother and sister-in-law, **Visal Sean** and **Sovath Ngin**. Thank you for your love, support, laughter, and your genuine hospitality whenever I visited you in Seattle. You are great travel buddies, and I was fortunate to have travelled with you to many beautiful places in Washington state.

I owe a great debt of gratitude to my beloved father and mother, **Seng An Sean** and **Kunthea Tiv**, who constantly love and support me throughout my life. You are the ones who I can always turn to at anytime. You are my best counselors. I know it has never been easy as I moved thousands of miles away to pursue this dream, but you

have been nothing but supportive. Your confidence in me and encouragement were in the end what made this journey possible. For that, and everything else, I eternally thank you.

VITA

Viseth Sean

EDUCATION

Doctor of Philosophy in Computational and Data Sciences — **2020**

Chapman University — *Orange, California*

Master of Science in Data Science — **2016**

Worcester Polytechnic Institute — *Worcester, Massachusetts*

Bachelor of Science in Computer Science — **2012**

Royal University of Phnom Penh — *Phnom Penh, Cambodia*

Global UGRAD Exchange Program in Computer Science — **2011**

East Tennessee State University — *Johnson City, Tennessee*

RESEARCH AND WORK EXPERIENCE

Graduate Research/Teaching Assistant — **2017–2020**

Chapman University — *Orange, California*

Data Science Intern — **2019**

Alignment Healthcare — *Orange, California*

Research Intern — **2018**

Travelers Insurance Company — *Hartford, Connecticut*

Research Intern — **2017**

HP Labs — *Palo Alto, California*

Research Intern **2016**

HP Labs *Palo Alto, California*

Software Developer **2013–2014**

ACLEDA Bank Plc. Headquarters *Phnom Penh, Cambodia*

SELECTED HONORS AND AWARDS

Graduate Fellowship and **2017–2020**

Teaching/Research Assistantship

Chapman University *Orange, California*

Fulbright Scholarship **2014–2016**

The US Department of State *United States of America*

Hesselbein Global Leadership Academy **2015**

University of Pittsburgh *Pittsburgh, Pennsylvania*

Dean's Award for Excellence and **2014**

Full Scholarship

Worcester Polytechnic Institute *Worcester, Massachusetts*

Outstanding Project Pitch Award **2014**

YSEALI Workshop initiated by *Kuala Lumpur, Malaysia*

President Barack Obama

Gold Medal of The Honda Y-E-S Award **2013**

Honda Foundation *Tokyo, Japan*

Full Scholarship **2008–2012**

Royal University of Phnom Penh *Phnom Penh, Cambodia*

LIST OF PUBLICATIONS

PUBLICATIONS

Viseth Sean, Jazette Johnson, Franceli L. Cibrian, Hollis Pass, Eliza DelPizzo-Cheng, Sara Jones, Karen Lotich, Brandon Makin, Deanna Hughes, and LouAnne Boyd. 2020. Designing, developing, and evaluating a global filter to work around local interference for children with autism. *In CHI '20.*

Franceli L. Cibrian, Jazette Johnson, **Viseth Sean**, Hollis Pass, and LouAnne E. Boyd. 2020. Combining eye tracking and verbal response to understand the impact of a global filter. *In CHI '20.*

Viseth Sean, Deanna Hughes and LouAnne Boyd. 2020. Combining HCI and AI in designing and evaluating a global filter to empower people with autism to see the big picture. *In CHI '20 Workshop AI4HCI.*

Viseth Sean, Franceli Linney Cibrian, Jazette Johnson, Hollis Pass, and LouAnne E. Boyd. 2019. Toward digital image processing and eye tracking to promote visual attention for people with autism. *In UbiComp/ISWC '19.*

Viseth Sean. 2016. Exploration Framework For Detecting Outliers In Data Streams. *Master's Thesis, Worcester Polytechnic Institute.* https://digitalcommons.wpi.edu/etd-theses/395

TECHNICAL PRESENTATIONS

Sunil Kothari, **Viseth Sean**, Juan Catana, Jun Zeng, Gary Dispoto. 2018. A data-driven framework for HP's MJF 3D printer for correlation and prediction of part mechanical properties. *In the 29th Annual International Solid Freeform Fabrication Symposium.*

PATENTS

Sunil Kothari, Kristopher Li, **Viseth Sean**, Jun Zeng, Lihua Zhao, Goffril Obegi, Gary J. Dispoto, Tod Heiles. 2019. Predicting distributions of values of layers for three-dimensional printing. *US Patent.*

Sunil Kothari, Jun Zeng, Kristopher Li, Goffril Obegi, Lihua Zhao, Gary J. Dispoto, **Viseth Sean**, Tod Heiles. 2019. Design rules for printing three-dimensional parts. *US Patent.*

ABSTRACT

Connecting the Dots for People with Autism: A Data-driven Approach to Designing and Evaluating a Global Filter

by Viseth Sean

"Social communication is the use of language in social contexts. It encompasses social interaction, social cognition, pragmatics, and language processing" [3]. One presumed prerequisite of social communication is visual attention–the focus of this work. "Visual attention is a process that directs a tiny fraction of the information arriving at primary visual cortex to high-level centers involved in visual working memory and pattern recognition" [7]. This process involves the integration of two streams: the global and local streams; the global stream rapidly processes the scene, and the local stream processes details. This integration is important to social communication in that attending to both the global and local features of a scene are necessary to grasp the overall meaning. For people with autism spectrum disorder (ASD), the integration of these two streams can be disrupted by the tendency to privilege details (local processing) over seeing the big picture (global processing) [66]. Consequently, people with ASD may have challenges integrating visual attention, which may disrupt their social communication. This doctoral work explores the hypothesis that visual attention can be redirected to the features of an image that contain holistic information about a scene, which when highlighted might enable people with ASD to see the forest as well as the trees (i.e., seeing a scene as a whole rather than parts). The focuses are on 1) designing a global filter that can shift visual attention from local details to global features, and 2) evaluating the performance of a global filter by leveraging eye-tracking technology. This doctoral work manipulates visual stimuli in an effort

to shift the visual attention of people with ASD.

This doctoral work includes two development life cycles (i.e., design, develop, evaluate): 1) low-fidelity filter, and 2) high-fidelity filter. The low-fidelity filter life cycle includes the design of four low-fidelity filters for an initial experiment which was tested with an adult participant with ASD. The performance of each filter was evaluated by using verbal responses and eye-tracking data in terms of visual analysis, fixation analysis, and saccade analysis. The results from this cycle informed the decision for designing a high-fidelity filter in the next development life cycle. In this second cycle, ten children with ASD participated in the experiment. The performance of the high-fidelity filter was evaluated by using both verbal responses and eye-tracking data in terms of eye gaze behaviors. Results indicate that baseline conditions slightly outperform global filters in terms of verbal response and the eye gaze behaviors.

To unpack the results in more details beyond group comparisons, three analyses (e.g., luminance, chroma, and spatial frequency) of image characteristics are performed to ascertain relevant aspects that contribute to the filter performance. The results indicate that there are no significant correlations between the image characteristics and the filter performance. However, among the three characteristics, spatial frequency is depicted as the most correlated factor with the filter performance. Additional analyses using neural networks, specifically Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN), are also explored. The result shows that CNN is more predictive of the relationship between an image and visual attention than MLP. This is a proof of concept that neural networks can be employed to identify images for future experiments, by avoiding any variance or bias in terms of unbalanced characteristics of images across the experimental image pool.

TABLE OF CONTENTS

Page

DEDICATION **IV**

ACKNOWLEDGMENTS **V**

VITA **VIII**

LIST OF PUBLICATIONS **X**

ABSTRACT **XII**

TABLE OF CONTENTS **XIV**

LIST OF TABLES **XVIII**

LIST OF FIGURES **XX**

1 Introduction **1**

1.1 Innovation of Assistive Technologies 2

1.2 Characteristics Associated with Autism 3

1.3 Nonverbal Behavior 5

1.4 Autism and Social Perception Challenges 6

1.5 Visual Attention 7

1.6 Autism and Global-Local Processing 9
1.7 Hypothesis 11
1.8 Research Questions 12
1.9 Contributions of the Dissertation 12
1.10 Summary of the Following Chapters 13

2 Related Work on Eye-tracking Technology 15

2.1 Eye-tracking Hardware 16
2.1.1 Screen-based Eye Tracker 16
2.1.2 Eye-tracking Glasses 17
2.1.3 Eye tracking-enabled VR Headsets 18
2.2 Eye-tracking Data 19
2.2.1 Gaze Points 19
2.2.2 Fixations 20
2.2.3 Saccade/Scanpath 21
2.2.4 Area of Interest (AOI) 22
2.2.5 Heat Maps 23
2.2.6 Video Logs 25
2.3 Cluster Fix Algorithm 26
2.4 Eye-tracking Technology and Autism 27

3 Preliminary Work: Designing Low-fidelity Filters 30

3.1 Methods 30
3.1.1 Design of Filters 31
3.1.2 Participants 38
3.1.3 Low-fidelity Prototype 38

3.1.4 Procedure . . . 39
3.2 Evaluating Filters with Eye-tracking Data . . . 39
3.2.1 Visual Analysis . . . 39
3.2.2 Fixation Analysis . . . 43
3.2.3 Saccade Analysis . . . 45

4 High-fidelity Filter 50
4.1 Design Methods . . . 51
4.1.1 Providing Filter Options to SLPs . . . 53
4.1.2 Participants . . . 56
4.1.3 Study Procedure . . . 56
4.1.4 Study Design . . . 59
4.1.5 Running the Sessions . . . 60
4.1.6 Study Setup . . . 62
4.2 Data Analyses . . . 62
4.3 Results . . . 65
4.3.1 Verbal Responses (Subjective Data) . . . 65
4.3.2 Eye Gaze Behaviors (Presumed Objective Data) . . . 65
4.4 Discussion . . . 66

5 Analysis on Characteristics of Experimental Images 68
5.1 Luminance . . . 69
5.2 Chroma . . . 72
5.3 Spatial Frequency . . . 72
5.4 Regression Analysis . . . 77
5.5 Machine Learning/Deep Learning Analysis . . . 80

5.5.1 Multi-Layer Perceptron (MLP) 82
5.5.2 Convolutional Neural Network (CNN) 86
5.6 Summary 89

6 Conclusion and Future Work **92**

BIBLIOGRAPHY **96**

APPENDICES **103**

LIST OF TABLES

Page

3.1 Navon's test result for P1. 38

3.2 Performance score of different filters in visual analysis. 42

3.3 Fixation analysis–the number of overlaps between P1's fixations and the areas of increasing visual angles from NT fixations as seen in Figure 3.8. 46

4.1 Participant demographics in the high-fidelity filter study as described by their SLPs. 58

4.2 Verbal responses to the prompt: "What was the picture about?" for the baseline/original picture in Figure 4.4 Top. For this trial, baseline was seen first. 60

4.3 Verbal responses for the filtered picture in Figure 4.4 Bottom. 61

4.4 Results of verbal responses across conditions and sessions in the high-fidelity study. 65

5.1 The correlation matrix between hit count and the characteristics of the images: luminance, chroma, and spatial frequency. 78

5.2 The summary result of multiple linear regression with hit count as the output variable, and the estimator variables are the standardized values of luminance and spatial frequency (R-squared=0.035). 79

5.3 The performance comparison between multiple linear regression and Poisson regression. Lower AIC and BIC values are preferred. 79

5.4 The performance comparison between individual-image CNN model and combined-images CNN model in terms of the mean and standard deviation (std.) of absolute percentage difference between predicted and actual hit count. 89

A.1 The name of each image that is shown below this table. 103

A.2 The sorted order of baseline and filtered images from the highest to the lowest hit count. 116

A.3 The values of spatial frequency for each image. 153

LIST OF FIGURES

Page

1.1 Sample items similar to items on the Navon's test, 1977. The top item is the target. The bottom items are the choices to choose from when a person is asked to find the match. The bottom left shares local features with the top whereas the bottom right figure shares the global feature. 10

2.1 Examples of screen-based eye trackers. Top image is the EyeLink 1000 Plus device by SR Research. Bottom image is the EyeLink Portable Duo device by SR Research [1]. 17

2.2 A participant, during the experiment with high-fidelity filters, is wearing the eye-tracking glasses by Positive Science. The participant's face is blurred for privacy concern. 18

2.3 An example of VR eye-tracking device by Pupil Labs. 19

2.4 An example of gaze points. Green x's denote gaze points of a participant that were captured in the experiment for that particular image. 20

2.5 An example of fixations. Blue x's denote fixations. Green x's denote eye gazes. 21

2.6 An example of saccade. Green lines denote saccade between fixations, which are denoted by blue x's. 22

2.7 Top is a baseline condition of a living room image. Bottom is the corresponding heat map of the living room image where dark blue represents low to no fixations. 24

2.8 Screenshots of an eye-tracking video log captured during high-fidelity filter experiment. 25

3.1 A sample of image from the experiment in Baseline condition (raw image). 31

3.2 This image shows the AOIs by overalying NT heat map on the corresponding image in Figure 3.1. It shows the global features which are the man and the bird. 32

3.3 An example of Lined Edges filter. 33

3.4 An example of White Background filter. 34

3.5 An example of Grey Blurred filter. 35

3.6 An example of Animation filter, starting with Lined Edges, White Background, Grey Blurred, and finally Baseline. 37

3.7 A heat map (of Figure 3.1) represents the AOIs (yellow regions with a green boundary) that are aggregated from 15 NT people from [87], overlaid by eye gazes (blue x's) of the man with ASD (P1). 41

3.8 An animation of evaluating P1's fixations compared to NT fixations. Red dots represent fixations of NT people. Blue dots represent fixations of the man with ASD, P1. Blue dots turn to green when they are in the areas of the incremental visual angle of NT people. 44

3.9 A White Background image (top), a raw image (middle), and a Grey Blurred image (bottom), overlapping with the hotspots (in green/yellow) and red arrows showing the second saccade of P1. 47

4.1 Top image is the raw or original version of two boys sitting on the beach. Middle image is the heat map based on eye fixations of NT people. Bottom image is the filtered image that is desaturated and blurred. This technique was applied to the semantic heat map. 52

4.2 Samples of the automated filters with a blurred background: a) mild blur, b) moderate blur, and c) severe blur. Mild and moderate blurs were chosen for further automation. 54

4.3 Samples of the automated filters: a) desaturated without blur, b) desaturated with mild blur, and c) desaturated with moderate blur. The desaturated with moderate blur filter was selected for the high-fidelity filter experiment. 55

4.4 Top: Screenshot of P2's eye-tracking video in baseline with cross hairs outside the image at the top left of the screen. Bottom: Screenshot of P2's eye-tracking video in the filtered condition with cross hairs in the center and within the AOIs indicated by the heat map. 57

4.5 Flowchart of global filter experiment paradigm where Part A and B occur in different sittings. The 50 images shown in Part A are the same images as Part B but are counterbalanced to be in the other condition they appear in Part A. The order is randomized for both Part A and B. 63

5.1 The top 3 highest average luminance images: a) filtered image of two boys sitting on the beach, b) baseline image of the two boys sitting on the beach, and c) filtered image of two men playing baseball in the field. These images also have the highest average chroma value. . . . 70

5.2 The top 3 lowest average luminance images: a) baseline image of a dog sitting at a table, b) baseline image of two people walking on the beach with two sailboats, and c) baseline image of a puppy with his toys. These images also have the lowest average chroma value. 71

5.3 An example of low spatial frequency image with five exact same bars in horizontal space. 73

5.4 An example of high spatial frequency. The image contains twice as many bars as in the low spatial frequency image in Figure 5.3. 73

5.5 The top 3 highest mean spatial-frequency images: a) five puppies with vertical stripes in the background, b) a man is walking by a brick wall of a grocery store, and c) a man and a dog are running on the beach. These images are all baseline images. 75

5.6 The top 3 lowest mean spatial-frequency images: a) a bathroom with a toilet, b) a boy with a goat , and c) a puppy standing on a barrel. These images are all filtered images. 76

5.7 An example of a neuron of ANN. Variables x_1, x_2, and x_3 are the inputs into the neuron; y is the output from the neuron. 81

5.8 The architecture of MLP used in this analysis. It consists of input layer, 2 hidden layers, and output layer. The input layer takes in inputs: luminance, chroma, and spatial frequency. The first hidden layer consists of 8 neurons and the second hidden layer consists of 4 neurons. The output layer takes in outputs from the previous layer (i.e., second hidden layer) and outputs the final result which is the hit count. 83

5.9 The learning curves of training and testing loss versus the number of epochs for the two-variable MLP model. The x-axis represents the number of epochs, and the y-axis represents the loss (i.e., MAPE). . . 86

5.10 The architecture of CNN used in this analysis. It consists of an input layer, three convolutional layers, three fully connected layers, and an output layer. Note that the normalization layers and pooling layers are not shown due to limited space. 88

5.11 The learning curves of train and test loss versus the number of epochs for the combined-images CNN model. The x-axis represents the number of epochs, and the y-axis represents the loss (i.e., MAPE). 90

A.1 The detailed CNN architecture that is employed in section 5.5.2. . . . 156

Chapter 1

Introduction

Over the past decade, Data Science has become a revolutionary technology in both academic research and the technology industry. It has proven to be beneficial in many fields such as supply chain optimization, finance, biomedicine, bioinformatics, natural sciences, social networks, smart cities, education, energy, sustainability and climate, health science, etc. The goal of Data Science is to extract, analyze, and visualize data to create insights which help make powerful data-driven decisions. This dissertation applies Data Science to human challenges, specifically in social communication in individuals with autism spectrum disorder (ASD), to explore the insights from data analysis to inform the design of innovative assistive technologies. As computing has become ubiquitous in the age of mobile computing, personal and assistive devices have turned to big data for solutions to problems that range not only from new factors such as wearable devices, to new sensors such as IoT sensory devices, but also to new insights from Data Science collected by these same devices.

1.1 Innovation of Assistive Technologies

Data Science drives the design of assistive technologies. Historically, assistive technology has addressed functional limitations beginning with rehabilitating soldiers after war to functional skill training for the disabled [86]. In terms of skill development for disabled people, functional goals have historically taken precedence over improving life experiences [86]. Still today, more often than not, new technologies are designed to diagnose or otherwise differentiate autistic from typical behavior as is the focus of behavior intervention journals [74]. Novel technology for autism has focused primarily on addressing medical symptoms that shape deficit behavior into more normative behavior. In a recent meta analysis of Human-Computer Interaction (HCI) projects for autism found that the majority of assistive technology projects in the field of HCI are interested in mediating or re-mediating core symptoms [74]. A critical review of the research reveals most technologies aimed to support users with ASD attempt to change social behavior to be more normative [74]. Projects aim to have therapeutic value at a behavioral or cognitive level to change specific social skill deficits to resemble normative behavior in skills such as emotion recognition, conversation skills, and turn taking. In other words, assistive technology tends to focus on supporting functional skills rather than improving life experiences.

Recent work calls for attention to be paid to supporting issues such as emotional and sensory regulation, communication, motor coordination, executive functioning, and sensory processing [85]. A few studies have addressed these areas. For example, the design of wearable applications has been explored to support self-control of behavior of children with Attention-Deficit/Hyperactivity Disorder (ADHD) [55]. Other projects have explored the use of multisensory interactive displays as a therapeutic device to support the motor development and sensory processing of children with ASD [55] and found that natural user interfaces in tandem with multisensory

stimuli are easy to use and useful for children with severe autism. However, there are less explored opportunities to use technology to help users with ASD in areas of challenge such as visual attention. These emerging efforts in HCI address a broader range of skills and experiences and aim to improve the quality of life. The evolution of assistive technology has benefited from the Data Science revolution in that access to information allows for new types of assistive technologies–smart technologies that can respond in real time or with algorithmic decision-making capabilities. It is the intersection of innovation and insight from which this work springs.

1.2 Characteristics Associated with Autism

The current clinical definition of autism in the Diagnostic and Statistical Manual 5^{th} edition (DSM-5) states that autism is a spectrum disorder, indicating that there are three levels of severity across two axes of behavior [8]. The three diagnostic levels of autism are: 1) severely impacted, 2) moderately impacted, and 3) mildly impacted. The two axes of behavior are: restrictive, repetitive behavior and social communication impairments. Restrictive, repetitive behaviors are episodes during which a person repeatedly engages in a motor movement such as lining up toy cars, with no obvious purpose beyond meeting an unidentified, internal need of the person [8]. Some behavioral clinicians interpret these behaviors as problematic because they appear unusual and are often distressing for parents and teachers to witness. Engaging in these self-stimulating behaviors, on the other hand, often provides relief to people with ASD. Tension is created between the person needing to regulate themselves and the people in the environment who are distressed by the autistic-looking behavior. Many behavioral programs attempt to decrease the repetitive, non-functional behaviors as they are deemed to be disruptive. The assumption is if someone is engaged in

self-stimulation, then they are not available to learn from the environment.

The second axis in the diagnostic criteria for autism are social communication impairments. Social communication impairments refer to understanding as well as using spoken and nonspoken (nonverbal) communication. An example of a social communication skill is joint attention–including "sharing attention (e.g., through the use of alternating eye gaze), following the attention of another (e.g., following eye gaze or a point), and directing the attention of another" [22]. Joint attention is a skill that at 3 to 4 years of age, has been used to distinguish children with ASD from those with development delays [21]. Social communication differences have been described in literature working with people who have challenges with empathizing. When a person does not act in an expected way to demonstrate their understanding of others' experiences, society assumes that the person lacks empathy [9, 38]. The challenge of demonstrating expected social behaviors persists across a lifetime. The social perception of people with ASD has been characterized as: early in life "aloofness," school age years "socially avoidant," and lastly adulthood as simply "odd" [31]. To adjust to this broad range of social challenges over a lifetime, many researchers, educators, and clinicians have focused on developing social skill interventions through a range of delivery agents including parents, peers, highly trained therapists, and technology (e.g., video modeling) [62]. This variation of delivery agents is desirable, because the challenges with social communication manifest in a variety of ways across the lifespan requiring varying support by stakeholders. Therefore, social communication supports require flexibility over the course of a day and over a lifetime to be adaptive to dynamic social contexts and an individual's changing needs.

A dynamic part of social interactions is the nonverbal communication that consists of both reading as well as using body language, gestures, facial expressions, and tone of voice. Nonverbal deficits in autism have been described as difficulty with:

making eye contact, entering a group, reading body language, using body language, and understanding facial expressions [62]. This collection of ephemeral behaviors could provide a context for intervention systems as some form of nonverbal behavior is constantly present in face-to-face interactions. Intervention systems are capable of constant monitoring while permitting the user to engage only when needed.

1.3 Nonverbal Behavior

Nonverbal skills are a critical point for communication and intervention [48]. Nonverbal communication begins as soon as a person approaches another person, makes eye contact, positions their body in relation to others, and continues as one speaks and listens. Therefore, nonverbal behavior may occur as the first behavior in an interaction and requires global local integration to make sense of the interaction. One's tone of voice and use of gestures and body language all convey messages about the intentions of the people in the interaction. For example, an attempt to exert dominance over a group might be apparent when someone remains standing while others are sitting. These unspoken dynamics are difficult to understand and implement for those with ASD. Emerging technologies need to support an understanding of the dynamic nature of social interactions across the variety of stakeholders. Therefore, the focus of assistive technology for nonverbal communication is concerned with the "interplay between agents," incorporated through multiple modes of interactions that result in the quality of engagement and reciprocity [61]. Additionally, consideration needs to be given to the temporal patterns of interaction [77], implicit in face-to-face interaction.

Technologies that support nonverbal skills and support one's interaction partner could build a bridge between the normative and neurodiverse experiences. Nonverbal com-

munication is the first point of contact in face-to-face interaction, thus an ideal place to target supports. Successful interactions are the result of both parties, sometimes with a greater degree of effort being made on one's conversational partner; this work may be hidden by computer mediated communication [13]. As we are living in "neuro-shared spaces," both on-line and off-line, supports for society are very much needed because "(e)ngagement with the majority culture is often necessary to enable broad cultural change and potentially acceptance" [68]. This work aims to support non-verbal communication, which requires global processing, by leveraging findings in ASD research, eye-tracking research, digital image processing technologies, and eye-tracking technology.

1.4 Autism and Social Perception Challenges

Differences in eye-tracking patterns have been found in infancy between children who are diagnosed with ASD and typically developing children. Researchers have identified genetic links to actively seeking social stimuli such as human faces [19] thus, demonstrating the potential for biological causes for differences in the behavior of eye contact. Whether biology plays a role or not, these differences are important in that reduced attention to social stimuli has been linked to difficulties in developing communication abilities [72]. In fact, current research is examining the correlations between receptive language ability and attention to social stimuli. Researchers suggest there is a "possibility of training social attention allocation to promote the development of other abilities, including those related to understanding and using language" [72]. Precedence for viewing autism through a cognitive, biological, behavioral, and communication lens is well established. This work considers the sensory perception differences–a more recent addition to the diagnostic label. Sensory challenges can sig-

nificantly impact behavior and cognition [66]. Therefore, this work takes the stance that sensory issues should be supported even before behavioral of cognitive issues are addressed.

This doctoral works supports sensory perception in a way that is specific to visual attention to local and global features of an image–with the assumption that cognitive abilities are intact and can be utilized if/when the hurdle of visual attention is minimized. This charge is taken up in this current study. For example, researchers who know where a person is looking when speaking about a television program can make these aspects visually brighter, which improves the comprehension of the person who is listening to the speaker [64]. So, the manipulation of the visual stimuli could have an impact on the cognitive process [64]. Leveraging these insights, this work aims to work around local interference (local features overwriting global features) that has been reported in decades of psychology literature [39, 72, 66, 34, 14]. Ultimately, this work aims to facilitate visual attention to the semantic (global) features to support social communication at the onset of an interaction.

1.5 Visual Attention

Challenges with sensory integration have been linked to the social challenges in autism. Social interactions are comprised of rich, dynamic exchanges of information that command differential attention. Deciding what and who to attend to in a scene is based on what aspects of the scene are relevant for that moment and over time. What is relevant can change rapidly. Therefore, with each new task (e.g., social exchange) most people quickly decide what is relevant to attend to and what can be ignored. Attending to relevant features of a task requires an overall understanding of the task, or a sense of the "big picture." From there, one's attention can focus on

only the details that are required based on the nature of a task. This process occurs automatically as sensory information is continually processed in both a top-down and bottom-up manner [5]. The global meaning, or gist, of a scene occurs pre-attentively, with relevant local details being processed if the task requires, and irrelevant details being suppressed [73]. This distinction is important because it is a critical point that can be leveraged for people with sensory processing challenges. Specifically, manipulating parts of images that are processed in the pre-attentive stage opens up new opportunities to support sensory integration through novel assistive technologies [66].

Characteristics of both the viewer and task (i.e., top-down) as well as features of the stimulus (i.e., bottom-up) draw one's initial eye gaze. These first few fixations of a scene are believed to be involuntary. Once a gist is formed, attention shifts between details and the big picture. Attentional gaze shifts have been found to also be influenced by the emotional content of a scene and its visual structure–specifically, spatial frequency which captures the scale of the contrast density [34]. This tendency to attend to emotion and spatial frequency differentially has been found to not occur in the case of ASD [23]. Although eye gaze in a live social interaction has recently been found to be functional in a few autistic children, [41, 15], autistic eye gaze has been repeatably reported as different. Previous research employing participants with ASD revealed perceptual differences in attending to spatial frequency suggesting local (high spatial frequency) over global features (low spatial frequency) are visually attended to in both social and nonsocial stimuli [11]. This work aims to modify the stimulus features as sensory processing occurs first and "(t)he bottom-up control of attention by those [sensory] features is largely involuntary" [83]. Therefore, the environment needs to be changed to shift visual attention towards global features. This work aims to provide a sensory-perceptual "work-around" to guide the viewer to attend visually to the pre-processed stimuli thereby augmenting bottom-up processing by reducing the visual processing workload. This work differs from most eye-tracking

interventions for ASD in that it does not target behavior change for the purpose of shaping neurotypical (NT) behaviors. Rather, this work aims to manipulate visual attention of static scenes to promote global processing (e.g., pre-process) in an effort to promote learning and communication. The conceptualization, design, development, and evaluation of a global filter is described here within.

1.6 Autism and Global-Local Processing

Global processing is the rapid processing of a scene to get a holistic understanding of an object, event, or scene (i.e., seeing a forest before seeing the trees). Detecting overall shape, proximity, and context provides a sense of the "holistic view" [54]. In contrast, local processing allows for rapid processing of the details, and has been referred to as "analytic processing" [54]. The global and local streams of information become integrated during cognitive processing to produce a complete mental representation of the stimuli [20]. This guides visual attention, which is a set of cognitive processes that filter relevant from the irrelevant information in a visual scene. Many researchers claim that the average person first processes global information by taking in the gist or holistic view, and then integrates the global view with the local details within a fraction of a second [10, 16, 52, 51]. In artificial conditions, such as Navon's hierarchical letters [52] (see Figure 1.1), one's default precedence can be determined by contrasting local and global features. For example, Figure 1.1 top image depicts a hierarchical letter that has a global feature (i.e., overall shape) that is the letter F, and a local picture is the many small-sized letter Z's. People are typically quicker at detecting the F than Z [52]. The task in Navon's test is to choose one of the bottom images that matches to the top image (see Figure 1.1). The first image at the bottom shares local features with the top image while the second image share global features

with the top image. If a person is asked to find a match between the top and bottom images and the answer is the first image at the bottom, they are focusing on local features; and if they select the second, they are focusing on global features.

Figure 1.1: Sample items similar to items on the Navon's test, 1977. The top item is the target. The bottom items are the choices to choose from when a person is asked to find the match. The bottom left shares local features with the top whereas the bottom right figure shares the global feature.

However, research has shown that some people with ASD visually process the world differently, and the global and local streams may not be integrated smoothly due to a tendency to prioritize local details [66]. This has been verified repeatedly by using the Navon's test in research studies [28, 30, 34, 39]. The lack of attention to the global features may result in missed social information. For example, interpreting face to face social interactions requires rapid integration of global details (e.g., body

language, emotion recognition) with the relevant local details (e.g., attending to a bleeding scratch on someone's face instead of number of freckles). As social interaction contains multiple streams of information across multiple modalities, people with ASD can miss key elements of social information. Missing global information can result in challenges with a variety of social communication skill such as entering and leaving conversations, responding on a topic, interpreting a speaker's intent, making social connections and friendships [66]. This work aims to filter out less socially-relevant local details and highlight the global features of visual scenes to help people with ASD understand the global information of an image (i.e., gist).

This work contributes both a low and high fidelity global filter that manipulates images to highlight global aspects. The design phases were completed in collaboration with a field site. The prototypes were tested with autistic children who display developmental language delay. The children viewed and verbally responded to baseline and filtered images. Findings reveal that this particular group of children performed well in the baseline condition as well as in the filtered condition–particularly in the high-fidelity user study. Additional analysis of image features was conducted and a predictive model was created. Implications for future work is provided.

1.7 Hypothesis

By manipulating the stimuli using the eye-tracking findings of NT people as a template for global features, a global filter can guide people with ASD to shift their visual attention.

1.8 Research Questions

1. How to design a global filter to shift visual attention?
   (a) How to identify the global area of interest for a given image?
   (b) How to create filter design (low fidelity)?
   (c) How to automate/create a data-driven filter?
2. How to effectively evaluate the performance of visual attention of people with ASD with the use of a global filter?
   (a) Which methods and approaches identify visual attention of people with ASD through eye-tracking technology?
   (b) Which concepts connect eye fixations and/or saccades to global vs. local processing?
   (c) Which statistical methods and/or machine learning/deep learning approaches best predict visual attention for a given image?

1.9 Contributions of the Dissertation

The contributions of this dissertation work include:

1. A proof of concept that demonstrates the feasibility of using a filter to shift gaze path to global features. Since poor global and local integration can lead to social challenges [21], filtering a scene could serve as an assistive technology that augments the process thus empowering individuals with ASD through access to visual information.

2. To the best of my knowledge, this is the first work in designing a technology to work around local interference (local features overwrite global features) for individuals with ASD. Implications of this concept for filtering images could inspire HCI researchers to consider new ways to support other neurodiverse conditions like ADHD, dyslexia, etc.

3. An application of Data Science/AI applied to HCI by predicting the performance of a given image.

1.10 Summary of the Following Chapters

In chapter 2, I provide the summary of related work including the details of eye-tracking technology, eye-tracking data, the types of technology and data that are used in this dissertation work, the Cluster Fix algorithm that is used to identify important eye-tracking metrics including saccades and fixations, experimental image dataset, atypical visual saliency in ASD, and the effect of spatial frequency on eye gaze shifting.

In chapter 3, I discuss the preliminary work–low-fidelity design. It contains two important tasks for designing and evaluating the proposed global filters. The first task is about designing low-fidelity filters: how the filters were created and the rationale of choosing each filter for experimentation. The second task focuses on the evaluation of each filter by using verbal responses and leveraging eye-tracking technology; in particular, eye-tracking data of an adult with ASD was captured while he viewed the filters.

In chapter 4, I discuss the high-fidelity filter study including design methods, data analyses, results, and discussion.

In chapter 5, I discuss further analysis that is completed to unveil the characteristics of images in the experimental image pool in terms of luminance, chroma, and spatial frequency. Also, I provide additional analyses using Machine Learning techniques, specifically two Deep Learning models namely Multi-Layer Perceptron and Convolutional Neural Network.

In chapter 6, I provide the conclusion around all the studies collectively. And, I lay out the future studies that can push this work forward.

Chapter 2

Related Work on Eye-tracking Technology

Eye-tracking technology is primarily used to detect and capture the activities of eye movements such as eye gazes, fixations, and saccade (see section 2.2), that are analyzed to understand human visual behaviors. Recent advances in technology have made great strides for improving eye trackers in terms of precision and affordability, and they have been used in different fields of research such as neuroscience, psychology, industrial engineering and human factors, marketing/advertising, and computer science [24]. The range of eye-tracking behaviors have been used for various purposes. For example, eye gaze has been used in online information retrieval research to understand the pattern of human behavior in navigating the information on the web [33], whereas fixations provide information about which details were taken in, and saccades show the movement from one area to another. Gaze path provides information regarding the course of moving the eyes across a scene.

2.1 Eye-tracking Hardware

Eye trackers are categorized into three main types: screen-based eye trackers, eye-tracking glasses, and eye tracking-enabled Virtual Reality (VR) headsets. Each of them is recommended to use in different settings. The following subsections provide more information on the eye-tracking devices available on the market and specify those that were used in the experiments.

2.1.1 Screen-based Eye Tracker

A screen-based eye tracker, also known as desktop eye tracker, detects visual attention in controlled environments. It allows understanding of visual attention by tracking where one looks on a screen such as images, videos, websites, games, software interfaces, etc. This type of eye tracker is best used for observations of any screen-based stimuli in a lab setting because the respondent has to be seated in front of the eye trackers. Oftentimes, a chin rest is used with screen-based eye tracker to minimize participant's head movements while keeping the participant in focal range of the eye tracker. Sometimes, additional equipment like a forehead rest is also used in combination with a chin rest to stabilize participant's head. Figure 2.1 shows some screen-based eye trackers, and a participant is using a chin rest in the bottom image. This screen-based eye tracker is used in the preliminary study.

Figure 2.1: Examples of screen-based eye trackers. Top image is the EyeLink 1000 Plus device by SR Research. Bottom image is the EyeLink Portable Duo device by SR Research [1].

2.1.2 Eye-tracking Glasses

Eye-tracking glasses, also known as a head-mounted eye tracker, are wearable devices that allow us to understand how one views and interacts in the physical world. Researchers use this type of eye tracker to measure visual attention outside of a laboratory setting; respondents are able to walk around freely with the glasses. See Figure 2.2 for a sample of eye-tracking glasses. These eye-tracking glasses are used in the high-fidelity study.

Figure 2.2: A participant, during the experiment with high-fidelity filters, is wearing the eye-tracking glasses by Positive Science. The participant's face is blurred for privacy concern.

2.1.3 Eye tracking-enabled VR Headsets

Eye tracking-enabled VR headsets are wearable devices that capture visual attention in virtual environments. VR eye trackers offer the possibility of conducting experiments in a world that is no longer bound by factors such as time, safety, and budget. Researchers in [18] provide an introduction to eye tracking in VR and a guide to set up experiments with eye tracking-enabled VR headsets. See Figure 2.3 for a sample device of VR eye tracker currently on the market.

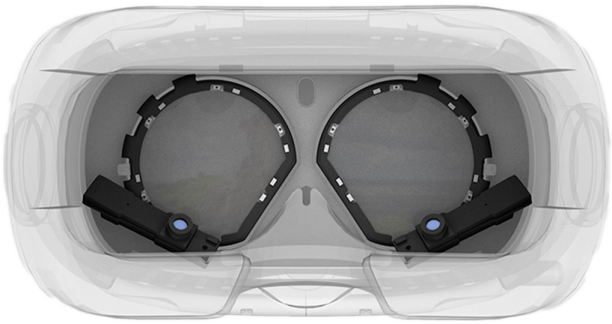


Figure 2.3: An example of VR eye-tracking device by Pupil Labs.

2.2 Eye-tracking Data

Depending on the type of eye tracker, oftentimes, the eye-tracking data outputs include eye positions, also known as gaze points, that allow us to assess visual attention. Instead of the raw gaze points (described in section 2.2.1), many researches have used different eye-tracking data metrics in their analyses. A comprehensive guide to the metrics can be found in this book [35]. The eye-tracking data metrics that are used in this dissertation are described next.

2.2.1 Gaze Points

Gaze points are the positions of the eyes, that show where/what the eyes are looking at. If an eye tracker captures data with a sampling rate of N (number) Hz, that means N gaze points per second. Gaze points are then clustered into metrics such as fixations and saccades (described in sections 2.2.2 and 2.2.3 respectively). Figure 2.4 show an example of gaze points.

Figure 2.4: An example of gaze points. Green x's denote gaze points of a participant that were captured in the experiment for that particular image.

2.2.2 Fixations

A fixation is a cluster of a series of gaze points that is very close in time and/or space; it denotes a period of time where the eyes are fixated towards a particular object. Fixation is a popular eye-tracking metric because it shows what grabbed the attention of a participant, or what is focused for a period of time when the eyes are relatively stationary because they are taking in information [60]. Also, fixations are considered the most informative metric mainly because compared to other metrics like saccades (see section 2.2.3), which happen too rapidly for the eyes to assimilate information [60]. As a result, the use of fixations minimizes the complexity in analyzing eye-tracking data while maintaining its important characteristics for understanding cognitive and visual behavior [69]. See Figure 2.5 for an example of eye fixations. As you can see,

a cluster of eye gaze points (green x's) makes up a fixation (blue x's).

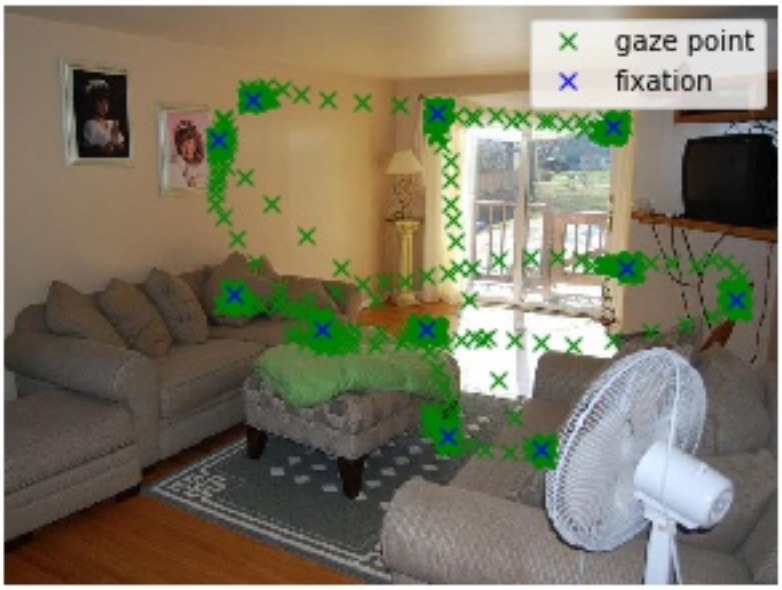


Figure 2.5: An example of fixations. Blue x's denote fixations. Green x's denote eye gazes.

2.2.3 Saccade/Scanpath

Saccade, also known as a scanpath, refers to rapid eye movements between fixations. A saccade is the fastest movement in human body; visual information is suppressed during this movement, i.e., visual stimuli are not processed when the eyes are in rapid motion. Figure 2.6 shows an example of saccade between each fixation. Saccades have been used in research to indicate which areas of an image have been scanned. Scan path is used to understand the sequence of scanning. These metrics have not been frequently used in autism research but could yield insight into global-local processing in this work (as described in section 3.2.3).

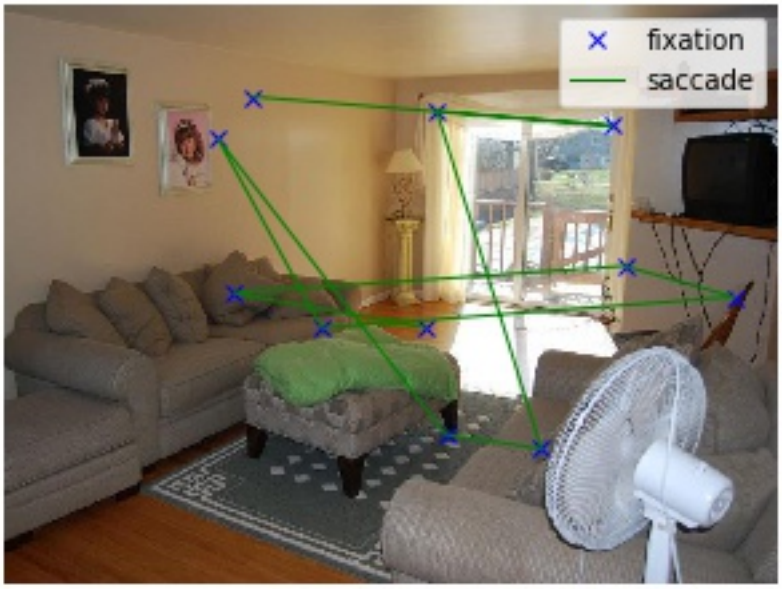


Figure 2.6: An example of saccade. Green lines denote saccade between fixations, which are denoted by blue x's.

2.2.4 Area of Interest (AOI)

An area of interest (AOI), also called a "hotspot," is a region of an image that is identified before a study as a target for visual attention. These areas are then used as a variable in the analysis. AOIs help researchers understand where people might be looking at and therefore thinking about. The purpose is to gather insight specifically for that region. In this work, the AOIs are where early NT fixations lay on a given image. The bright yellowish/greenish areas in Figure 2.7 represent the AOIs in the context (as no red areas are present).

2.2.5 Heat Maps

Heat maps are visualizations of the general distribution of gaze points. Heat maps are typically seen as a gradient overlay on a picture where color spectrum is used to indicate amount. For the heat maps in this work, yellow is the most viewed area and blue is the least. See Figure 2.7 for an example of a heat map.

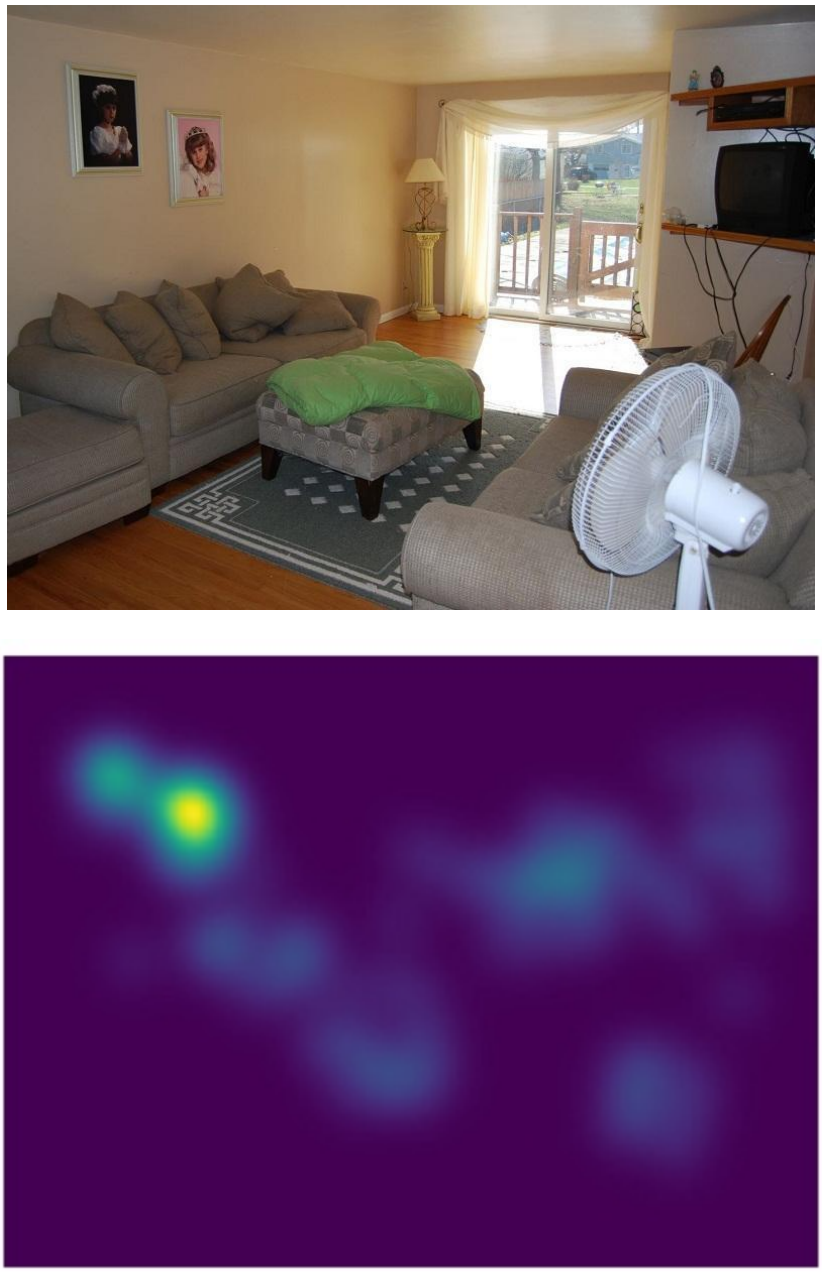


Figure 2.7: Top is a baseline condition of a living room image. Bottom is the corresponding heat map of the living room image where dark blue represents low to no fixations.

2.2.6 Video Logs

Video logs show visual behavior of viewers in video format. Figure 2.8 shows a video log sample of two screenshots from the high-fidelity filter experiment.

Figure 2.8: Screenshots of an eye-tracking video log captured during high-fidelity filter experiment.

2.3 Cluster Fix Algorithm

Eye gaze behavior is commonly parsed into fixations and saccades by using different algorithms. Traditional algorithms use eye acceleration, dispersion, or velocity thresholds to separate and label visual behavior into saccades and fixations. Sulvucci et al. classified algorithms for computing saccades and fixations into three categories: velocity-based, dispersion-based, and area-based [69]. To identify the occurrences of saccades, researchers have employed velocity and/or acceleration thresholds as eye velocity and acceleration are much greater during a saccade compared to a fixation [36, 53, 56]. Berg et al. and Liston et al. also employed a velocity-based algorithm for identifying saccades, then applied a principal component analysis technique for differentiating between saccades, smooth pursuit and noise [12, 45] (smooth pursuit and noise are not in the scope of this dissertation work). Another eye gaze algorithm identifies saccades and fixations of a viewer for a given image by employing dispersion and projection clustering [79].

However, König and Buffalo pointed out the limitations of these algorithms [40]. First, using velocity and acceleration thresholds for detecting saccades and fixations are not enough for complex oculomotor tasks including viewing of natural scenes and dynamic stimuli without any constrain, i.e., more variables have to be included in identifying saccades and fixations. Also, there is inconsistency in results due to the use of arbitrary thresholds in previous algorithms. As a result, König and Buffalo proposed a new algorithm called Cluster Fix, to address the limitations of previous algorithms.

Cluster Fix was employed to build the Object and Semantic Images and Eye-tracking (OSIE) dataset [87] that contains fixations of 15 NT people viewing 700 images. This current work also utilized the Cluster Fix algorithm to generate saccades and fixations

of the participants to be consistent with the OSIE dataset that was employed to build global filters and evaluate them. The details of the procedural outline of Cluster Fix algorithm can be found in the original paper [40], and the source code in MATLAB can be found here [42].

The Cluster Fix algorithm employs k-means clustering techniques based on four eye movement variables: distance, velocity, acceleration, and angular velocity, to identify fixations. Distance is "Euclidian distance between the position of the scan path at a time point to the position of the scan path two time points later" [40]. Velocity and acceleration are calculated "as the first and second derivative of position, respectively" [40]. Angular velocity is computed "as the difference in the angle of the scan path from one time point to the next" [40]. Cluster Fix utilizes the average silhouette width (a built-in MATLAB function SILHOUETTE) to automatically determine the number of clusters (k). Similar to general k-means clustering intuition, Cluster Fix first globally assesses the whole scan path, identifies saccades and fixations, and then locally re-assesses each saccade and fixation pair in order to detect small, short saccades, and the start and end of saccades.

2.4 Eye-tracking Technology and Autism

Visual attention has been tracked by using eye gaze patterns to predict where people will look in natural scenes. Researchers have focused on attention to characteristics of the images such as a pixel, object, and semantic levels (e.g. "features that relate to humans: face, emotion, touch"; "objects with implied motion"; "relating to other senses of humans: sound, taste, touch, smell"; "designed to attract attention or for interaction with humans: text, watchability, operability") [87]. For example, in [87], researchers built a dataset of 700 images with eye-tracking data of 15 viewers and

annotation of the image about object and semantic attributes. When researchers evaluated the eye-tracking data, they found for NT viewers that object and semantic information from scenes are the most important aspects and are seen first. For example, in the image with a man and a bird (see Figures 3.1 and 3.2), the man's face and the bird are global features.

The database was then used again to speculate where people with ASD would look in a scene [81]. Eye-tracking data for 20 people with ASD was compared to the eye-tracking data from 19 NT people [81]. A stronger image center bias in people with ASD was a new insight. Specifically people with ASD looked at the center of the image longer–regardless of how many objects were in the scene [81]. Additionally, the researchers found that there was a reduced saliency for faces and locations. They deduced this finding from the social gaze of the people in the images. Overall, they found that the tiny (pixel-sized) but drastic (high-contrast) changes drew the eyes of people with ASD more than did the whole objects or the overall scene [81].

Previous research also investigated the link between the visual behavior of people with ASD and their atypical visual processing of spatial frequencies (i.e., visual characteristics that can be distilled from an image to reveal the varying degrees local to global information) [23]. Additionally, spatial frequency has been explored as a characteristic of stimuli that is processed across both streams: global stream–low frequency, and local steam–high frequency. The experiment in this study involved 30 people with ASD and 30 NT people focusing on gaze shifts that cue the location of targets in different spatial frequencies. The results revealed that people with ASD were biased toward the use of high spatial frequencies (local information).

Even though these existing research projects leveraged eye-tracking metrics (e.g., fixations for [87], fixations for [81]) in studying the differences between people with ASD and NT people, they do not aim to work around local interference for people with

ASD. This dissertation work aims to work around local interference by shifting the eye gaze of people with ASD to the global features in a given scene, by manipulating the stimuli through the use of a global filter.

Chapter 3

Preliminary Work: Designing Low-fidelity Filters

3.1 Methods

To test the hypothesis that filtered images could help assist in shifting eye gaze from local to global features, a variety of filters were created. For the initial study presented first, basic tools such as PowerPoint and PhotoScape X were used to visually highlight the global aspects of 40 images. The images were taken from the open-source dataset previously described [87]. All images can be found at this github repository [2]. Four filters to highlight global features in different ways were designed (for details, see Design of Filters section).

3.1.1 Design of Filters

The filters were designed to highlight global image features by altering high contrast in brightness or color in four different ways. For example, the global information (or main idea) of the image in Figure 3.1 is the man and the bird, which is illustrated by the NT heat map overlaying on corresponding image in Figure 3.2. Each of the filters was therefore used to bring attention to global concepts such as the man and the bird.

Figure 3.1: A sample of image from the experiment in Baseline condition (raw image).

Figure 3.2: This image shows the AOIs by overalying NT heat map on the corresponding image in Figure 3.1. It shows the global features which are the man and the bird.

1. **Baseline (Raw Image):** Ten raw images were used as a baseline condition. The baseline images were not altered (e.g., see Figure 3.1).

2. **Lined Edges Filter:** The Lined Edges filter converts the entire image to a black and white line drawing that emphasizes the shapes of and boundaries between objects. By removing the shading and detail within the objects, the spatial-frequency is reduced, with the intention of transforming the image to be only global features (e.g., see Figure 3.3). This filter is inspired by and similar to icons used in Augmentative and Alternative Communication (AAC) systems. AAC devices provide a means of expressive communication for people with complex communication needs in producing or understanding speech [29]. For example, Proloquo2Go is a popular AAC software that uses line drawing

icons commonly deployed on tablets [4]. Line drawings simplify and promote global processing because all features that would be processed by sensory neurons that process local details such as texture are removed, leaving just the shape of objects. Some global features are removed as well such as color and shading. This filter prototype enables examination of whether removal of several features would support leaner [82], more efficient visual processing of global information. To make images appear as line drawings, the spatial frequency was set at a consistent level across the whole image, rather than varying for contours or high-contrast areas. For this purpose, commercial filters that affected the whole image rather than specific features were used. The intent was to minimize details beyond those that provided an object's edge. Because object edges are perceived automatically to aid shape identification, the hypothesis is that this filter would support object recognition [63].

Figure 3.3: An example of Lined Edges filter.

3. **White Background Filter:** The White Background filter highlights the image's assumed primary object, which is presented with original color and shad-

ing. This object was determined by the researchers. The background is presented as white space. The hypothesis is that a primary object with no background would promote global processing, as the less-relevant local information that could distract the observer has been removed. This approach assumes that primary objects are sufficient for an observer to produce a main idea, gestalt, or holistic comprehension of an image. The idea of removing all the images that are not primary images as determined by the design team creates a condition similar to the "errorless learning" method of ensuring a student gets a response correct by only providing a correct response over the course of prompts or when a task is particularly difficult. Of course, many of the main objects still contain details that could distract participants who display local interference [78]. See Figure 3.4 for an example of White Background filter.

Figure 3.4: An example of White Background filter.

4. **Grey Blurred Filter:** Unlike the White Background filter that whitens the background in an image, the Grey Blurred filter highlights the primary object by presenting it with its original color and shading, against a greyed-out, blurred

background that reveals a hint of context. The hypothesis is that this filter would elicit visual attention to the raw main objects as they are richest areas in terms of intensity while still providing partial information about the context. Also, some of the background would be preserved as the image is altered by removing the color and distorting the pixels of the background, the goal is to remove pixel-level contrasts (e.g., a high contrast in hue intensity where a reflection appeared) in the background that would not contribute to identifying the main object. See Figure 3.5.

Figure 3.5: An example of Grey Blurred filter.

5. **Animation:** The animation consists of a sequential presentation of sequencing of the four filters, using a graphical interchange format (GIF). The GIF cycles through each filter over three seconds. All baseline and filtered images were also presented for three seconds each, as were the images in the eye-gaze studies referenced earlier. The hypothesis is that the animation can simulate the global to local progression. Specifically, by showing the global object with no background first to support errorless learning which is having only the correct choice

available (only the main object(s)), followed by the other filters that added in background and details over time, the global-to-local progression described in the visual processing literature [5] would be created in the environment. The visual processing literature claims that neurons are specialized and geographically localized to process certain features (e.g., movement, color) [59]. In other words, different sensory neurons specialize in detecting edges (visual cortex), color (cones), luminance (rods). As the neuronal activity in the visual cortex disperses through the neural network to cognitive processing, global and local streams of visual information are integrated. In the case of autistic visual attention, the global processing neurons are thought to be intact but overridden by local processing. The goal of the filters in this work is to activate global neurons first (e.g., give global processes a "head start"). In the first filter, the details are removed to attempt to eliminate local interference. Then by adding details back into the image progressively through animation, the filter is integrating details and eventually a complete picture is presented. This preliminary low-fidelity work did not exhaust ways to create a low-fidelity prototype animation, but simply used the filters already created to suffice as the global-to-local progression. See Figure 3.6 for the screenshots of the Animation filter.

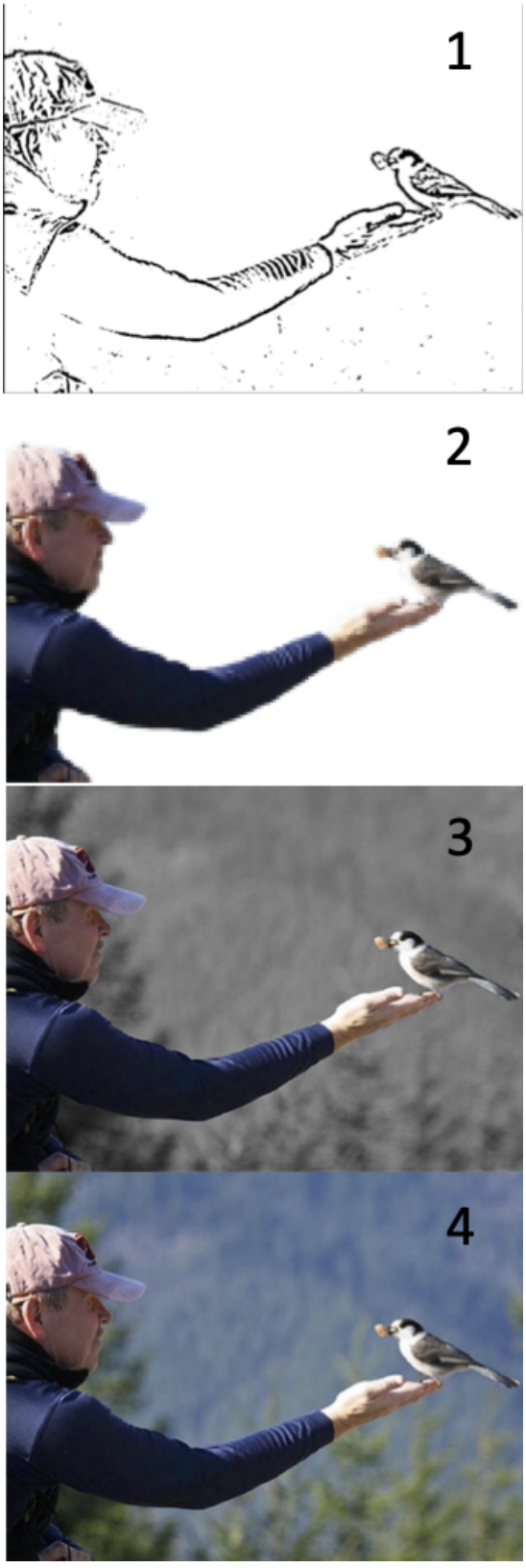


Figure 3.6: An example of Animation filter, starting with Lined Edges, White Background, Grey Blurred, and finally Baseline.

Table 3.1: Navon's test result for P1.

Condition	Reaction time per experimental condition	Error count per experimental condition
Global level	1,656 ms	4 errors
Local level	1,991 ms	1 error
No target at all	2,126 ms	2 errors

3.1.2 Participants

In this preliminary experiment, a 40-year old man with ASD, P1, completed the user test. To provide an objective understanding of visual attention, his eye-tracking data was collected as well as verbal responses to identify the impact of the global filter. As an inclusion criterion, P1 took the Navon's test (www.psytool.org). The test revealed that P1 made 4 errors in identifying global level while he made only 1 error for local level. The detailed Navon's test result for P1 can been found in Table 3.1. As a result, P1 demonstrated local precedence, as there are no cut off scores to verify local interference, it is speculation that he struggles to some degree with global processing. P1 also reported having difficulty with social communication, mainly nonverbal communication. The Navon's test was used to screen the participant for local precedence or interference because researchers in [39, 47, 57, 65, 80] also used Navon-type stimuli to explore atypical visual behaviors (global/local processing) in people with ASD.

3.1.3 Low-fidelity Prototype

The images were taken from the first 50 of 700 images of real-world scenes in the OSIE dataset [87], i.e., images 1-10 are baseline images, images 11-20 are White

Background filters, images 21-30 are Grey Blurred filters, images 31-40 are Animation filters, and images 41-50 are Line Edged filters. Low-fidelity prototypes were created in a PowerPoint slideshow consisting of the 50 images, which were separated by a transition slide. The auto-advance was set for 10 seconds. After each image was presented for 3 seconds, there was a slide with the prompt, "What was the picture about?" followed by a blank screen.

3.1.4 Procedure

Using a chin rest for stability, the lab technician conducted a calibration test using the Eyelink 1000 system. Next, the study began where P1 viewed the low-fidelity prototype that was comprised the 50 images that automatically advanced over the 10 minute session. His eye-tracking data were recorded at a rate of 500 Hz, yielding 500 gaze points per second.

3.2 Evaluating Filters with Eye-tracking Data

The performance of the 4 low-fidelity filters were compared to each other and the baseline images. The aim of using eye-tracking data, which is presumed to be objective data, is to validate the subjective verbal responses. The evaluation was done by using 3 different techniques: visual analysis, fixation analysis, and saccade analysis.

3.2.1 Visual Analysis

A cursory visual analysis of the eye-tracking data is presented here. This preliminary evaluation is based on the 50 pictures containing the heat maps of eye tracking from

an aggregate of 15 NT people from [87] and the eye fixations from the participant with ASD (P1) that take into account the number of eye's fixations that falls in and out of the AOIs.

In this visual analysis, an example of the AOIs can be seen as the yellow regions with a green boundary, shown in Figure 3.7. An example of a fixation can be visually seen as any dense cluster of eye gazes (blue x's), also shown in Figure 3.7. The accuracy score was calculated by counting the overlaps between the AOI and P1's fixations. A penalty score was calculated by counting the number of fixations that falls out of the AOIs. The preliminary proposed formulas for a given set of P1's fixations are shown in formula 3.1 and 3.2. Formula 3.1 calculated accuracy score by counting the overlaps between the AOIs and P1's fixations. The penalty score takes into account the non-overlaps between the AOIs and P1's fixations, and the number of fixations that falls out of the AOIs. The purpose of having these two scoring metrics is to take into consideration both correct and incorrect fixations as each image has a variable amount of hotspots.

$$\textbf{accuracy\_score} = \frac{overlapped\_AOI\_count}{total\_AOI\_count} \tag{3.1}$$

$$\textbf{penalty\_score} = \frac{non\_overlapped\_AOI\_count}{total\_AOI\_count} \times incorrect\_fixation\_count \tag{3.2}$$

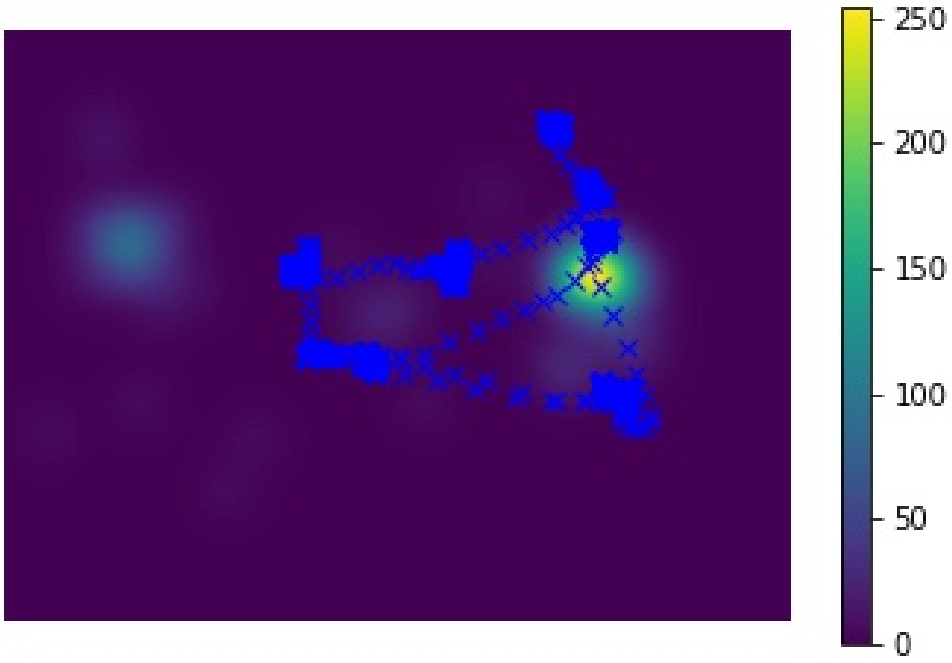


Figure 3.7: A heat map (of Figure 3.1) represents the AOIs (yellow regions with a green boundary) that are aggregated from 15 NT people from [87], overlaid by eye gazes (blue x's) of the man with ASD (P1).

Findings and Limitations

The AOI is defined for this analysis as any yellow regions with a green boundary, which corresponds to where the fixations of NT people are. The fixations of P1 are defined as those blue points that are densely overlaid on top of each other. For example, Figure 3.7 shows the heat map of the picture from Figure 3.1, and there are 2 AOIs and 9 fixations. Two researchers viewed the images independently to determine overlaps or non-overlaps and compared results. Disagreements were discussed until criteria for each image were agreed upon. One of the fixations fell into 1 AOI. As a result, P1's score is 50% accuracy and eight times 50% penalty for this particular

Table 3.2: Performance score of different filters in visual analysis.

Filter	Average Accuracy (%)	Average Penalty (%)
Baseline (Raw Image)	56.6	39.4
Lined Edges	48.3	48.6
White Background	**60.0**	**38.2**
Grey Blurred	47.5	49.5
Animation	48.3	47.4

image. Among the five experimental filters, the White Background filter yielded average result marginally better than the baseline and significantly better than the other filters in term of both accuracy percentage (60%) and also penalty percentage (38.2%) as shown in Table 3.2. This initial analysis was to provide an initial indication of how the filters performed [70].

There are some limitations in the preliminary visual analysis. First, the approach does not use the numerical data which can be more precise when determining a fixation as well as the boundaries for an AOI. Second, equal scores were given for each AOI regardless of their sizes. Third, the number of fixations that falls into the same AOI was not considered. Fourth, the penalty score is not applied when all AOIs are already covered by fixations. That is, no matter how many fixations are outside the AOIs, as long as there are fixations inside all the AOIs, then the penalty score will be zero. These limitations are addressed in the analysis on the numerical data in the next section: 3.2.2–Fixation Analysis.

3.2.2 Fixation Analysis

In this subsection, performance of the filters was measured based on fixations that are computed from raw gaze points of P1, using Cluster Fix algorithm [1] in MATLAB [40]. The number of overlap/non-overlap between P1's fixations and the AOIs was counted.

The AOIs from the NT heat maps have been expanded by increasing the degrees of visual angle to explore if or when there is the most overlap. Results show that there was no specific degree of visual angle that aligns with the AOI in any conditions. Figure 3.8 shows the animation of overlap/non-overlap between P1's fixations and the NT heat maps.

[1] Used with permission.

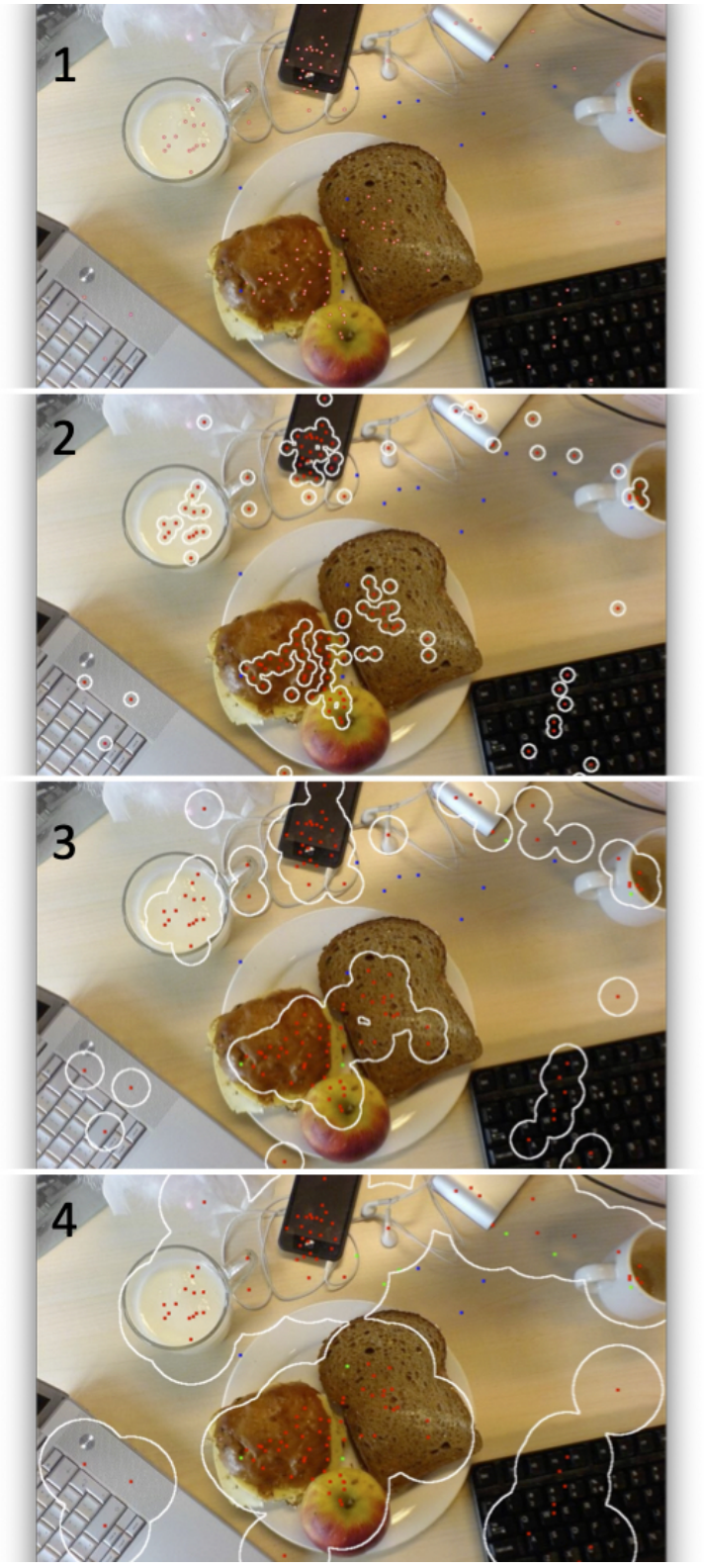


Figure 3.8: An animation of evaluating P1's fixations compared to NT fixations. Red dots represent fixations of NT people. Blue dots represent fixations of the man with ASD, P1. Blue dots turn to green when they are in the areas of the incremental visual angle of NT people.

Even if increasing the visual angles (i.e., the angle a viewed object is displayed on the retina) to the maximum possible angle, there is no significant overlap between P1's fixations (13 fixations) and the AOIs. The purpose of increasing the angle is to consider viewing the image by using peripheral vision. See Table 3.3 for an example of the number of overlaps while increasing the visual angles that corresponds to the Figure 3.8. Even though this result does not provide evidence of effectiveness for the filters, this procedure could be useful in determining emerging shifts in eye gaze in future work as a metric for peripheral vision (i.e., vision beyond the fixation point) that anecdotally seems relevant yet has not been discussed or designed by the research community. This leads us to the next measurement using saccade in the next section: Saccade Analysis.

3.2.3 Saccade Analysis

It is thought that the global processing of visual attention occurs in the first few fixations and saccades [76]. Given that this work is focused on global visual attention, further analysis of the early saccades of P1 was conducted.

A center bias in the gaze occurs as viewers tend to look at the center before a new image appears [81]. To address this bias, the first fixation was not counted but rather the second and third fixations were analyzed for the global aspects. After viewers get the global meaning, then they tend to look at details; thus, the saccade between the second and third fixations, called the "second saccade," is considered to be the representative of global processing. With the existing fixation data, a vector was created to indicate the second saccade (i.e., a segment of scan path with direction; see Figure 3.9).

Table 3.3: Fixation analysis–the number of overlaps between P1's fixations and the areas of increasing visual angles from NT fixations as seen in Figure 3.8.

Degree of visual angle	Overlap count	Degree of visual angle	Overlap count
1	0	21	6
2	0	22	6
3	0	23	6
4	0	24	6
5	0	25	6
6	0	26	7
7	0	27	7
8	0	28	7
9	0	29	7
10	1	30	7
11	2	31	7
12	2	32	7
13	2	33	7
14	2	34	7
15	2	35	7
16	2	36	7
17	3	37	7
18	4	38	7
19	5	39	7
20	5	40	7

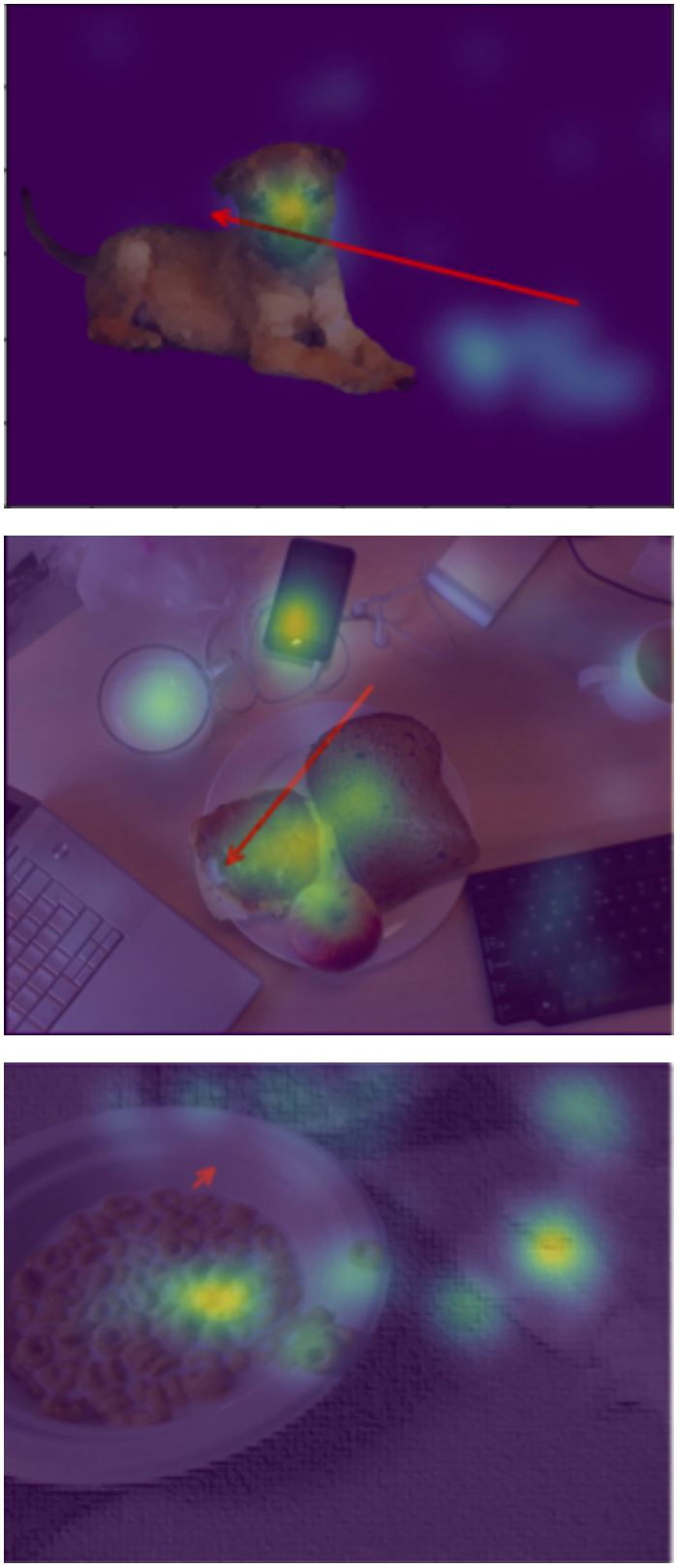


Figure 3.9: A White Background image (top), a raw image (middle), and a Grey Blurred image (bottom), overlapping with the hotspots (in green/yellow) and red arrows showing the second saccade of P1.

A "hit" is when the second saccade of P1 cuts through any hotspots in a given image. Three evaluation metrics were used to identify the filter performance.

1. Comparing the number of overlaps between the vectors of early saccades and hotspots–the overlap demonstrates a shift in attention to global features [44, 83]. The purpose of this method is to figure out if saccade vectors are oriented toward and traveled through the AOIs (hotspots).

2. Classifying the length of the vector below or above the median saccade length, as it has been associated with global processing [84]. The purpose of this method is to figure out if global saccades occur by classifying saccades as long (global) or short (local) [44].

3. Computing the combination of the parameters of length and overlap, which is called "multiplicative parameters."

The second saccade of each image ranged from 23 pixels to 465 pixels with a median of 222 pixels. Each image was classified as 0 if it was smaller than the median saccade length and 1 otherwise. Once each vector was labeled, a χ^2 test was conducted for each image viewed in baseline compared to each filter's 10 images (see example in Figure 3.9). The Animation filter begins with a Lined Edges filter, which resulted in only the presentation of the Lined Edges filter. Therefore, there is no analysis of the Animation filter.

Results

The White Background images demonstrated more overlaps between the second saccades and the hotspots than the baseline images (9 out of 10 vectors overlapping the hotspots compared to 2 out of 10 in baseline; $\chi^2(3, N = 40) = 12, p = 0.007$; Figure

3.9). Saccade vectors appear to be oriented toward as well as traveled through the global areas (hotspots). This pattern occurred remarkably more often with the White Background filter than the baseline condition.

More global saccades (longer) were found in the White Background filter (7 out of 10 images) compared to the baseline (1 out of 10 images), $\chi^2(3, N = 40) = 9.92, p = 0.019$. The Grey Blurred filter also had the same performance as the White Background filter (7 of 10 images had long saccades). These combined results show that when the global objects are emphasized, more global saccades occurred (saccades tend to be longer).

Taking both metrics of saccade length and overlap of hotspots together, the White Background filter continues to perform better than the baseline (7 of 10 images with global saccades and overlaps, compared with 1 out of 10 images), $\chi^2(3, N = 40) = 9.71, p = 0.02$. This analysis revealed that the White Background filter yielded the best results [17], which also conforms to the preliminary visual analysis above [70]. P1 provided more global responses with the White Background filters, thus suggesting that he shifted his eye gaze to important areas of the image. Overall, this preliminary study produced a proof of concept and a rich data set of eye-tracking–thus enabling us to view global processing from different perspectives. The study revealed that the White background filter was the best performer for producing global behavior. However, the White Background filter is not a sufficient filter as the global meaning may require the context of the image. This consideration leads to the iteration, discussed in the next chapter: High-fidelity Filter.

Chapter 4

High-fidelity Filter

In the low-fidelity phase, the filters were manually implemented to highlight global image features by altering high contrast in brightness or color in four different ways. Also, the design team worked closely alongside Speech-Language Pathologists (SLPs) who work with children with ASD having significant speech and language disorders. The same team continued design for the high-fidelity filter using the results of low-fidelity study. This high-fidelity filter aimed to meet the requirements laid out in the low-fidelity prototype, proof of concept study. The result was a filter that was desaturated and blurred. This filter is an automatically rendered filter that is comprised of the following inputs: the original image, the heat map, the amount of blur, and desaturation. The resulting output is the modified image used as the global filter in this high-fidelity phase.

4.1 Design Methods

Heat Maps

In the digital image processing research, algorithms have been developed to differentiate three levels of visual processing in humans [81] which are: semantic (i.e., pertaining to the context), object, and pixel (a very small picture element of color on a screen). The authors found that NT's focused more on the socially-relevant or semantic-level features while participants with ASD focused more on areas of shape and contrast at the pixel level. For this current work, the global filter was created using the NT heat maps to identify AOIs at a semantic level. By doing so, the aim is to direct visual attention to semantic features; thus the NT's heat maps serve as global level, see Figure 4.1. As the point of focus in the visual system can sense approximately 1° of visual angle which equates to an object's height of 0.4 inches when at a distance of 23 inches from the monitor [26, 25], the hotspot is intended to guide the eye to that spot as a focal point and minimize the background.

Blurring

On a per-pixel basis, for every point on the OSIE image, blurring was created using the Gaussian blur algorithm. Specifically, a corresponding point was taken from the image's heat map and its level of luminance was used to blur the image's point. In other words, a darker point on the heat map means that the Gaussian blur will be applied with a higher value up the maximum blur level (which is an arbitrary number). For example, all white is blurred at 0% of maximum blur level, 50% gray is blurred 50%, 75% gray is blurred 75% and black is blurred 100%.

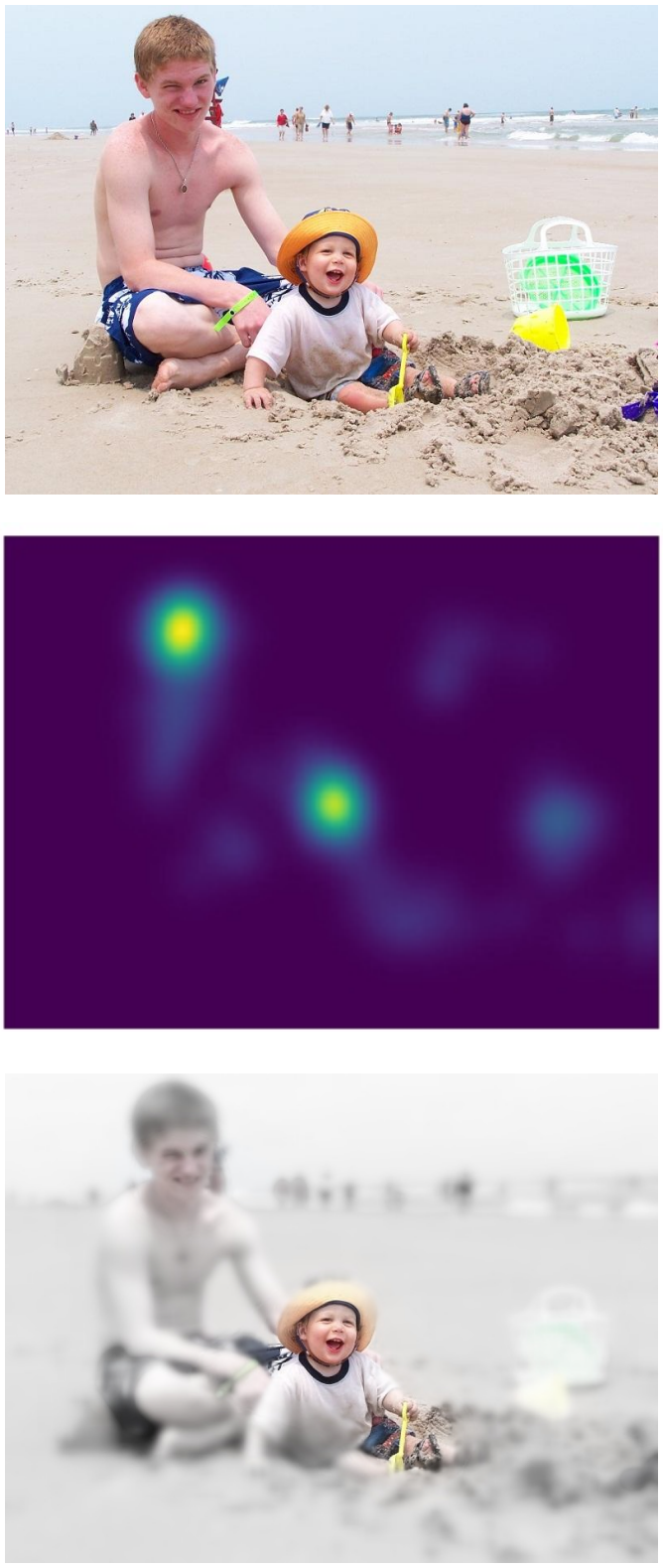


Figure 4.1: Top image is the raw or original version of two boys sitting on the beach. Middle image is the heat map based on eye fixations of NT people. Bottom image is the filtered image that is desaturated and blurred. This technique was applied to the semantic heat map.

Desaturation

For the desaturation process, the same process was employed where every point on the OSIE image, the corresponding point from the image's heat map was referenced. Once referenced, the level of luminance was used to implement the same degree of desaturation. In other words, a dark point on the heat map means that the filter will desaturate more until the image is fully grayscale. For example, all white is the image's original color (saturation), 50% gray is half of the original saturation, and black is fully grayscale.

4.1.1 Providing Filter Options to SLPs

To determine how to filter the images for the high-fidelity study, the findings from the low-fidelity probe were reviewed with the school site team. The design team explained to the SLPs that the white background and blurred gray background were most effective for the group of children at the school as well as for the one adult lab participant. Then, the design team showed them options for the automated filter including a blurred background with mild, moderate, and severe blur, as shown in Figure 4.2. Mild and moderate blurs were selected for further design with desaturated backgrounds, and lastly one filter that was desaturated but not blurred; see Figure 4.3. After brainstorming, the whole group agreed to use the filter that was moderately desaturated and blurred as the intervention.

Figure 4.2: Samples of the automated filters with a blurred background: a) mild blur, b) moderate blur, and c) severe blur. Mild and moderate blurs were chosen for further automation.

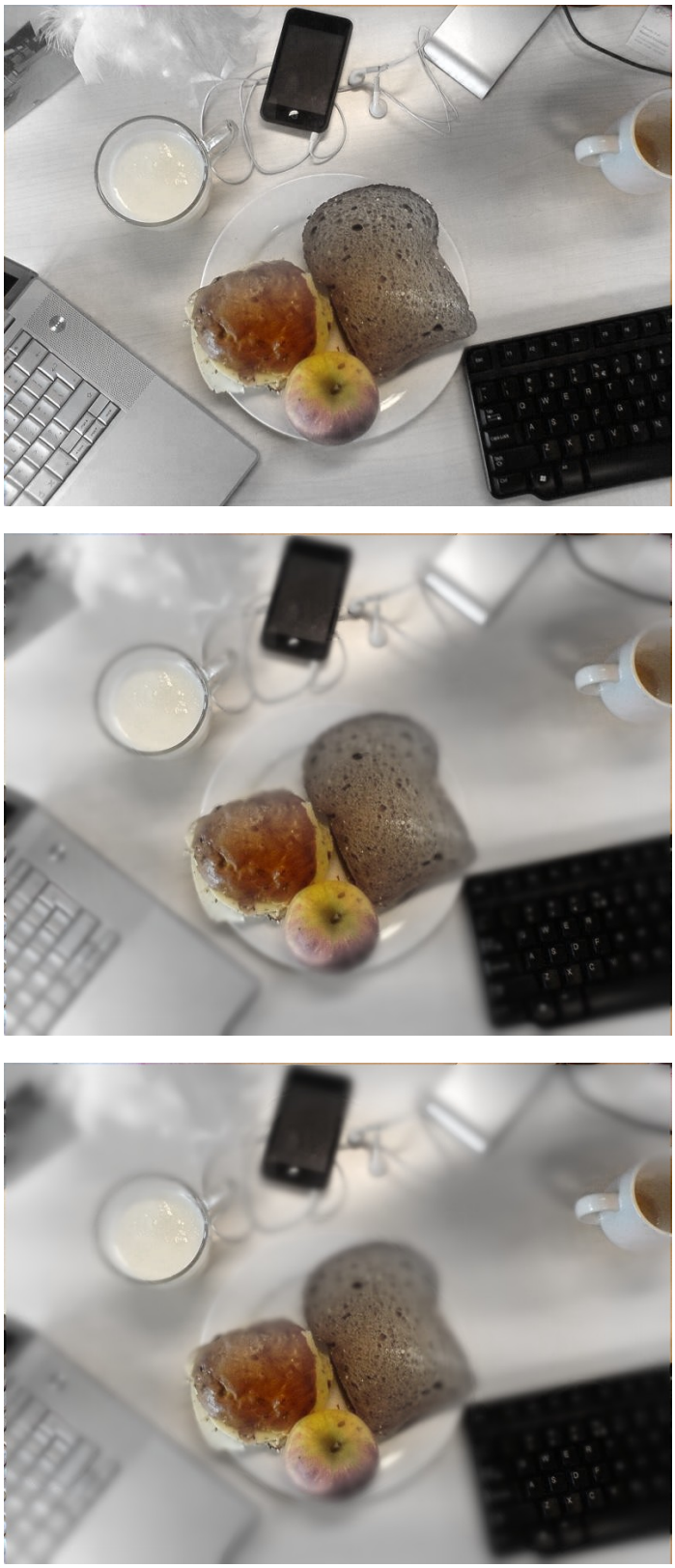


Figure 4.3: Samples of the automated filters: a) desaturated without blur, b) desaturated with mild blur, and c) desaturated with moderate blur. The desaturated with moderate blur filter was selected for the high-fidelity filter experiment.

4.1.2 Participants

Ten male students participated in this study. They each have a diagnosis of ASD and represented a wide range of speech and language issues described in Table 4.1. They were not screened for local precedence or local interference using the Navon's test as the time was not afforded to the research team nor was it clear if the participants could complete the screening. However, understanding if the participants demonstrate local precedence before the onset of the experiment would be helpful in the future. This work was approved by Chapman IRB 19-167.

4.1.3 Study Procedure

Over the course of four days (two rounds of 2-day sessions), this study was conducted at a non-public school that specializes in speech and language disorders. Children who receive highly specialized speech and language services were recruited. All children who attend the school automatically participate in intervention programs that support appropriate language and social interactions. All participant children receive a full range of speech and language, counseling, and behavioral services that focus on everyday uses of language, including listening, understanding, and responding appropriately to others. Parents returned a signed consent form prior to the study, and children were asked to provide assent at the onset of the first session. Two sessions were conducted across consecutive days except for P9 who completed one session in the morning and the other one in the afternoon as he was otherwise offsite for the duration of the study. One assumption made by the team was that children with ASD in a specialized school for speech and language would demonstrate local interference but did not directly screen for this. The team recruited children whom the team believed would tolerate wearing the eye gaze headset and be able to work with their

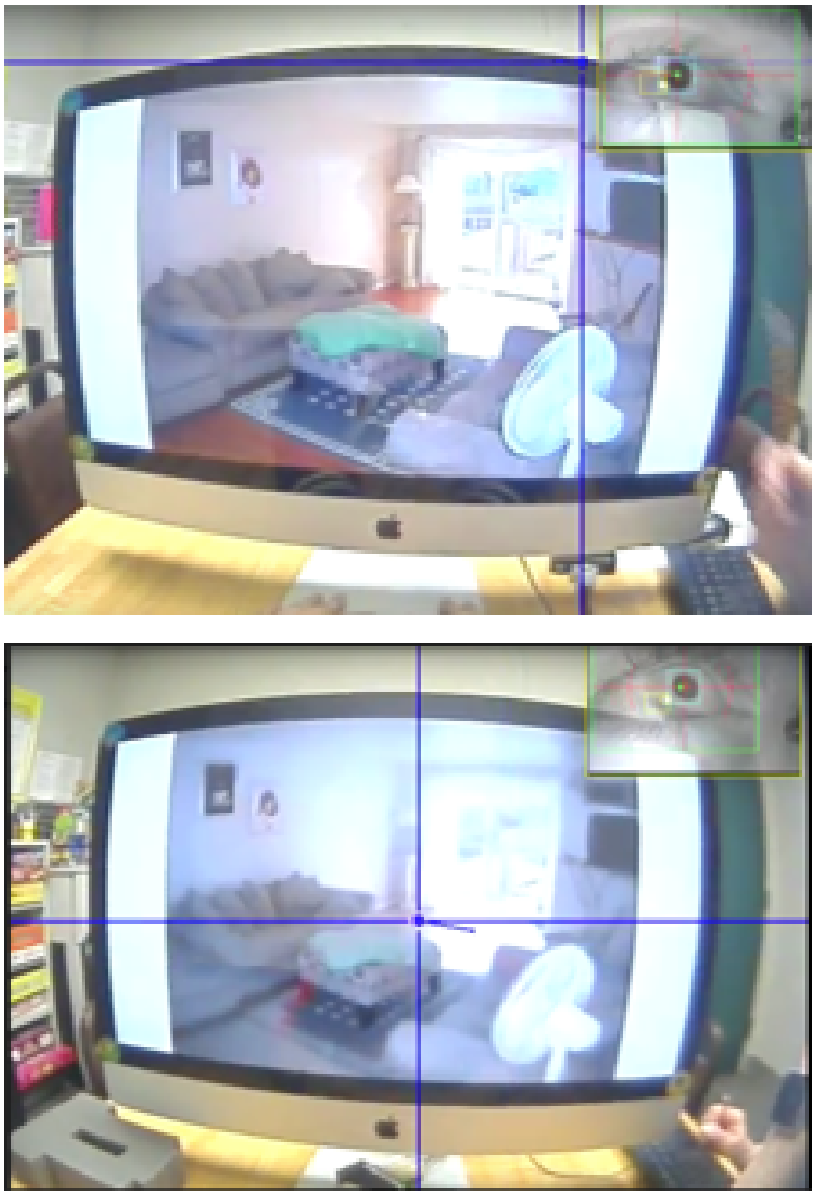


Figure 4.4: Top: Screenshot of P2's eye-tracking video in baseline with cross hairs outside the image at the top left of the screen. Bottom: Screenshot of P2's eye-tracking video in the filtered condition with cross hairs in the center and within the AOIs indicated by the heat map.

Table 4.1: Participant demographics in the high-fidelity filter study as described by their SLPs.

P#	Age	Specific speech related challenges
P1	18	Social referencing difficulties, Perspective-taking difficulties
P2	18	Speech articulation difficulties, Phonological challenges, Grammar and syntax difficulties
P3	12	Delayed receptive and expressive language, Perspective-taking difficulties
P4	16	Cognitive rigidity, Perspective-taking difficulties, Expressive language formulation difficulties, Vocab difficulties: understanding and use
P5	12	Expressive formulation difficulties, Word finding difficulty, Perspective-taking difficulty
P6	12	Delayed response for language, Expressive narrative language difficulty, Language processing difficulties, Perspective-taking difficulties – genuine in what he says but cannot provide rationale
P7	11	Cognitive rigidity, Word finding difficulty, Narrative language difficulty
P8	15	Delayed response time due in part to dysfluency, Expressive semantic challenges
P9	9	Cognitive rigidity, (difficulty with transition – would not put on eye tracker), Early sequencing difficulties, Delayed response for language
P10	19	Delayed response time, Language processing difficulties, Pragmatic difficulties (difficulties with holding conversations, turn taking)

SLPs for two 10-15 minutes test sessions. Another assumption made by the team was that eye gaze, a measure of one's overt attention (observable through eye-tracking technology), reflects the participant's covert attention (unobservable) [44].

The experimental design of the study was a 2 X 2 factorial design (baseline X filter, session one X session two) where 50 images were presented in their original form and the same 50 images in the filtered condition, in a randomized fashion across 2 sessions where the same image was presented in the opposite session in a counterbalanced way to control for order effects. Two session were required to provide time between image presentation for the participants to forget the image as members of the research team were concerned about a confound. Specifically, the concern was that showing the same (or filtered in our case) image close in time would result in familiarity of the image and hence lead to a different gaze path because they have had an opportunity to get familiar with the image. Therefore, each image was only shown once per day to minimize the impact of memory. Verbal responses and video logs of eye gaze behavior were collected.

4.1.4 Study Design

A simple randomization method [6] was employed to determine the order of images, whether it was baseline or filter in the first presentation. The hypothesis is that: 1) eye gaze fixations will shift from locally salient areas (high pixel contrast) to globally salient areas (hotspots in heat maps) in the filtered condition, and that 2) this will lead to a shift in communication via a verbal response to the prompt, "What was the picture about?" to more global responses in the filtered condition as well.

Table 4.2: Verbal responses to the prompt: "What was the picture about?" for the baseline/original picture in Figure 4.4 Top. For this trial, baseline was seen first.

P#	**Verbal response**	**Score**
P1	Living room	2
P2	Fan	1
P3	It's a couch in the living room	2
P4	Living room	2
P5	A door	1
P6	Living room	2
P7	Living room with fan and foot sofas	2
P8	Chairs	1
P9	It's a room	2
P10	Bed is couching	1

4.1.5 Running the Sessions

The treating SLP for each child was the person who conducted the child's session. The SLP sat next to the child, provided the introductory explanation at the start of the session, and any redirection to look at the fixation point throughout the session–a possibly confounding issue to address in future work. Calibration to align aspects of the screen with the eye tracker occurred at the onset of every session (except for P8 who did not choose to wear the eye tracker in session 1). The calibration included gazing at 5 points on a calibration screen and took approximately 2 minutes per participant. Video from a head-mounted device of the user's view was captured. Each session lasted approximately 9 minutes.

Table 4.3: Verbal responses for the filtered picture in Figure 4.4 Bottom.

P#	**Verbal response**	**Score**
P1	Living room	2
P2	Living room	2
P3	It's a sleeping bag in the living room	1
P4	Living room	2
P5	Couch	1
P6	Living room	2
P7	Living room with a fan, a blanket, and a sofa	2
P8	A chair	1
P9	A room	2
P10	Sitting on the couch	1

4.1.6 Study Setup

The room was set up with a table, chairs, and a monitor. Participants sat approximately 23 inches from the 27-inch screen. A set of practice slides were introduced at the beginning of the PowerPoint presentation of images. Additionally, a repeated transition slide was presented between each image. The transition slide only contained a small plus sign in the center to draw the participants gaze to the center before the next image was presented. This procedure is used to attempt to control for the location of the first fixation. The presentation of images auto-advanced so that images were presented for 3 seconds followed by a 7-second repeated textual prompt slide that read "What was this picture about?" See Figure 4.5. The participants were asked to verbalize their response during the 7-second "What was this picture about?" slide. This time frame was determined by the SLP's who worked with the participants on a regular basis and were their instructors during the study. The participants' responses were transcribed by a SLP live and were deemed reliable via review of video recordings by a team member for 20% of trials.

4.2 Data Analyses

An analysis on the verbal responses to the prompt "What was the picture about?" as well as video logs of eye gaze behavior was conducted. To determine a score for the global or local nature of the verbal responses in relation to a specific image, an SLP was hired as an independent contractor to develop a rubric. The first version of scoring rubric for verbal responses was made by an independent SLP with the score range from 0 to 4 at the onset of this study. The rubric was presented to a group of scorers that consisted of 2 senior authors (research faculty), the independent SLP, and three speech-language pathology students. The group met for a total of seven hours

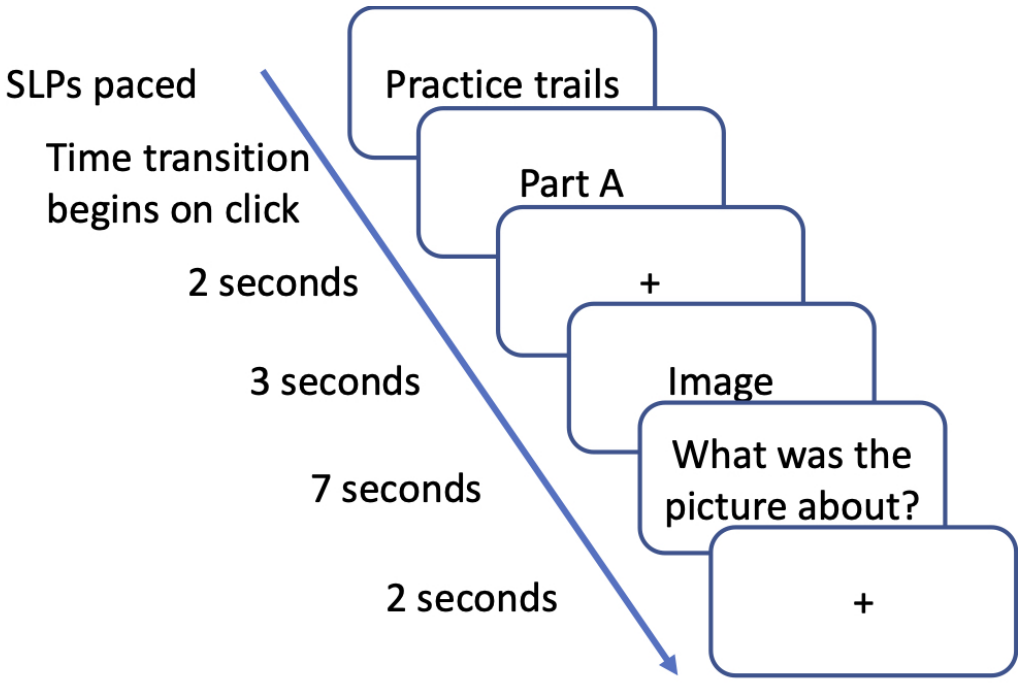


Figure 4.5: Flowchart of global filter experiment paradigm where Part A and B occur in different sittings. The 50 images shown in Part A are the same images as Part B but are counterbalanced to be in the other condition they appear in Part A. The order is randomized for both Part A and B.

over two days to refine the rubric and obtain inter-rater reliability. The resulting rubric was used by the 2 faculty and 2 speech-language pathology graduate students to score verbal responses independently; one outside rater was given the rubric to score independently. However, achieving sufficient reliability data with this rubric proved difficult (i.e., inter-rater reliability of 27%) so the rubric was revised. It was reasoned that the complex scoring requirements, combined with the wide range of responses, had resulted in too many possible interpretations of the data. As a result, a new, simpler scoring rubric with three possible scores was proposed: 0 (incorrect/unrelated responses); 1 (irrelevant or local details); and 2 (plausible global description).

The revised version of rubric was given to two independent speech-language pathology students. They first took the test themselves and their responses were transcribed. Then they scored the study responses. A third speech-language pathology student scored independently to provide inter-rater reliability which was found to be 87%. In every session, responses were presented without indicating who said what and which condition the responses were from. During the processing of the data, from the 550 possible response pairs, 75 were deleted for no responses resulting in a total of 848 (475 pairs) verbal responses for both conditions.

Next, six undergraduate research assistants viewed 848 three-second video clips at 0.25 playback speed that captured the gaze path that was recorded by the eye tracker. They scored each video as "hit" or "miss" based on the eye gaze path passing through any hotspots in a given image, with an inter-observer agreement (IOA) score of 82%.

Table 4.4: Results of verbal responses across conditions and sessions in the high-fidelity study.

Condition	**Session 1 average score**	**Session 2 average score**
Baseline	1.49	1.53
Filtered	1.45	1.36

4.3 Results

4.3.1 Verbal Responses (Subjective Data)

The results for the verbal analysis reveal a significant difference in baseline and filter conditions ($p = 0.001$)–but in an unexpected direction such that participants scored slightly higher in the baseline condition than the filtered, as shown in Table 4.4. Comparing the difference between baseline and filter is greater than the total difference between session 1 to session 2–which indicates an effect of the filter in the wrong direction.

An example of the types of utterances recorded shows most children used similar if not exactly the same words in each condition, (see Tables 4.2 and 4.3).

4.3.2 Eye Gaze Behaviors (Presumed Objective Data)

No significant difference (p=.07) was found in the "paired t-test comparing the scores of the baseline hits to the filtered hits. Further investigation revealed that for 41% of the images, participants did not look at the hotspots in either condition; whereas, they looked towards at least one hotspot in a given image for both conditions 6% of the time. Impressively, for 25% of the total images, their gaze passed through the

hotspots in the filtered condition. Only 28% of the baseline images resulted in hits to the hotspots, so it is feasible that the filter potentially worked in some cases" [71].

4.4 Discussion

Because participants were asked to verbalize "What was the picture about?" they were therefore prompted to employ global processing. This may be why the results showed high rates of global responding in baseline as well as in the filtered condition. Additionally, only the overall gaze path was analyzed, it was not broken down into local and global paths as described by [44], wherein the authors say that local and global gaze paths are represented by different eye-tracking behavior such as fixations and saccades respectively. This provides an additional potential way to analyze eye-tracking data when saccade data is available as done in section 3.2.3.

Instead of a screen-based eye tracker, a head-mounted eye-tracking device (i.e., glasses shown in Figure 2.2) was used in the high-fidelity filter study. This device was used because it was a portable alternative to conduct experiment offsite (i.e., outside of laboratory). Additionally, it is difficult to arrange for the participants with ASD just to put on the glasses, let alone having them use a chin rest. Due to the inherent limitations of this portable device, fixations and saccades are not available; only the video logs were obtained.

The results from this high-fidelity filter study do not support the hypothesis that manipulating the stimuli will guide visual attention to the global features. Given the positive results from the low-fidelity study and the potential negative impact of requiring verbal behavior along with eye-tracking behavior, further investigation of the stimuli was warranted. Some sensory aspects of visual processing and hence visual

attention were considered and further analyses on the characteristics of images in the experimental pool were conducted, specifically for luminance, chroma, and spatial frequency (see Chapter 5).

Chapter 5

Analysis on Characteristics of Experimental Images

The characteristics of images that are discussed here include: luminance, chroma, and spatial frequency. Luminance, chroma, and spatial frequency are attributes that are sensed early in human visual-processing. Therefore, these attributes are parts of the early global impression of an image, hence, areas to target for an intervention aimed to highlight global processing. Specifically, each characteristic could play an important role in forming a global filter because of their effects: the intensity of light, color, and the distribution of content inside the image. First, how to quantify each characteristic for a given image is discussed. Then, each feature is consolidated into a regression analysis to study the effects and correlation between each characteristic and the performance score (i.e., hit count). The 50 baseline/raw images from OSIE dataset [87] and the corresponding 50 filtered images can be found in appendix Supplemental Materials A.1–Baseline Images and A.2–High-fidelity Filtered Images, respectively. The sorted order of baseline images and high-fidelity filters from highest to lowest hit count can be found in appendix Supplemental Materials A.3–Baseline and High-

fidelity Filtered Images Sorted Based on Highest to Lowest Hit Count.

5.1 Luminance

One of the image characteristics used in the high-fidelity filter is the intensity of light, also known as luminance. The filter desaturates the images so that the brightness of non-global features are lowered down because researchers found people with ASD tend to focus on bright contrast at the pixel level in a given image [81]. Therefore, by dimming the areas of non-relevant content, the focus should be diverted to the targeted areas that maintain their original luminance.

To quantify the luminance of a given image, the images were first converted from RGB (Red, Green, Blue) to HLS (Hue, Lightness, Saturation) by utilizing a python function called *cvtColor(image, cv2.COLOR_RGB2HLS)* from a python library called OpenCV [46]. This step focused on the lightness so the lightness value was extracted from the HLS. Sample code can be found in the appendix Sample Code B.1–Luminance.

Among the experimental image pool, the top 3 images with the highest average luminance are shown in Figure 5.1, and the top 3 lowest average luminance are shown in Figure 5.2. Overall, each filtered image has a level of luminance higher than its corresponding baseline image, which is expected as the high-fidelity filter grey out most areas on the image, making them close to white color which corresponds to a high degree of lightness. The luminance frequency histogram for each image can be found in the appendix Supplemental Materials A.4–Luminance Frequency Histogram.

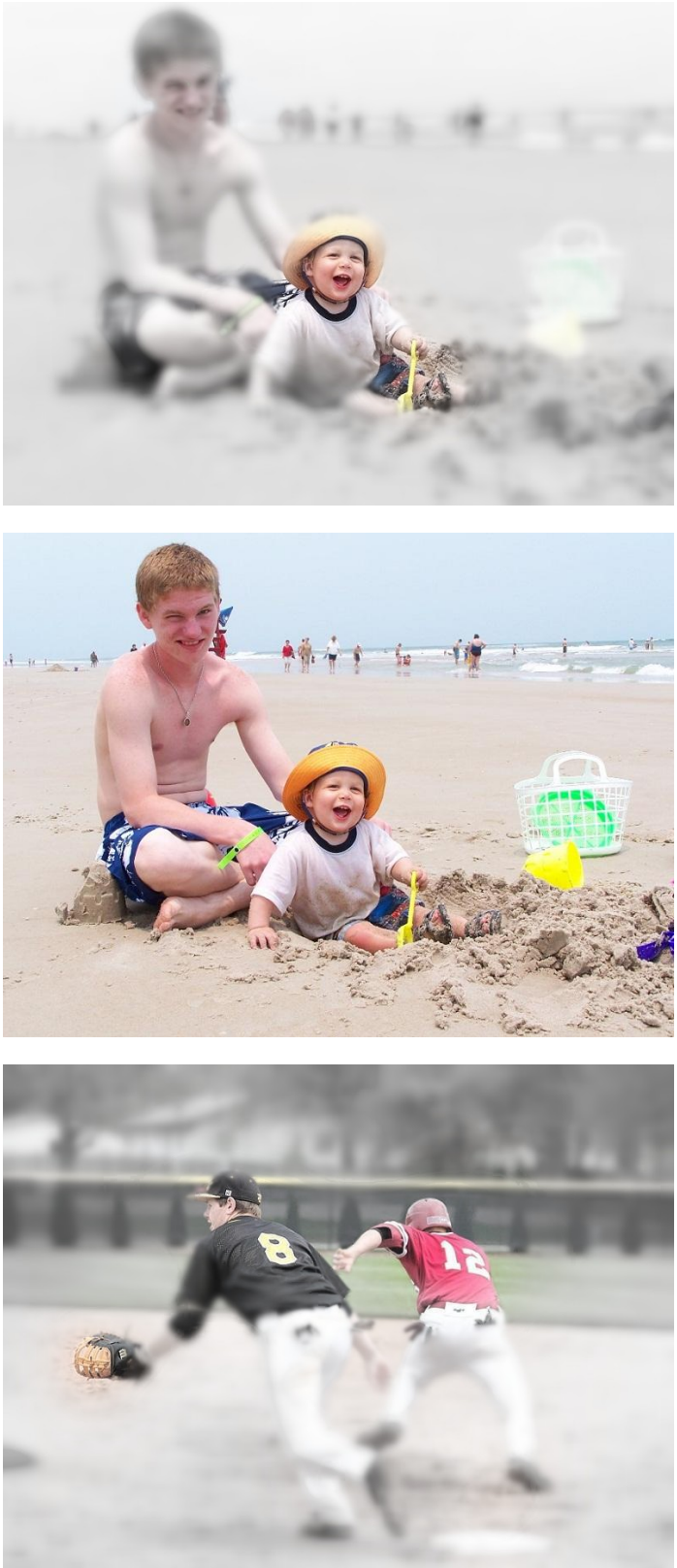


Figure 5.1: The top 3 highest average luminance images: a) filtered image of two boys sitting on the beach, b) baseline image of the two boys sitting on the beach, and c) filtered image of two men playing baseball in the field. These images also have the highest average chroma value.

Figure 5.2: The top 3 lowest average luminance images: a) baseline image of a dog sitting at a table, b) baseline image of two people walking on the beach with two sailboats, and c) baseline image of a puppy with his toys. These images also have the lowest average chroma value.

5.2 Chroma

Another characteristic of an image is color, also known as chroma. For the experiment, colorful images which contain three channels were used : blue, green, and red. To extract chroma value of each channel for each image, a python function called *imread(image)* from the same OpenCV library as for luminance was used. Sample code can be found in the appendix Sample Code B.2–Chroma.

Among the experimental image pool, the top 3 images with the highest average chroma values are the same as the images with the highest average luminance, shown in Figure 5.1, and the top 3 images with the lowest average chroma are also the same as the images with the lowest average luminance, shown in Figure 5.2. Luminance and chroma are highly positively correlated (correlation coefficient=0.99, shown in Table 5.1). However, the other images do not follow this same trend. Besides, because the images contain 3 channels, extra analysis was conducted, based on separate channel: 1) blue channel, 2) green channel, and 3) red channel. However, no significant findings were noted to indicate a contributing role of one color over another, that was related to the hit rate of an image. Image histogram for each image can be found in the appendix Supplemental Materials A.5–Image Histogram and A.6–Image Histogram in Separate Channels: Blue, Green, Red.

5.3 Spatial Frequency

For a given image, the overall activity level is measured by the spatial frequency of the image [27]. Spatial frequency describes the periodic distributions of light and dark in an image [49]. High spatial frequency refers to features such as sharp edges and fine details, whereas low spatial frequency refers to features such as global shape.

See Figures 5.3 and 5.4 for simple examples of a high spatial frequency image and a low spatial frequency image.

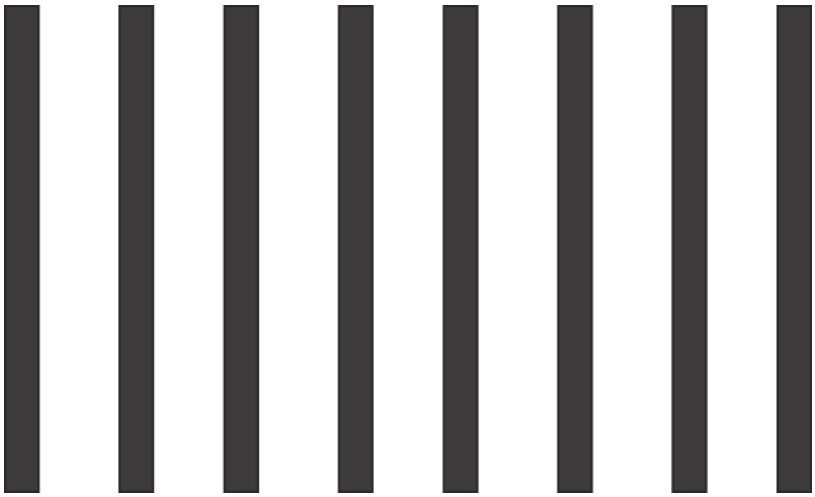


Figure 5.3: An example of low spatial frequency image with five exact same bars in horizontal space.

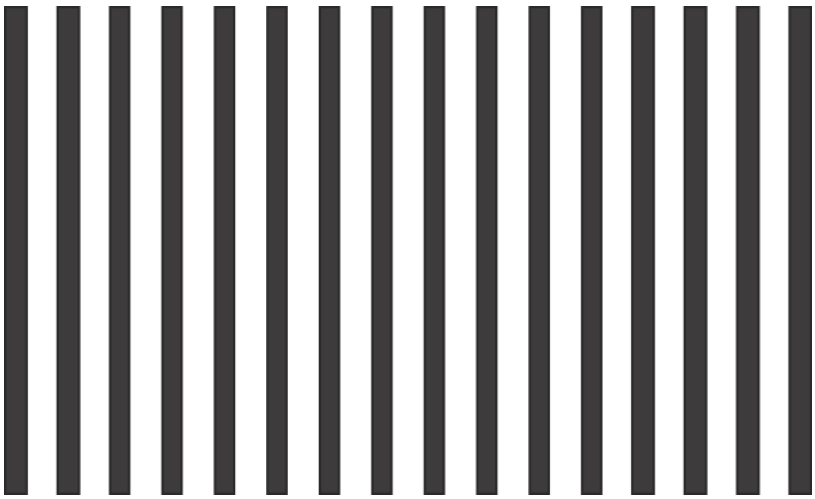


Figure 5.4: An example of high spatial frequency. The image contains twice as many bars as in the low spatial frequency image in Figure 5.3.

To quantify spatial frequency in each image, the formula from [27, 43] was employed. Spatial frequency is calculated by the following equations. For a given image F (M x N) with F(m,n) value at the position (m,n), the spatial frequency is define as

$$SF = \sqrt{(RF)^2 + (CF)^2}, \tag{5.1}$$

where RF is the row spatial frequency

$$RF = \sqrt{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=2}^{N}[F(m,n) - F(m,n-1)]^2} \tag{5.2}$$

and CF is the column spatial frequency

$$CF = \sqrt{\frac{1}{MN}\sum_{n=1}^{N}\sum_{m=2}^{M}[F(m,n) - F(m-1,n)]^2} \tag{5.3}$$

Among the experimental image pool, the top 3 images with the highest mean spatial frequency are shown in Figure 5.5; these images are in baseline condition. The top 3 images with lowest mean spatial frequency are shown in Figure 5.6; as expected, these images are in the filtered condition using the high-fidelity filter.

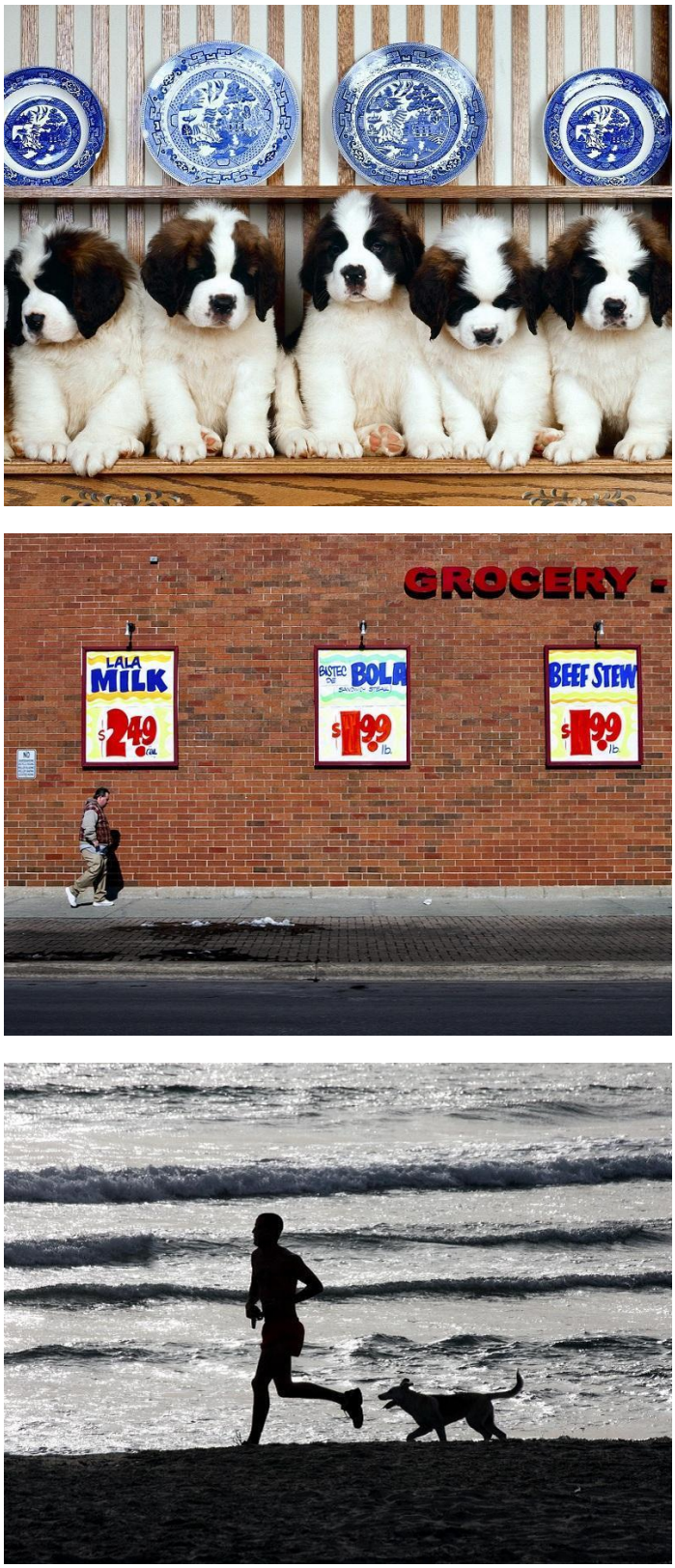


Figure 5.5: The top 3 highest mean spatial-frequency images: a) five puppies with vertical stripes in the background, b) a man is walking by a brick wall of a grocery store, and c) a man and a dog are running on the beach. These images are all baseline images.

Figure 5.6: The top 3 lowest mean spatial-frequency images: a) a bathroom with a toilet, b) a boy with a goat , and c) a puppy standing on a barrel. These images are all filtered images.

Overall, the spatial frequency in filtered images are considerably lower than the baseline images. Because the experimental images are in RGB, the spatial frequency in separate channels were also calculated in: 1) blue channel, 2) green channel, and 3) red channel. The result shows the same trend such that each filtered image has lower spatial frequency in terms of any channel, compared to its corresponding baseline image. Table of mean spatial frequency for each image can be found in the appendix Supplemental Materials A.7–Spatial Frequency. Sample code for calculating the spatial frequency can be found in the appendix Sample Code B.3–Spatial Frequency.

5.4 Regression Analysis

After quantifying the characteristics of each image, characteristics were assessed to see if they affected the number of hits (participants' gaze path cut through the hotspots). To conduct this analysis, a multiple linear regression model with the variables of luminance, chroma, and spatial frequency was built.

The equation of multiple linear regression is defined as

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{5.4}$$

where $\hat{y}$ is the dependent variable, $x_1, x_2, ..., x_n$ are independent variables with their corresponding coefficient $\beta_1, \beta_2, ..., \beta_n$, and β_0 is the constant or interception.

For the analysis, hit count is the $\hat{y}$, and the independent variables $x_i, i \in [1, 2, 3, ..., n]$, are mean luminance and mean spatial frequency. Mean chroma is dropped out of this analysis because mean luminance and mean chroma are highly positively correlated as shown in Table 5.1; i.e., including both variables is not necessary. Following the

equation (5.4), the multiple linear regression is defined as

$$hit\_count = \beta_0 + \beta_1 mean\_luminance + \beta_2 mean\_spatial\_frequency \tag{5.5}$$

Because each characteristic of the images represents different ranges of values, they were standardized before running the regression analysis in equation 5.5. Standardization is a data pre-processing technique used to transform data to standard normally distributed data, (i.e., zero mean and a standard deviation of one (unit variance)). If any characteristic of the images has a variance that is substantially larger than the others, it could make the feature unable to learn from the others correctly when performing regression analysis.

To see the correlation between each variable and the hit count, a correlation matrix needs to be computed. None of the image characteristics are highly correlated with the hit count. But, spatial frequency has the highest correlation with the hit count, compared to luminance and chroma (see Table 5.1 for the correlation matrix). Sample code for performing regression analysis can be found in appendix Sample Code B.4– Regression Analysis.

Table 5.1: The correlation matrix between hit count and the characteristics of the images: luminance, chroma, and spatial frequency.

	Hit count	Luminance	Chroma (RGB)	Spatial frequency
Hit count	1.000000	0.043295	0.039833	0.173309
Luminance	0.043295	1.000000	0.999053	-0.154905
Chroma (RGB)	0.039833	0.999053	1.000000	-0.157920
Spatial frequency	0.173309	-0.154905	-0.157920	1.000000

Table 5.2: The summary result of multiple linear regression with hit count as the output variable, and the estimator variables are the standardized values of luminance and spatial frequency (R-squared=0.035).

	Coefficients	Standard error	p-values	95% CI
constant	5.2347	0.175	0.000	[4.887, 5.582]
mean_luminance	0.1248	0.177	0.483	[-0.227, 0.477]
mean_spatial_frequency	0.3203	0.177	0.074	[-0.031, 0.672]

Table 5.3: The performance comparison between multiple linear regression and Poisson regression. Lower AIC and BIC values are preferred.

Model	AIC	BIC
Multiple linear regression	388.81	396.56
Poisson regression	402.98	410.73

Table 5.2 shows the result of the multiple linear regression. For one unit change (standard deviation) in mean luminance and mean spatial frequency, the hit count is changed by 0.12 and 0.32 respectively. We can see that none of the p-values of each variable shows significance; i.e., each characteristic of the image is not significantly predictive of hit count. However, spatial frequency is marginally significant (p=0.074). Specifically, higher spatial frequency is more likely to have a hit. Also, the upper and lower confidence limits (95% confidence interval) cover a wide range of values (and include 0 for each variable), which is why their effects are not significant. Note: A Poisson regression analysis was also conducted because the hit count is discrete data. However, the linear regression outperforms the Poisson regression based on the metrics AIC and BIC, see Table 5.3. A model with a lower AIC and BIC is preferred because then the model is likely to be closest to the true pattern of the data.

5.5 Machine Learning/Deep Learning Analysis

In regression analysis, the 3 characteristics (i.e., luminance, chroma, and spatial frequency) are extracted from individual image as variables to explore the relationship between the images and the hit count. However, none is found to have significant relationship. As a result, additional analysis using modern algorithms is conducted.

Machine Learning (ML) has become popular at the present time because it has proven to outperform conventional methods like regression model because of the ability to learn non-linear/complex relationships (which cannot be discerned by linear separation approaches) generally occurring in real-world problems. A sub-field of ML is Artificial Neural Network (ANN). ANN has a sub-field called Deep Learning (DL) that has grabbed attention from almost every field with remarkable success in both academic and industry sectors even though DL/ANN is generally considered as a blackbox approach in terms of interpretability. To explore if a given image is predictive of the hit count, two DL analyses are used, particularly Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN).

Before moving onto MLP and CNN, it is important to understand the basic component of ANNs. ANNs are biologically inspired by the brain neural architecture [32]; as a result, the basic component of ANN is also named a neuron, as the one in the brain. A neuron, also known as a node, receives inputs and fires an output. Conceptually, a neuron is like a placeholder with a mathematical function, used to apply on the provided inputs for generating an output, see Figure 5.7. The function used in a neuron is generally called an activation function. There are a number of activation functions and each of them is used for different purposes. The activation function that is used in the following analyses is called Rectified Linear Unit (ReLU) [50]. ReLU is a non-linear activation function and currently the most widely-used in

DL models [58]. See Equation 5.6 for the ReLU activation function; if the input x is positive, the function returns that value back, otherwise, it returns 0.

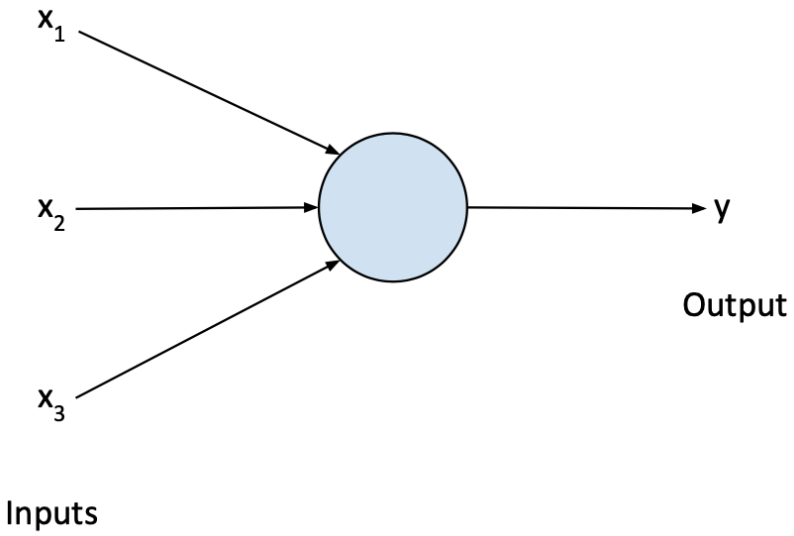


Figure 5.7: An example of a neuron of ANN. Variables x_1, x_2, and x_3 are the inputs into the neuron; y is the output from the neuron.

$$f(x) = max(0, x) \tag{5.6}$$

A more thorough introduction of DL can be found in a popular textbook called "Deep Learning" [32] by Goodfellow et al. Both MLP and CNN analyses are implemented in Python with Keras library. An easy-to-follow tutorial can be found at [67].

5.5.1 Multi-Layer Perceptron (MLP)

MLP Architecture

Intuitively, an MLP is "a mathematical function mapping some set of input values to output values" [32]. MLP is a typical feed-forward ANN because there are no connections in which the outputs from the MLP model are fed back into itself. MLP comprises at least three layers of neurons: an input layer, a hidden layer and an output layer. A layer is a group of neurons that take in inputs and provide outputs. The neurons apply the activation function assigned to them on the inputs to produce the outputs. Figure 5.8 shows an example of the MLP that is used in this analysis. The MLP architecture consists of 4 layers: a) the input layer comprises the three characteristics/variables of each image (using all three variables instead of two variables yields better result here), b) the first hidden layer consists of 8 neurons and each neuron employs ReLU activation function, c) the second hidden layer consists of 4 neurons and also use ReLU activation functions, d) the output layer has one neuron which employs linear activation function (i.e., a simple one-degree polynomial) in order to output a number for the hit count. The number of hidden layers and nodes in each layer are randomly chosen.

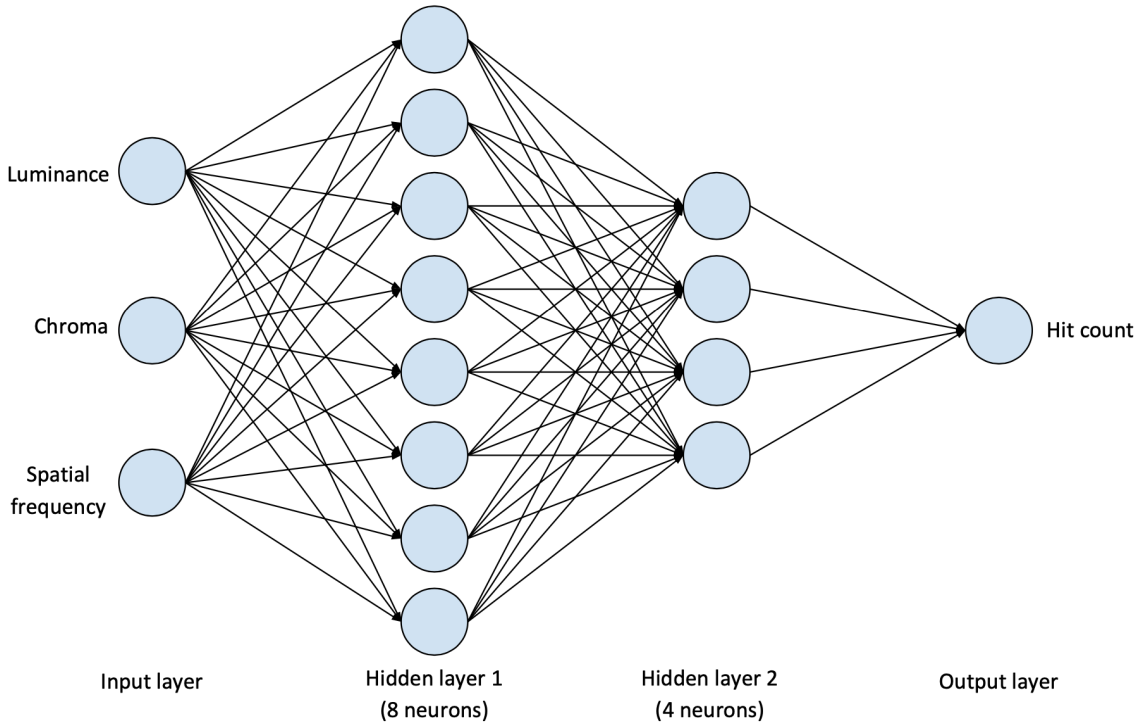


Figure 5.8: The architecture of MLP used in this analysis. It consists of input layer, 2 hidden layers, and output layer. The input layer takes in inputs: luminance, chroma, and spatial frequency. The first hidden layer consists of 8 neurons and the second hidden layer consists of 4 neurons. The output layer takes in outputs from the previous layer (i.e., second hidden layer) and outputs the final result which is the hit count.

It would be good practice to apply a regularization technique like "dropout," a technique in which some randomly selected neurons are ignored during training to avoid overfitting [75]. However, dropout is not employed in this MLP model, considering the size of the current available dataset. Instead, a light-weight MLP (i.e., 2 hidden layers) is employed in this analysis to avoid overfitting. An overfitting model (e.g., MLP in this section) is a model that fits the quirks and/or random noises in a given sample of a dataset rather than reflecting the overall population. In other words, overfitting can be seen when a model performs well with a training dataset but poorly with a testing dataset. The training dataset is the data that is used to build/train a model. The testing dataset is the unseen dataset that is used to test

the performance of a model.

There are a few good practices for training a model. First, both input variables (i.e., luminance, chroma, and spatial frequency) and output variable (i.e., hit count) are scaled to range [0, 1] before training the MLP model. Scaling these variables allows the MLP model to more easily train and converge. Second, an important objective in training a model is to minimize the error, i.e., the difference between the predicted value and the actual value. So, an objective function, also known as loss function/cost function, has to be employed in training a model. Mean Absolute Percentage Error (MAPE) is utilized as the loss function in the MLP model; i.e., MAPE minimizes the mean percentage difference between the predicted hit count and the actual hit count. Third, minimizing the loss function is completed by updating the parameters (i.e., weights of each input) of the MLP model. To accomplish this, an optimizer is needed to decide how to update the parameters, i.e., by how much, and when. Adam [37] is currently the most popular optimizer because of its superior performance compared to others. It is employed in training the MLP model for this work. Last but not least, it is important to limit the number of iterations through the entire training dataset for the MLP model because otherwise it can either take an unnecessarily long time to train the model or affect the performance of the model. A common practice to identify the number of epochs is by plotting the number of epochs along the x-axis and the error (i.e., loss function) of the model on the y-axis. This plot, also known as learning curves, can help to identify the number of epochs.

Methods

The dataset is divided into 75% train set and 25% test set. Two MLP models are built on all 3 variables (i.e., luminance, chroma, and spatial frequency) and 2 variables (luminance and spatial frequency) respectively; each MLP model is trained for 200

epochs. MAPE is a metric used for the loss function.

Results

The average hit count is 5.23 and the standard deviation is 1.75. The three-variable MLP model results in a mean and a standard deviation of absolute percentage difference between predicted and actual hit count of 32.89% and 23.03% respectively. The two-variable MLP model results in a mean and a standard deviation of absolute percentage difference between predicted and an actual hit count of 31.31% and 22.02% respectively. So, there is a slight difference between the two models and the simpler one (i.e., with 2 variables) should be the chosen model. Also, Figure 5.9 shows the two-variable MLP model learning curves of training and testing loss versus the number of epochs. Both loss curves start off by dropping over the number of epochs significantly; it can be seen that between 40 to 50 epochs are likely to be enough number of epochs for training the two-variable MLP model.

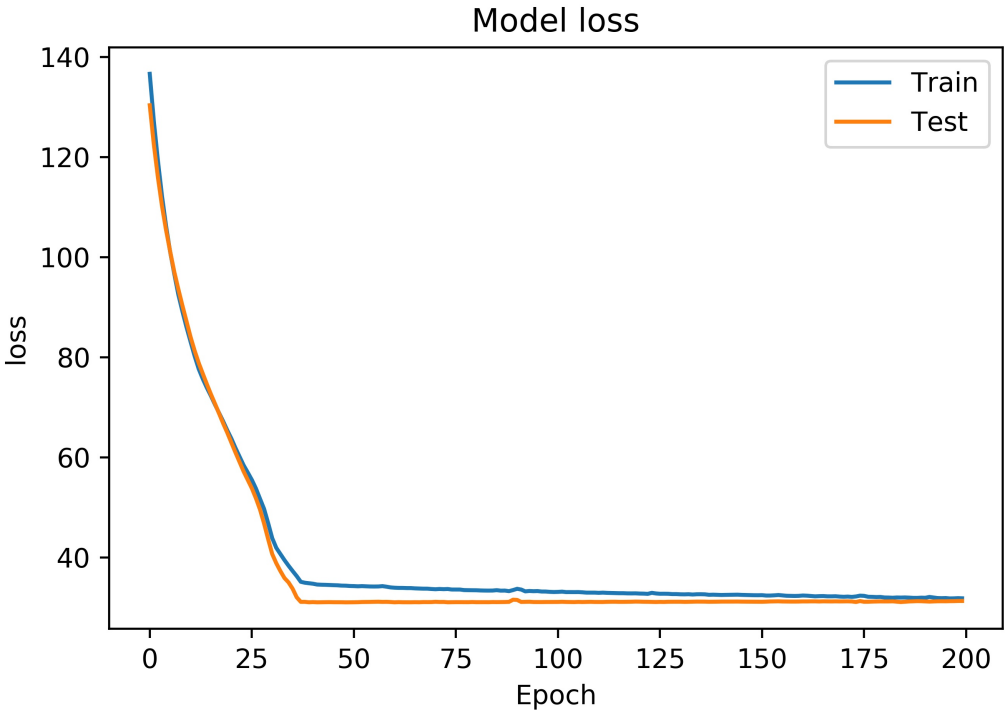


Figure 5.9: The learning curves of training and testing loss versus the number of epochs for the two-variable MLP model. The x-axis represents the number of epochs, and the y-axis represents the loss (i.e., MAPE).

Overall, the MLP model obtains a mean absolute percentage error of 31.31%, implying that, on average, the hit count predictions will be off by 31.31%. Remember that the dataset that is used in training the MLP models are originally extracted from images. Thus, the next analysis is to leverage the raw/original dataset (i.e., images) to see if there is any improvement over the MLP models.

5.5.2 Convolutional Neural Network (CNN)

The original format of the data, which is in the form of an image, contains rich information that requires sophisticated algorithms for revealing the hidden insight. To address this problem, further investigation was conducted using a DL model,

particularly, CNN.

CNN Architecture

CNN is currently the most popular class of DL that is used heavily in Computer Vision, a scientific field that seeks to understand digital images/videos. The name CNN is derived from the type of hidden layers that are used in the architecture. The hidden layers can contain multiple types of convolutional layers, normalization layers, pooling layers, and fully connected layers. Simply put, instead of using the normal activation functions described above in section 5.5, convolution and pooling functions, for example, are employed as activation functions. In a convolutional layer, convolution takes in two inputs (i.e., an image and a filter, also known as a kernel, for the input image), and outputs a third image which is the result of applying the filter on the input image, specifically multiplies the input image with the filter to get the modified image. A normalization layer is utilized to normalize the input (i.e., the mean close to 0 and the standard deviation close to 1) to reduce the training time. A pooling layer is employed to reduce dimensionality of the input, by applying pooling functions such as max-pooling (i.e., selecting the maximum value in a filter region), or average pooling (i.e., selecting the average value in a filter region). The fully connected layers are basically like MLP layers. Figure 5.10 shows an example of CNN architecture that is used in this analysis. Note that the normalization layers and pooling layers are not shown in Figure 5.10 due to limited space. For the detailed architecture of CNN that is implemented, see appendix Supplemental Materials A.8–CNN Architecture. Because this CNN is a more complex model (i.e., consists of more input variables from the images), a regularization of 50% rate dropout is put into place in the first fully connected layer (see Figure A.1 in appendix Supplemental Materials A.8–CNN Architecture).

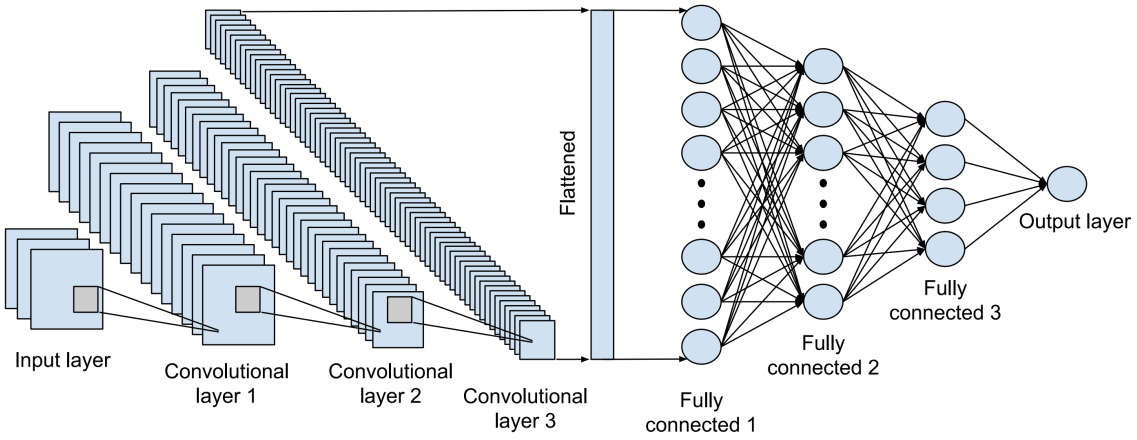


Figure 5.10: The architecture of CNN used in this analysis. It consists of an input layer, three convolutional layers, three fully connected layers, and an output layer. Note that the normalization layers and pooling layers are not shown due to limited space.

Methods

Like the MLP analysis, the dataset (i.e., images for CNN analysis) is also divided into 75% train set and 25% test set. Two analyses are conducted: a) individual image, and b) combining the corresponding baseline and filtered images into a single image. The images were combined to determine if the added features would yield a better result.

Results

For individual-image model, the mean and standard deviation of absolute percentage difference between prediction and hit count are 31.64% and 32.52% respectively. For the combined-images model, the mean and standard deviation of absolute percentage difference between prediction and hit count are 22.22% and 10.14% respectively. So, the combined-images model yields better results. Table 5.4 shows the results of the

two analyses. Note: for combined-images analysis, the dataset decreases compared to individual image analysis, from 98 images to 48 images. Also, from the learning curves of the combined-images CNN model in Figure 5.11, it can be seen that 150 epochs are likely to be enough for training because the test loss stays similar (i.e., no improvement) throughout 200 epochs. As expected, the test loss curve starts off with a smaller dip because each combined image share a majority of desaturated and blurred areas. Then the curve increases significantly when the CNN model really starts to learn the differences between images and drops to a lower level after it accumulates the learning of differences.

Table 5.4: The performance comparison between individual-image CNN model and combined-images CNN model in terms of the mean and standard deviation (std.) of absolute percentage difference between predicted and actual hit count.

Model	Mean difference	Std. difference
Individual-image CNN	31.64%	32.52%
Combined-images CNN	22.22%	10.14%

It would be a good idea to explore the analysis using a combination of all the available datasets that are not redundant. However, because the three characteristics are extracted from images, analysis on a combination of these variables and the original images would be redundant and unnecessary.

5.6 Summary

From the regression analysis, none of the image characteristics shows any significant effect on the hit count. Also, for additional exploration on seeing if the hit count can be predicted for a given image, ML approaches were employed. To do so, we chose DL

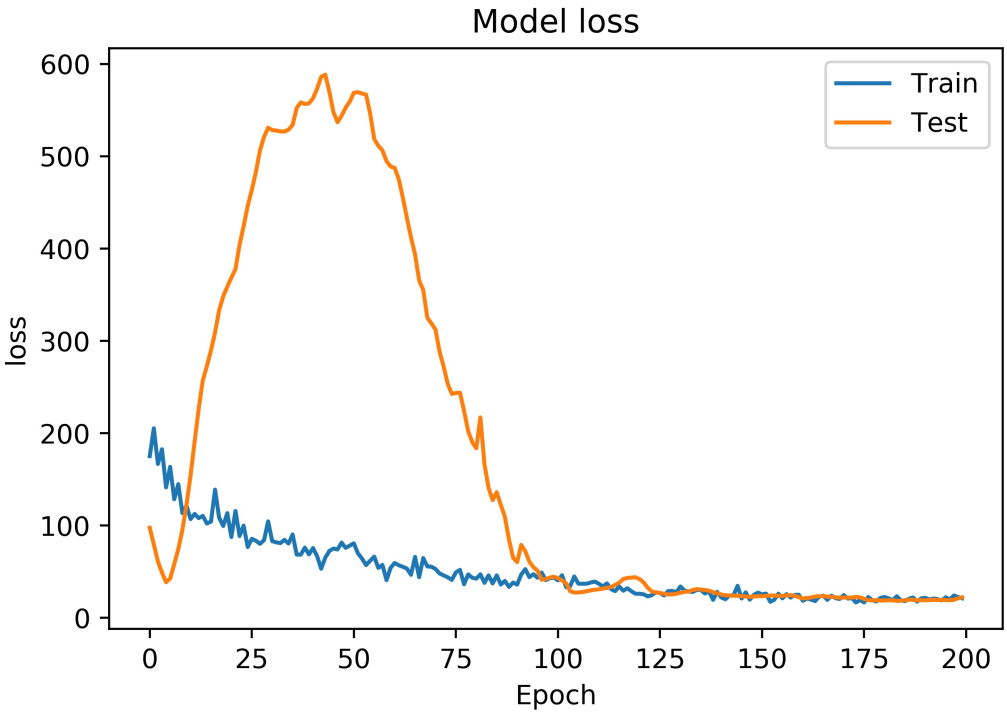


Figure 5.11: The learning curves of train and test loss versus the number of epochs for the combined-images CNN model. The x-axis represents the number of epochs, and the y-axis represents the loss (i.e., MAPE).

approaches: a) MLP, and b) CNN (particularly because the extracted characteristics of experimental stimuli were not as informative as the original data that is in the form of images). These DL experiments are a proof of concept that it is possible to identify the performance (hit count) for a given image. When it comes to images, CNN tends to perform better than typical MLP, and it did perform better in this work. Therefore, the CNN model can be used to avoid any variance or bias in terms of unbalanced characteristics of images across the experimental image pool.

Chapter 6

Conclusion and Future Work

In this preliminary work, four low-fidelity filters were designed and the performance of each filter was evaluated through three analyses: visual analysis, fixation analysis, and saccade analysis. The results from these analyses follow the same trend that the White Background filter outperforms the other filters. However, the White Background filter does not work all the time, especially when understanding the global meaning requires the context of the image. As a result, a high-fidelity filter was designed to systematically desaturate the images while blurring the background with the heat maps from the OSIE dataset to improve upon the White Background. The new design was a result from the success of the preliminary study as well as discussions with SLPs including what would be the best global filter to help shift eye gaze to the global features for people with ASD .

In the high-fidelity filter evaluation, the filter performance is based on two analyses: verbal responses and eye gaze behavior. The results from both analyses inform that the participants scored slightly higher in baseline than the filtered images, which does not conform with the hypothesis. The way that the participants were prompted

to respond verbally provided them a cue to respond globally (i.e., "What was the picture about?"). Additionally, participants may have been confused by the filtered images as no explanation as to the purpose of the filter was provided. This led to the analysis of image characteristics to see if there is any contributing factors to this result. Analyses of luminance, chroma, and spatial frequency for each image were conducted. The findings reveal that there is no significant correlation between each characteristic to the hit rate (performance). However, spatial frequency depicts higher correlation with the performance compared to the other characteristics. Therefore, suggesting that filtering low level characteristics of images outside of the AOIs at first blush does not appear to change eye gaze.

Additional analyses using ML algorithms were completed to understand the relationship between each image and hit count. CNN outperforms MLP because the nature of our data is in image format. This work demonstrates a proof of concept that we can build a predictive model for identifying images in the future experimental image pool.

There are a number of ways to improve this work. First, future experiments could be improved by separating verbal behavior from eye tracking–specifically by collecting eye-tracking data for the first session and the verbal behavior in a subsequent session. This method will allow the participants to view the images freely without being primed to think of global answers. The eye-tracking data will be a cleaner version of the current work. Then in a second session, verbal responses to the question "What was the picture about?" could be collected. If circumstances allow, collecting eye-tracking data during the second session could provide additional insight as to the impact of the verbal prompt on eye gaze.

Second, images could be screened rather than simply selected from the first 50 of 700 available images in this work. Because there is a stronger correlation between

spatial frequency and the hit count, compared to other characteristics of images in the analysis, images should be selected based on similar spatial frequency to avoid any variance or bias in terms of unbalanced spatial frequency across the experimental image pool.

Third, because of the wide variety of participants' language abilities, the verbal scoring was extremely difficult for the SLPs to reach consensus; future work with SLPs could continue to refine the definitions of global and local in the verbal rubric and add additional examples of specific language for each image. The aim is to improve the confidence of raters that these concepts are distinct and measurable. Also, additional scorers who do not identify as Caucasian or as women should be added to expand the voices that are reflected in the scoring.

Fourth, re-running the study with new participants using a screen-based eye tracker (instead of a head-mounted eye-tracking device) would allow for automatic data collection of eye gaze behavior from which more extensive data analyses could be conducted. For example, the eye gaze coordinates allow for numerically evaluating the performance of the global filter; this includes the analysis of fixations, gaze path (saccades), and a wide visual angle to accommodate peripheral viewing.

Fifth, as part of the re-run, the study will include typically developing children as additional participants. This group will allow for comparisons between the groups that could reveal the range of change and patterns across both groups of children, rather than comparing autistic children to NT young adults.

Sixth, adding typically developing children will also allow us to re-define the AOIs from the perspective of a child. Additionally, participants should be screened for their precedence–global or local.

Last, as the tools and methods are refined based on the additional data provided from

the improvements, efficacy studies to support a work-around for local interference will be conducted.

Alternatively, it is possible that filtering images does not redirect the eye gaze of viewers with local precedence. Given the amount of changes to be made to the current work, it is possible that significant change could occur given the extensive groundwork laid out here–spurring on the promise of Data Science as an encouraging tool to design, develop, and evaluate assistive technology. Future data-driven technology will allow for the deployment of real-time work-arounds providing assistance for people with ASD. With these real-time work-arounds, there is an opportunity to also build reciprocal, real-time systems that highlight local details for the NT and/or those with global precedence–thus opening up a new avenue for "neuro-shared spaces" [68].

Bibliography

[1] Eye trackers and eye tracking hardware. https://www.sr-research.com/hardware/. Accessed: 2020-02-19.

[2] Repository: Predicting human gaze beyond pixels. https://github.com/NUS-VIP/predicting-human-gaze-beyond-pixels. Accessed: 2020-02-23.

[3] Social communication disorder. https://www.asha.org/Practice-Portal/Clinical-Topics/Social-Communication-Disorder/. Accessed: 2020-03-07.

[4] Speak up with symbol-based aac. https://www.assistiveware.com/products/proloquo2go. Accessed: 2020-04-22.

[5] M. Ahissar and S. Hochstein. The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8:457–464, 2004.

[6] D. G. Altman and J. M. Bland. Statistics notes. treatment allocation in controlled trials: why randomise? *BMJ*, 318 7192:1209, 1999.

[7] C. H. Anderson, D. C. V. Essen, and B. A. Olshausen. Directed visual attention and the dynamic control of information flow. 2005.

[8] A. P. Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)*. 5 edition, 2013.

[9] S. Baron-Cohen. Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders*, 18:379–402, 1988.

[10] N. Baumann and J. Kuhl. Positive affect and flexibility: Overcoming the precedence of global over local processing of visual information. *Motivation and Emotion*, 29:123–134, 2005.

[11] M. Behrmann, C. Thomas, and K. Humphreys. Seeing it differently: visual processing in autism. *Trends in Cognitive Sciences*, 10:258–264, 2006.

[12] D. J. V. D. Berg, S. E. Boehnke, R. A. Marino, D. P. Munoz, and L. Itti. Free viewing of dynamic stimuli by humans and monkeys. *Journal of vision*, 9 5:19.1–15, 2009.

[13] M. Burke, R. E. Kraut, and D. Williams. Social use of computer-mediated communication by adults on the autism spectrum. In *CSCW '10*, 2010.

[14] K. Chawarska and F. Shic. Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 39:1663–1672, 2009.

[15] Y. Cheng. Book review: Terhi korkiakangas, communication, gaze and autism: A multimodal interaction perspective. 2019.

[16] J. Christie, J. P. Ginsberg, J. Steedman, J. Fridriksson, L. Bonilha, and C. Rorden. Global versus local processing: seeing the left side of the forest and the right side of the trees. *Frontiers in Human Neuroscience*, 6, 2012.

[17] F. L. Cibrian, J. Johnson, V. Sean, H. Pass, and L. Boyd. Combining eye tracking and verbal response to understand the impact of a global filter. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI '20, page 1–6, New York, NY, USA, 2020. Association for Computing Machinery.

[18] V. Clay, P. König, and S. U. König. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12(1), Apr. 2019.

[19] J. N. Constantino, S. Kennon-McGill, C. Weichselbaum, N. Marrus, A. Haider, A. Glowinski, S. E. Gillespie, C. Klaiman, A. Klin, and W. Jones. Infant viewing of social scenes is under genetic control and atypical in autism. In *Nature*, 2017.

[20] J. Davidoff, E. Fonteneau, and J. Fagot. Local and global processing: Observations from a remote culture. *Cognition*, 108:702–709, 2008.

[21] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. M. Estes, and J. M. Liaw. Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, 40 2:271–83, 2004.

[22] M. de Haan and M. R. Gunnar. Handbook of developmental social neuroscience. 2011.

[23] M. C. de Jong, H. van Engeland, and C. Kemner. Attentional effects of gaze shifts are influenced by emotion and spatial frequency, but not in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47 4:443–54, 2008.

[24] A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, Computers*, 34:455–470, 2002.

[25] C. C. Eriksen and J. T. Hoffman. The extent of processing of noise elements during selective encoding from visual displays. *Perception Psychophysics*, 14:155–160, 1973.

[26] C. W. Eriksen and J. E. Hoffman. Temporal and spatial characteristics of selective encoding from visual displays. *Perception Psychophysics*, 12:201–204, 1972.

[27] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Trans. Communications*, 43:2959–2965, 1995.

[28] N. E. V. Foster, T. Ouimet, A. Tryfon, K. A. Doyle-Thomas, E. Anagnostou, and K. Hyde. Effects of age and attention on auditory global–local processing in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 46:1415–1428, 2016.

[29] J. B. Ganz, E. R. Hong, F. D. Goodwyn, E. S. Kite, and W. Gilliland. Impact of pecs tablet computer app on receptive identification of pictures given a verbal stimulus. *Developmental neurorehabilitation*, 18 2:82–7, 2015.

[30] B. A. Gargaro, T. May, B. J. Tonge, D. M. Sheppard, J. L. Bradshaw, and N. J. Rinehart. Attentional mechanisms in autism, adhd, and autism-adhd using a local-global paradigm. *Journal of attention disorders*, 22 14:1320–1332, 2018.

[31] P. F. Gerhardt, F. Cicero, and E. A. Mayville. Employment and related services for adults with autism spectrum disorders. 2014.

[32] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[33] L. Granka, M. K. Feusner, and L. Lorigo. Eyetracking in online search. 2007.

[34] T. F. Gross. Global-local precedence in the perception of facial age and emotional expression by children with autism and other developmental disabilities. *Journal of Autism and Developmental Disorders*, 35:773–785, 2005.

[35] K. Holmqvist, J. van de Weijer, M. Nyström, R. Andersson, H. Jarodzka, and R. Dewhurst. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 1 edition, 2011.

[36] D. L. Kimmel, D. Mammo, and W. T. Newsome. Tracking the eye non-invasively: simultaneous comparison of the scleral search coil and optical tracking techniques in the macaque monkey. In *Front. Behav. Neurosci.*, 2012.

[37] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[38] A. Klin, C. A. Saulnier, S. S. Sparrow, D. V. Cicchetti, F. Volkmar, and C. Lord. Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: The vineland and the ados. *Journal of Autism and Developmental Disorders*, 37:748–759, 2007.

[39] K. Koldewyn, Y. V. Jiang, S. Weigelt, and N. Kanwisher. Global/local processing in autism: Not a disability, but a disinclination. *Journal of Autism and Developmental Disorders*, 43:2329–2340, 2013.

[40] S. D. König and E. A. Buffalo. A nonparametric method for detecting fixations and saccades using cluster analysis: Removing the need for arbitrary thresholds. *Journal of Neuroscience Methods*, 227:121–131, 2014.

[41] T. K. Korkiakangas and J. Rae. The interactional use of eye-gaze in children with autism spectrum disorders. 2014.

[42] S. D. König and E. A. Buffalo. Cluster fix for matlab. `https://buffalomemorylab.com/clusterfix`. Accessed: 2020-02-25.

[43] S. Li, J. T. Kwok, and Y. Wang. Combination of images with diverse focuses using the spatial frequency. *Inf. Fusion*, 2:169–176, 2001.

[44] J. C. Liechty, R. Pieters, and M. Wedel. Global and local covert visual attention: Evidence from a bayesian hidden markov model. *Psychometrika*, 68:519–541, 2003.

[45] D. B. Liston, A. E. Krukowski, and L. S. Stone. Saccade detection during smooth tracking. *Displays*, 34:171–176, 2013.

[46] A. Mordvintsev and A. K. Opencv-python tutorials. `https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_tutorials.html`. Accessed: 2020-03-03.

[47] L. Mottron, J. A. Burack, G. Iarocci, S. Belleville, and J. T. Enns. Locally oriented perception with intact global processing among adolescents with high-functioning autism: evidence from multiple paradigms. *Journal of child psychology and psychiatry, and allied disciplines*, 44 6:904–13, 2003.

[48] P. Mundy, M. Sigman, J. L. Ungerer, and T. Sherman. Defining the social deficits of autism: the contribution of non-verbal communication measures. *Journal of child psychology and psychiatry, and allied disciplines*, 27 5:657–669, 1986.

[49] M. S. Murphy, D. I. Brooks, and R. G. Cook. Pigeons use high spatial frequencies when memorizing pictures. *Journal of experimental psychology. Animal learning and cognition*, 41 3:277–85, 2015.

[50] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

[51] D. Navon and J. Norman. Does global precedence reality depend on visual angle? *Journal of Experimental Psychology: Human Perception and Performance*, 9(6):955–965, 1983.

[52] D. H. Navon. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9:353–383, 1977.

[53] M. Nyström and K. Holmqvist. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42:188–204, 2010.

[54] A. Oliva and P. G. Schyns. Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34(1):72 – 107, 1997.

[55] D. H. Ortega, F. L. Cibrian, and M. Tentori. Bendablesound: a fabric-based interactive surface to promote free play in children with autism. In *ASSETS*, 2015.

[56] J. Otero-Millan, X. G. Troncoso, S. L. Macknik, I. Serrano-Pedraza, and S. Martinez-Conde. Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of vision*, 8 14:21.1–18, 2008.

[57] K. C. Plaisted, J. Swettenham, and L. Rees. Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of child psychology and psychiatry, and allied disciplines*, 40 5:733–42, 1999.

[58] Q. V. L. Prajit Ramachandran, Barret Zoph. Searching for activation functions, 2018.

[59] F. H. Previc. Functional specialization in the lower and upper visual fields in humans: Its ecological origins and. 1990.

[60] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124 3:372–422, 1998.

[61] J. M. Rehg, G. D. Abowd, A. Rozga, M. Romero, M. A. Clements, S. Sclaroff, I. Essa, O. Y. Ousley, Y. Li, C. Kim, H. Rao, J. C. Kim, L. L. Presti, J. Zhang, D. Lantsman, J. Bidwell, and Z. Ye. Decoding children's social behavior. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3414–3421, 2013.

[62] B. Reichow and F. Volkmar. Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework. *Journal of Autism and Developmental Disorders*, 40:149–166, 2010.

[63] L. Rello and J. P. Bigham. Good background colors for readers: A study of people with and without dyslexia. In *ASSETS '17*, 2017.

[64] D. C. Richardson and R. Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29 6:1045–60, 2005.

[65] N. J. Rinehart, J. L. Bradshaw, S. A. Moss, A. V. Brereton, and B. J. Tonge. Atypical interference of local detail on global processing in high-functioning autism and asperger's disorder. *Journal of child psychology and psychiatry, and allied disciplines*, 41 6:769–78, 2000.

[66] C. E. Robertson and S. Baron-Cohen. Sensory perception in autism. *Nature Reviews Neuroscience*, 18:671–684, 2017.

[67] A. Rosebrock. Regression with keras. `https://www.pyimagesearch.com/2019/01/21/regression-with-keras`. Accessed: 2020-04-07.

[68] H. B. Rosqvist, C. Brownlow, and L. O'Dell. Mapping the social geographies of autism : online and off-line narratives of neuro-shared and separate spaces. 2013.

[69] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *ETRA*, 2000.

[70] V. Sean, F. Cibrian, J. Johnson, H. Pass, and L. Boyd. Toward digital image processing and eye tracking to promote visual attention for people with autism. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, pages 194–197, New York, NY, USA, 2019. ACM.

[71] V. Sean, J. Johnson, F. L. Cibrian, H. Pass, E. DelPizzo-Cheng, S. Jones, K. Lotich, B. Makin, D. Hughes, and L. Boyd. Designing, developing, and evaluating a global filter to work around local interference for children with autism. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI '20, page 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[72] D. P. Skwerer, B. H. Brukilacchio, A. L. Chu, B. Eggleston, S. D. Meyer, and H. Tager-Flusberg. Do minimally verbal and verbally fluent individuals with autism spectrum disorder differ in their viewing patterns of dynamic social scenes? *Autism : the International Journal of Research and Practice*, 2019.

[73] M. Spering and M. Carrasco. Acting without seeing: eye movements reveal visual processing without awareness. *Trends in Neurosciences*, 38:247–258, 2015.

[74] K. Spiel, C. Frauenberger, O. Keyes, and G. Fitzpatrick. Agency of autistic children in technology–a critical literature review. *ACM Trans. Comput.-Hum. Interact.*, 2019.

[75] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1929–1958, 2014.

[76] S. Tanggaard. Mental imagery in high-functioning autism spectrum disorder. 2016.

[77] M. Tentori and G. R. Hayes. Designing for interaction immediacy to enhance social skills of children with autism. In *UbiComp '10*, 2010.

[78] P. E. Touchette and J. S. Howard. Errorless learning: reinforcement contingencies and stimulus control transfer in delayed prompting. *Journal of applied behavior analysis*, 17 2:175–88, 1984.

[79] T. Urruty, S. Lew, N. Ihaddadene, and D. A. Simovici. Detecting eye fixations by projection clustering. *14th International Conference of Image Analysis and Processing - Workshops (ICIAPW 2007)*, pages 45–50, 2007.

[80] L. Wang, L. Mottron, D. Peng, C. Berthiaume, and M. Dawson. Local bias and local-to-global interference without global deficit: a robust finding in autism under various conditions of attention, exposure time, and visual angle. *Cognitive neuropsychology*, 24 5:550–74, 2007.

[81] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs, and Q. Zhao. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, 88:604–616, 2015.

[82] X. Wang. *BEING SOCIAL IN ISOCIAL: A Case Study of Youth with Autism Spectrum Disorder Learning Social Competence in 3D Collaborative Virtual Learning Environment.* American Academic Press, 1 edition, 2017.

[83] M. Wedel and R. Pieters. A review of eye-tracking research in marketing. 2008.

[84] M. Wedel, R. Pieters, and J. Liechty. Attention switching during scene perception: how goals influence the time course of eye movements across advertisements. *Journal of experimental psychology. Applied*, 14 2:129–38, 2008.

[85] R. M. Williams and J. E. Gilbert. Cyborg perspectives on computing research reform. In *CHI Extended Abstracts*, 2019.

[86] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich. Ability-based design: Concept, principles and examples. *TACCESS*, 3:9:1–9:27, 2011.

[87] J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao. Predicting human gaze beyond pixels. *Journal of Vision*, 14 1, 2014.

APPENDICES

A Supplemental Materials

A.1 Baseline Images

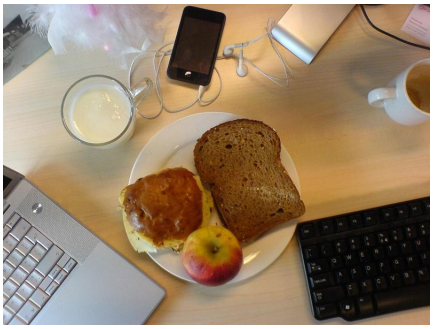
We used the first 50 images from OSIE dataset [87]. The name of images are shown as in Table A.1.

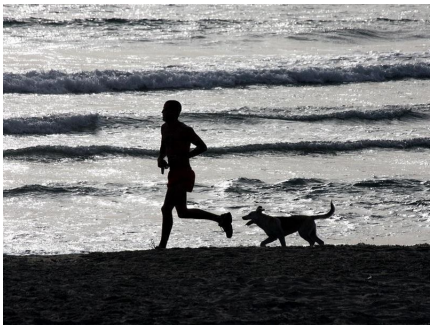
Table A.1: The name of each image that is shown below this table.

1001	1002
1003	1004

.....

1049	1050







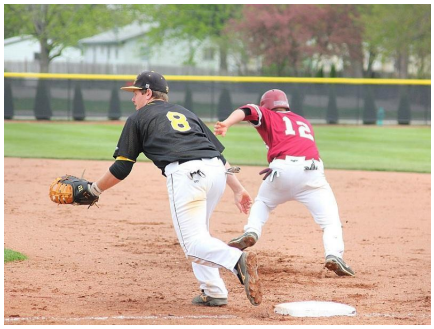




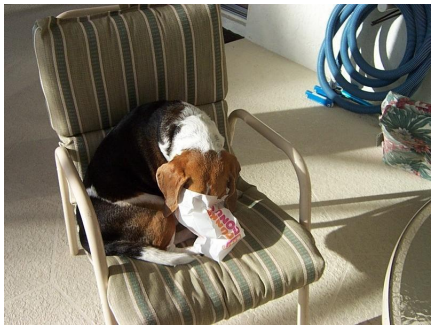
GROCERY
LALA
MILK
$2.49
BOLA
$1.99
BEEF STEW
$1.99


FOODLINK
Providing Jobs, Preventing Hunger and Training for the Future
1-800-283-9000


STOP




Your Executive
delivering for you

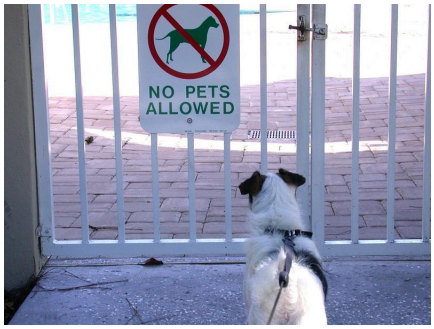






PT




NO PETS
ALLOWED





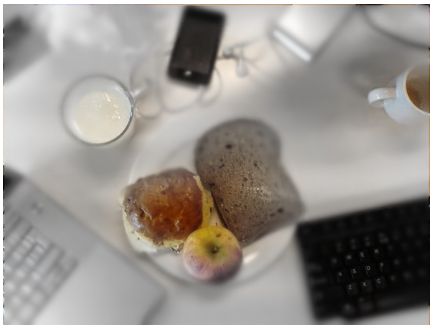
QUAKER
OATS

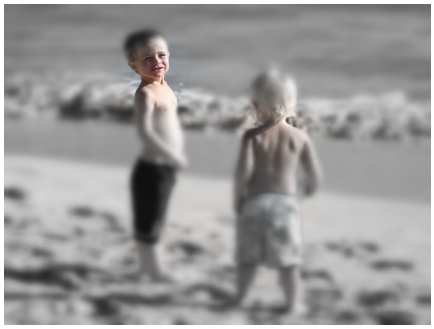
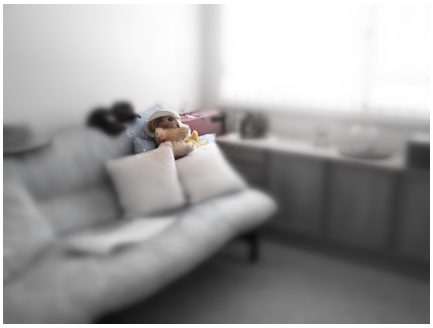
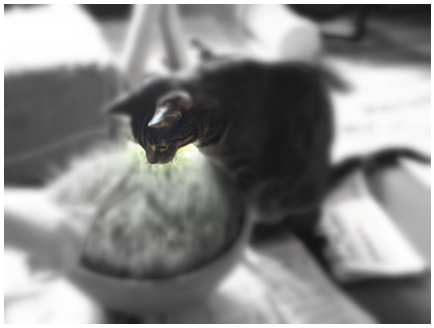
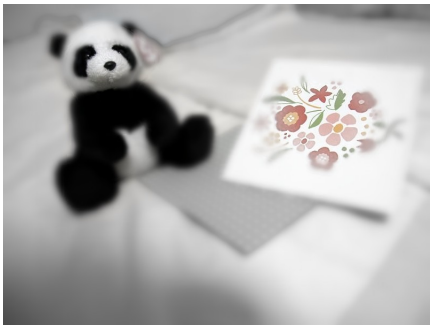


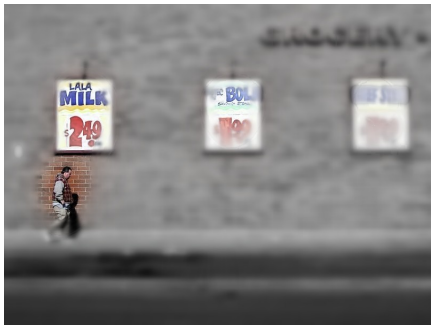

DOMUS

A.2 High-fidelity Filtered Images

Below are the 50 filtered images using the high-fidelity filter. The name of each image are also shown in Table A.1.





LALA
MILK









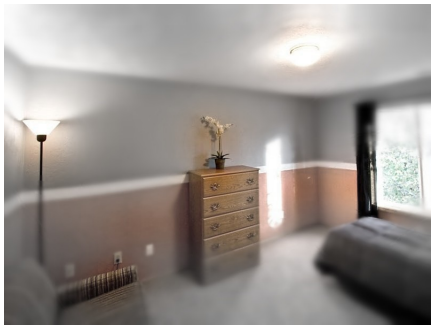
NO PETS
ALLOWED

TEB
NP P IBAS





FOODLINK
1-800-283-9000

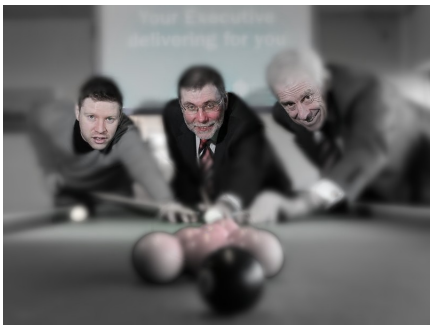
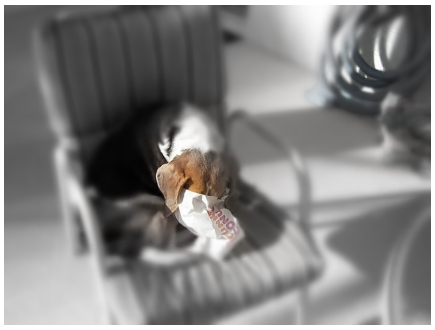
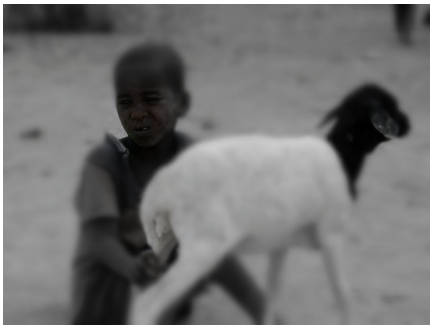



STOP

8
12

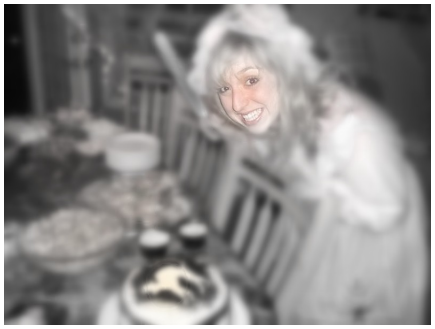

Die Olchis fliegen
in die Schule



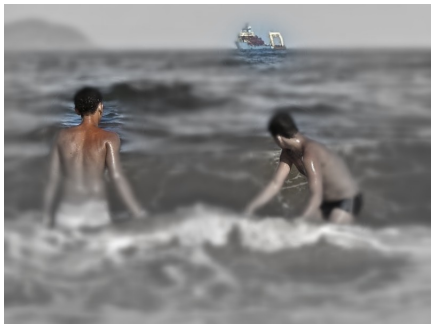




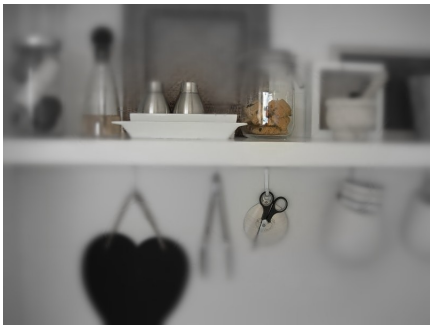







QUAKER
OATS




A.3 Baseline and High-fidelity Filtered Images Sorted Based on Highest to Lowest Hit Count

The hit count range from 0 to 10, which is the number of participants; i.e., 0 for a given image/filter means no hit count from any participant, and 10 means all participants have at least one hit for the given image. The name of images and their corresponding conditions and hit counts are shown as in Table A.2. The images from the experimental image pool are shown below, in a sorted order from the highest to the lowest hit count.

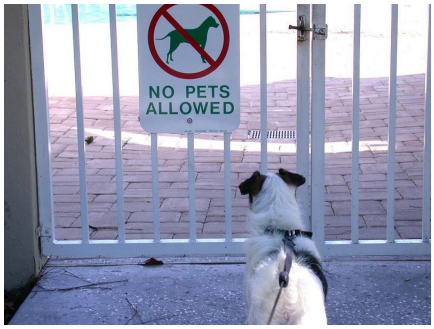
Table A.2: The sorted order of baseline and filtered images from the highest to the lowest hit count.

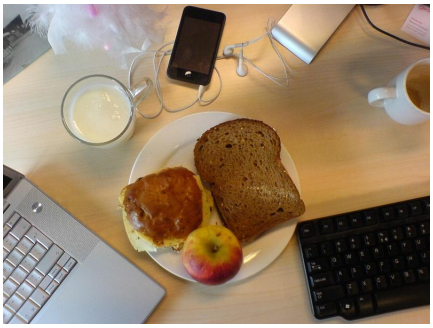
Image	Condition	Hit count	Image	Condition	Hit count
1002	Baseline	10	1030	Baseline	7
1043	Baseline	9	1007	Baseline	7
1001	Baseline	8	1032	Filtered	6
1047	Filtered	8	1017	Filtered	6
1031	Filtered	8	1033	Filtered	6
1020	Filtered	8	1002	Filtered	6
1038	Baseline	8	1030	Filtered	6
1040	Baseline	8	1026	Filtered	6
1041	Filtered	8	1021	Filtered	6
1042	Baseline	8	1003	Baseline	6
1029	Baseline	7	1027	Baseline	6
1044	Filtered	7	1019	Filtered	6
1036	Baseline	7	1018	Filtered	6
1020	Baseline	7	1003	Filtered	6
1028	Baseline	7	1012	Baseline	6
1050	Baseline	7	1048	Baseline	6
1034	Baseline	7	1039	Baseline	6
1044	Baseline	7	1004	Baseline	6
1023	Baseline	7	1049	Filtered	6
1046	Baseline	7	1008	Baseline	6

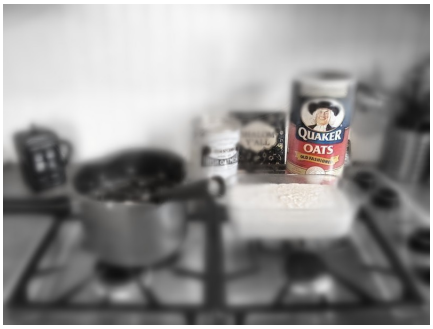
Image	Condition	Hit count	Image	Condition	Hit count
1004	Filtered	6	1016	Baseline	5
1009	Baseline	6	1008	Filtered	5
1011	Filtered	6	1007	Filtered	5
1045	Filtered	6	1013	Baseline	5
1046	Filtered	6	1050	Filtered	4
1031	Baseline	5	1010	Filtered	4
1047	Baseline	5	1009	Filtered	4
1036	Filtered	5	1038	Filtered	4
1037	Filtered	5	1016	Filtered	4
1043	Filtered	5	1017	Baseline	4
1035	Filtered	5	1019	Baseline	4
1041	Baseline	5	1035	Baseline	4
1049	Baseline	5	1006	Filtered	4
1026	Baseline	5	1006	Baseline	4
1018	Baseline	5	1033	Baseline	4
1014	Filtered	5	1023	Filtered	4
1005	Baseline	5	1032	Baseline	4
1010	Baseline	5	1005	Filtered	4
1024	Baseline	5	1025	Filtered	4
1015	Filtered	5	1029	Filtered	4

Image	Condition	Hit count	Image	Condition	Hit count
1013	Filtered	4	1022	Baseline	3
1048	Filtered	3	1022	Filtered	3
1011	Baseline	3	1039	Filtered	3
1012	Filtered	3	1015	Baseline	2
1028	Filtered	3	1027	Filtered	2
1042	Filtered	3	1021	Baseline	2
1014	Baseline	3	1024	Filtered	2
1040	Filtered	3	1001	Filtered	1
1037	Baseline	3	1025	Baseline	0
1034	Filtered	3	1045	Baseline	0


Die Olchis fliegen
in die Schule


NO PETS
ALLOWED




QUAKER
OATS


PT


Memoranda


QUAKER
OATS
SHALOM Y'ALL


Die Olchis fliegen in die Schule


Your Executive
delivering for you


STEVENSON
2








STOP

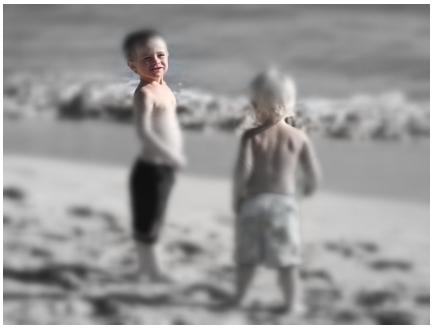



DOMUS


TEB
BNP PARIBAS

8
12

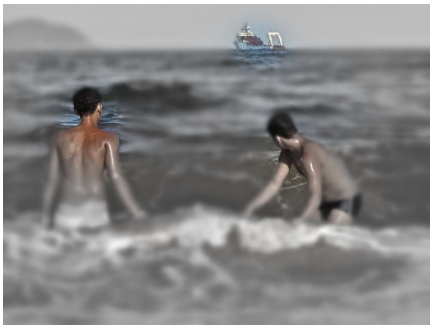








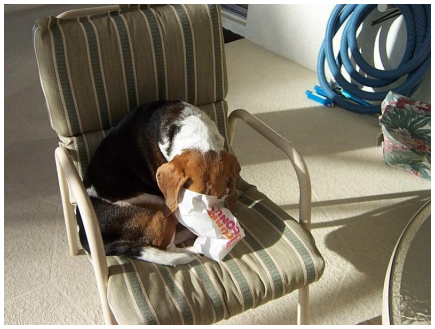
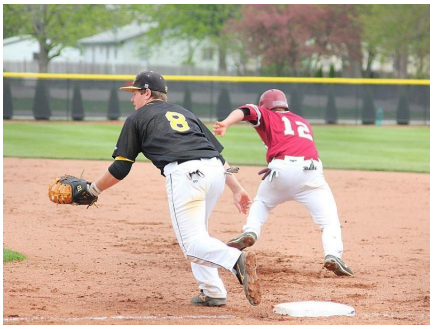




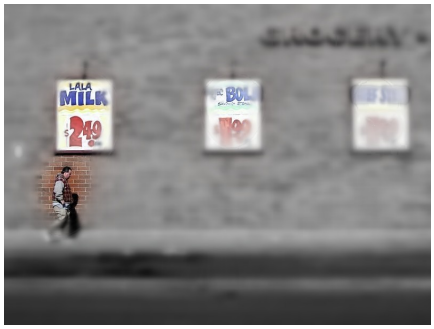
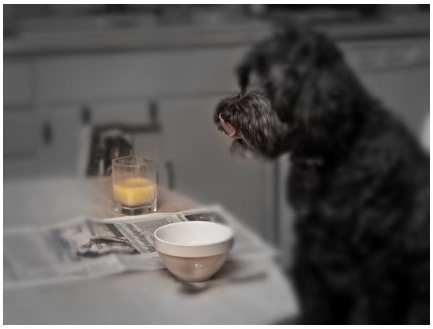


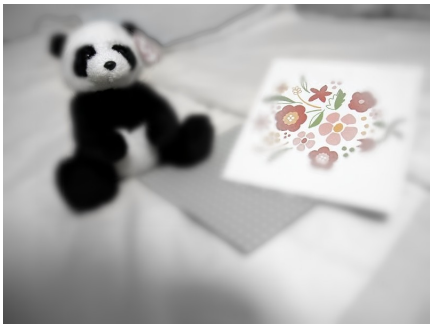



GROCERY
LALA
MILK
BOLA
BEEF STEW

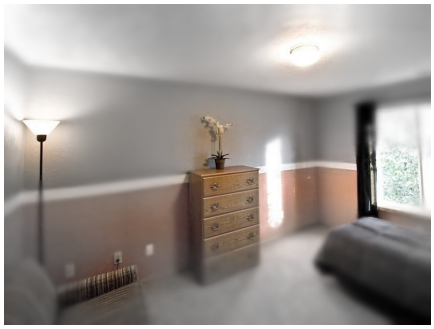





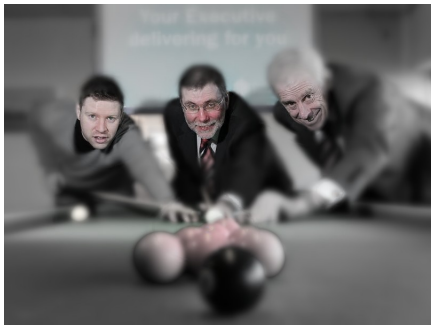

LALA
MILK


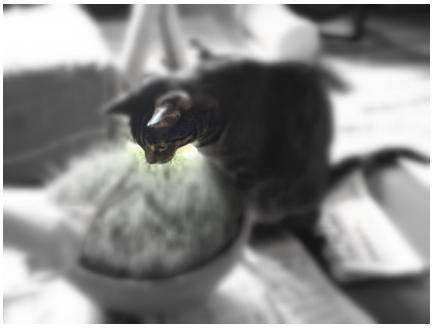
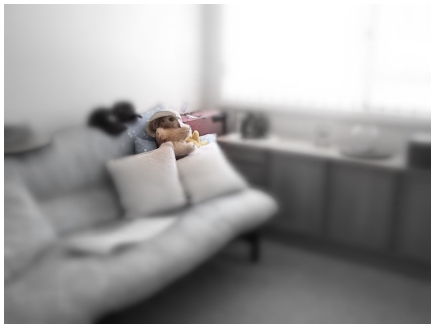


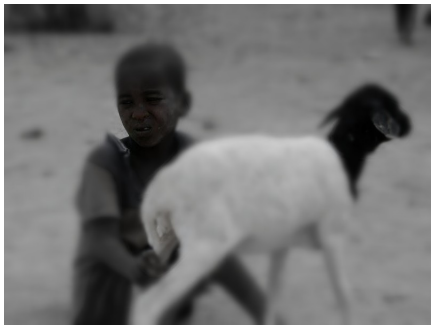
NO PETS
ALLOWED

FOODLINK
Providing Jobs, Preventing Hunger
and Training for the Future
1-800-283-9000


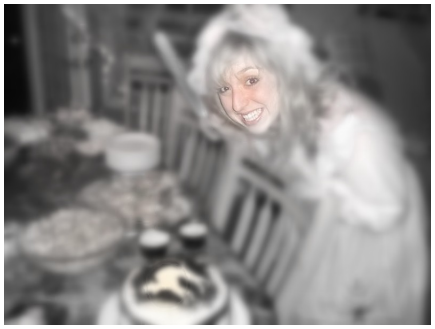

STOP

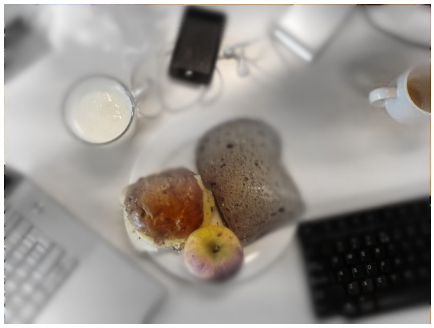



FOODLINK





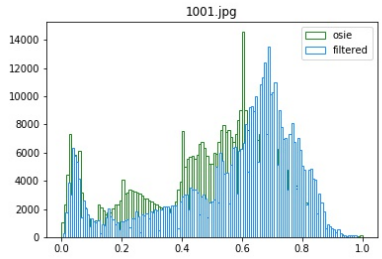


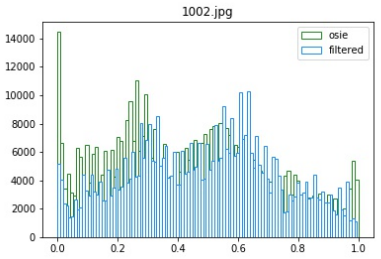

TEB
NP P IBAS


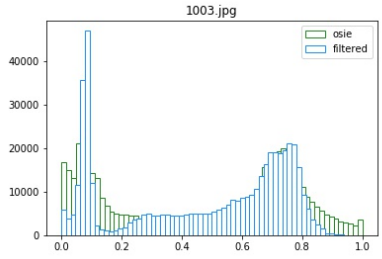


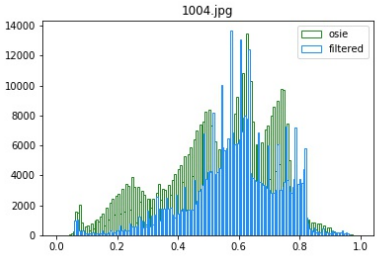
A.4 Luminance Frequency Histogram

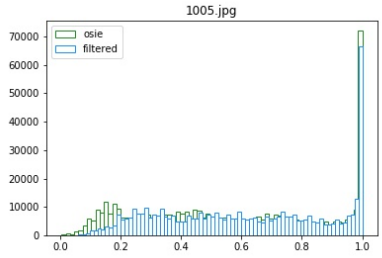
Luminance frequency histograms also follows the same order in Table A.1. The green histograms represent the baseline images from OSIE dataset. The blue histograms represent the high-fidelity filtered images. The range for luminance is between 0 and 1.

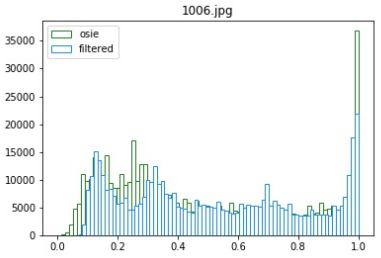

1001.jpg
osie
filtered

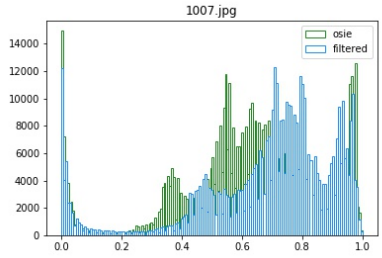

1002.jpg
osie
filtered

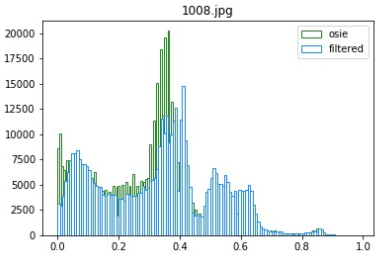

1003.jpg
osie
filtered

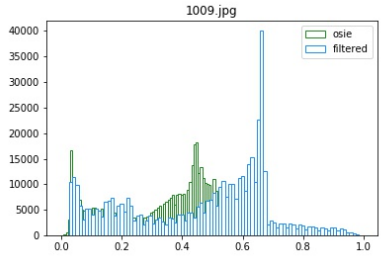

1004.jpg
osie
filtered

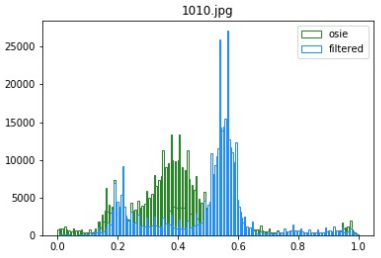

1005.jpg
osie
filtered

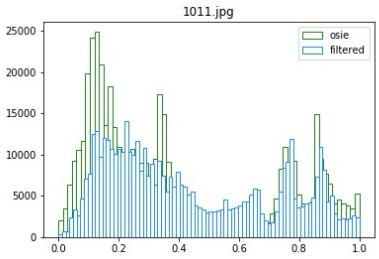

1006.jpg
osie
filtered

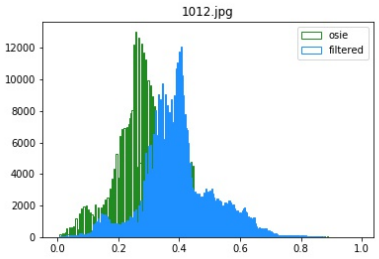

1007.jpg
osie
filtered

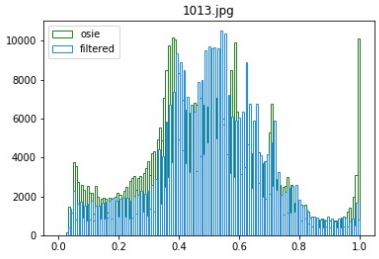

1008.jpg
osie
filtered

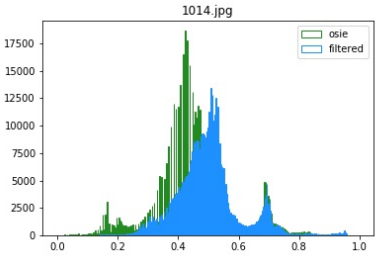

1009.jpg
osie
filtered

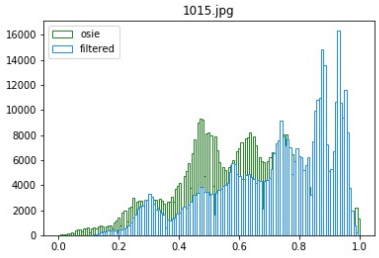

1010.jpg
osie
filtered

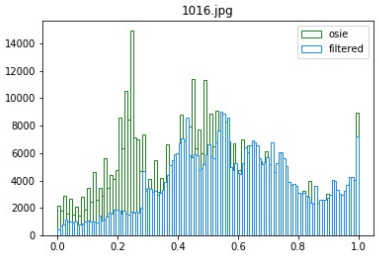

1011.jpg
osie
filtered

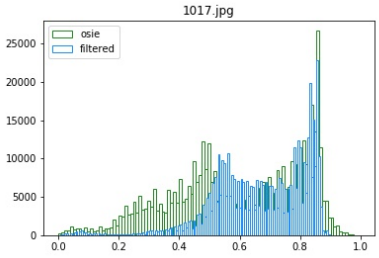

1012.jpg
osie
filtered

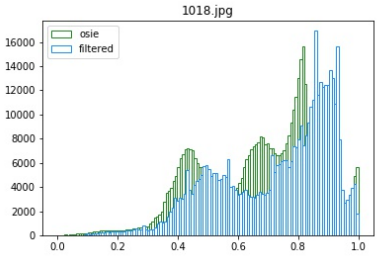

1013.jpg
osie
filtered

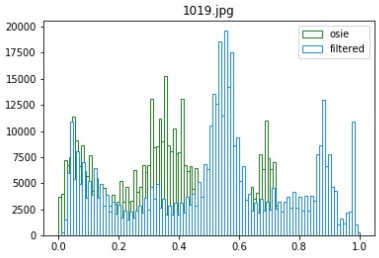

1014.jpg
osie
filtered

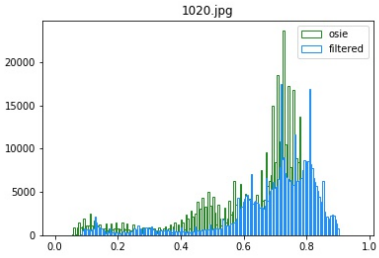

1015.jpg
osie
filtered

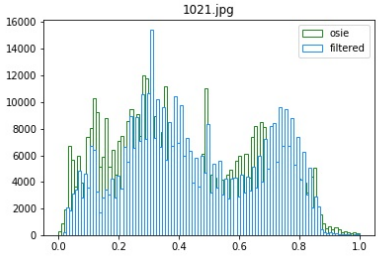

1016.jpg
osie
filtered

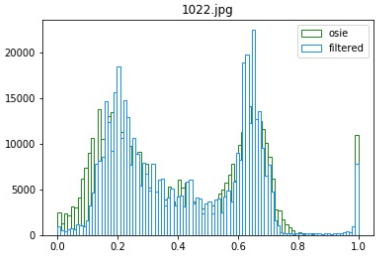

1017.jpg
osie
filtered

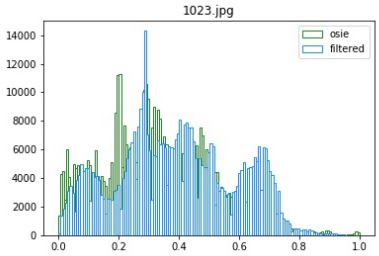

1018.jpg
osie
filtered

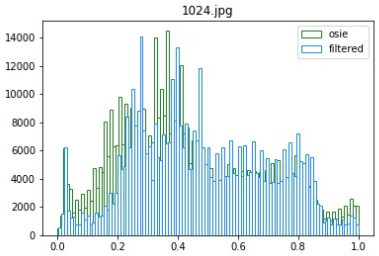

1019.jpg
osie
filtered

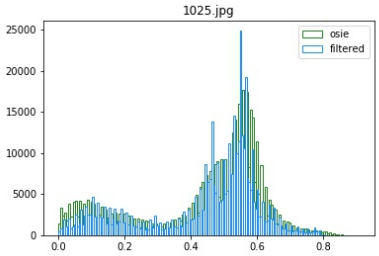

1020.jpg
osie
filtered

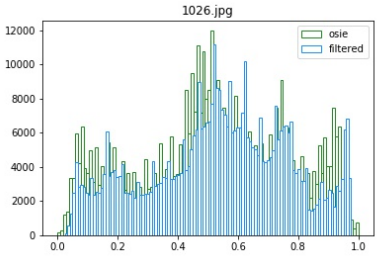

1021.jpg
osie
filtered

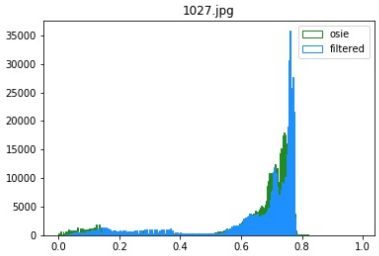

1022.jpg
osie
filtered

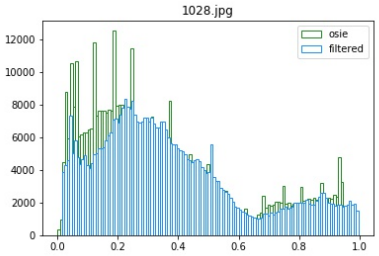

1023.jpg
osie
filtered

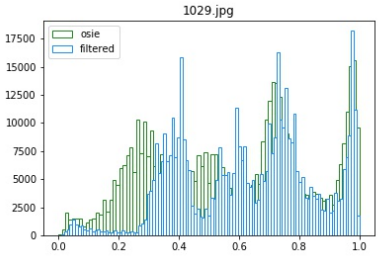

1024.jpg
osie
filtered

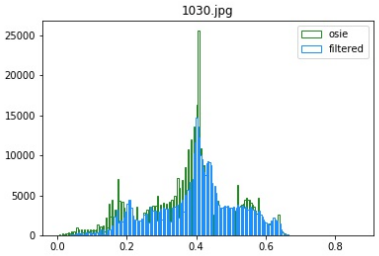

1025.jpg
osie
filtered

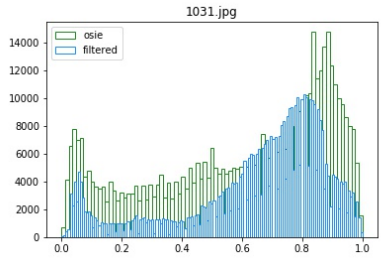

1026.jpg
osie
filtered

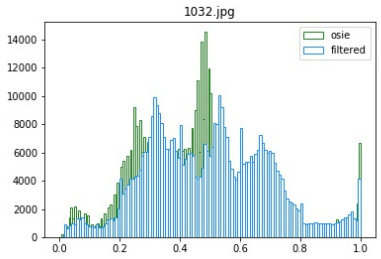

1027.jpg
osie
filtered

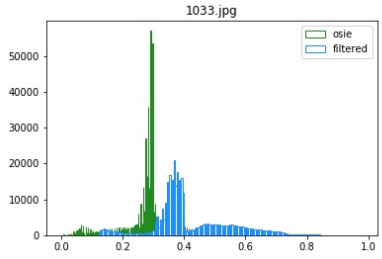

1028.jpg
osie
filtered

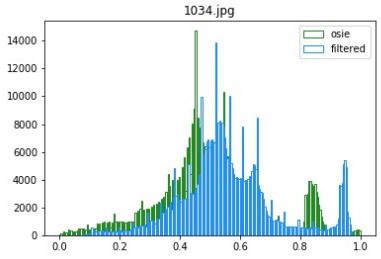

1029.jpg
osie
filtered

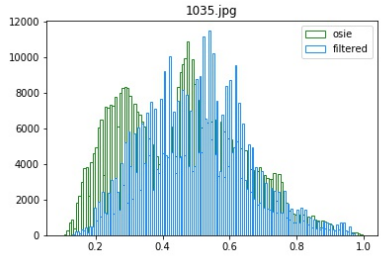

1030.jpg
osie
filtered

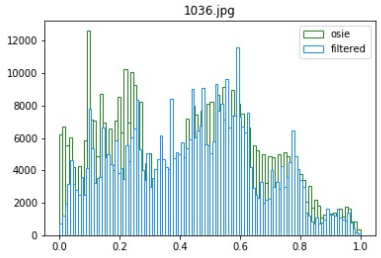

1031.jpg
osie
filtered

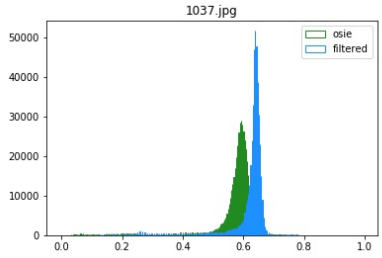

1032.jpg
osie
filtered

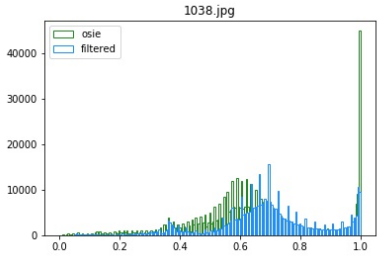

1033.jpg
osie
filtered

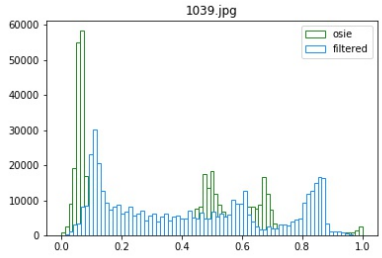

1034.jpg
osie
filtered

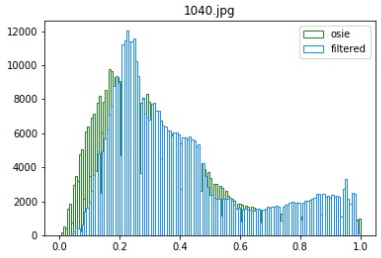

1035.jpg
osie
filtered

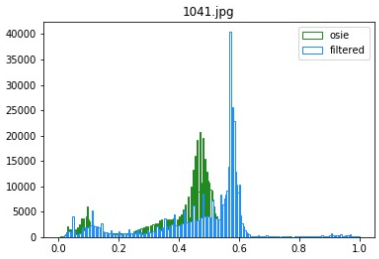

1036.jpg
osie
filtered

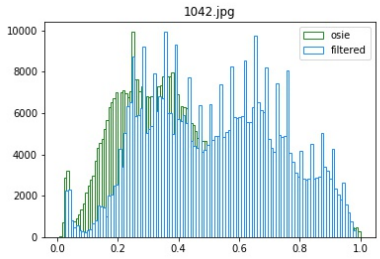

1037.jpg
osie
filtered

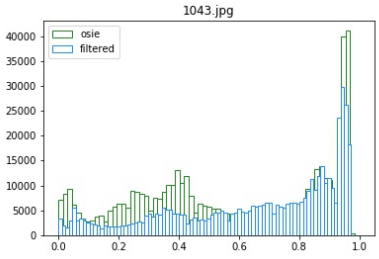

1038.jpg
osie
filtered

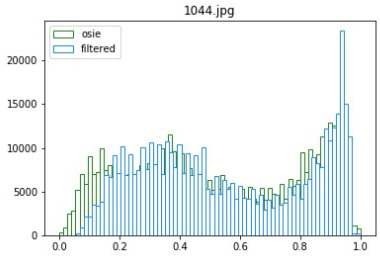

1039.jpg
osie
filtered

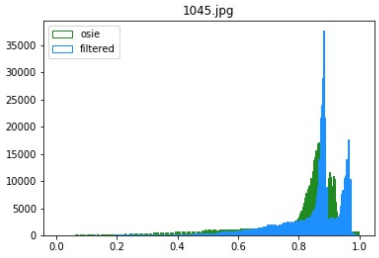

1040.jpg
osie
filtered

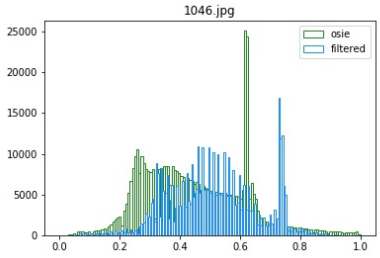

1041.jpg
osie
filtered

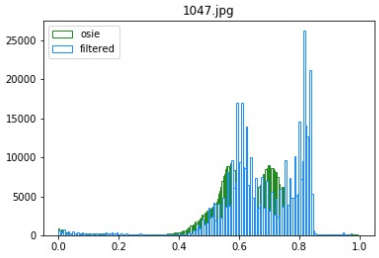

1042.jpg
osie
filtered

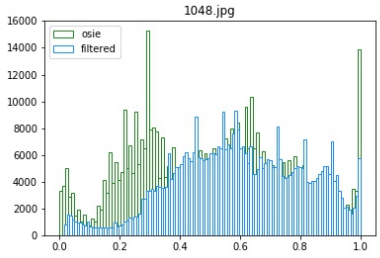

1043.jpg
osie
filtered

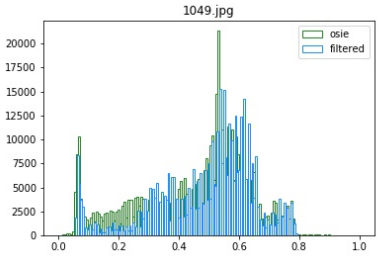

1044.jpg
osie
filtered

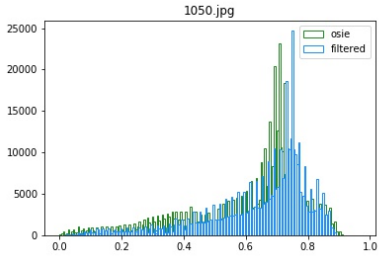

1045.jpg
osie
filtered


1046.jpg
osie
filtered


1047.jpg
osie
filtered

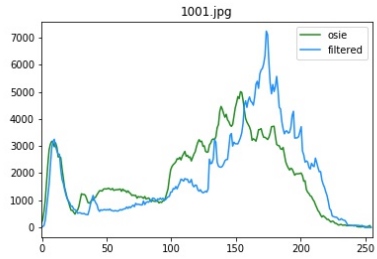

1048.jpg
osie
filtered

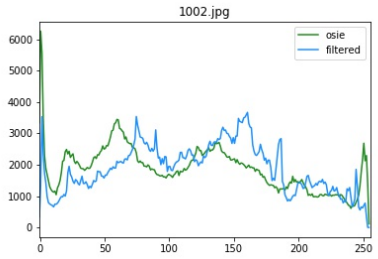

1049.jpg
osie
filtered

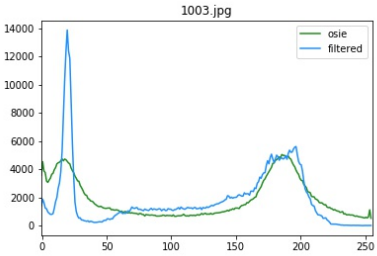

1050.jpg
osie
filtered

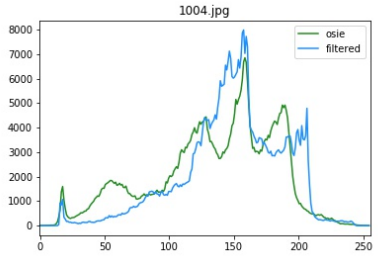
A.5 Image Histogram

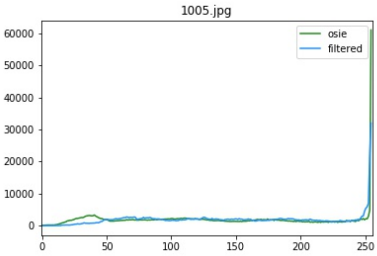
The image histograms below also follow the same order as in Table A.1. The green lines represent baseline images from OSIE dataset, and the blue lines represent the high-fidelity filtered images. The chroma value is between 0 and 255.

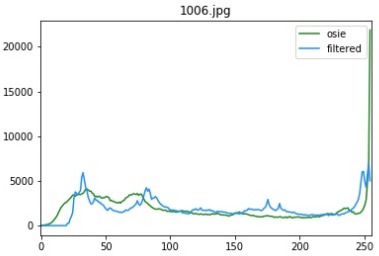

1001.jpg
osie
filtered

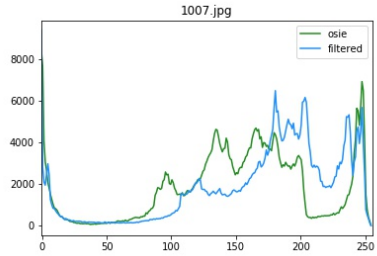

1002.jpg
osie
filtered

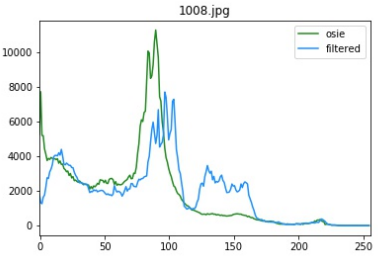

1003.jpg
osie
filtered

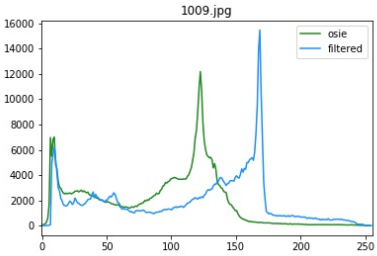

1004.jpg
osie
filtered

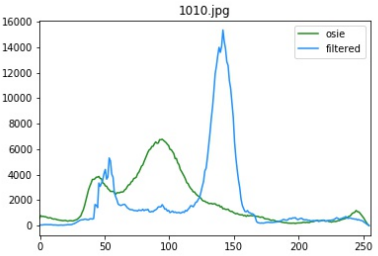

1005.jpg
osie
filtered

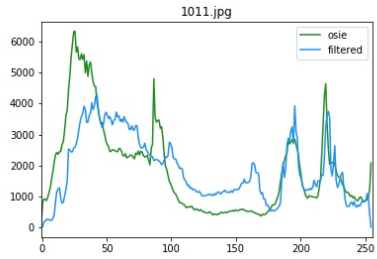

1006.jpg
osie
filtered

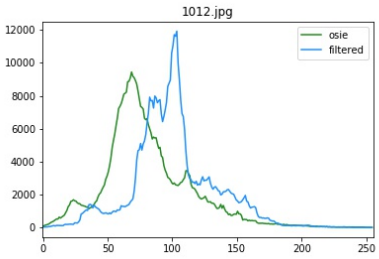

1007.jpg
osie
filtered

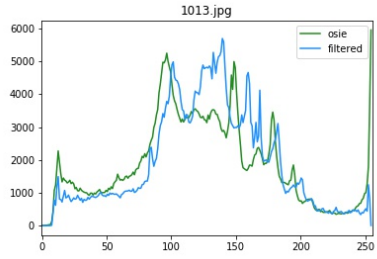

1008.jpg
osie
filtered

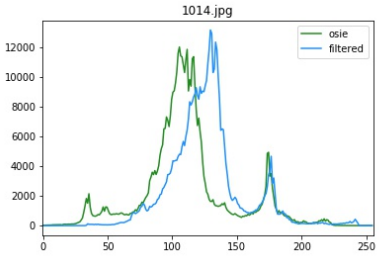

1009.jpg
osie
filtered

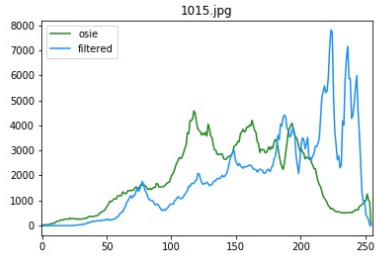

1010.jpg
osie
filtered

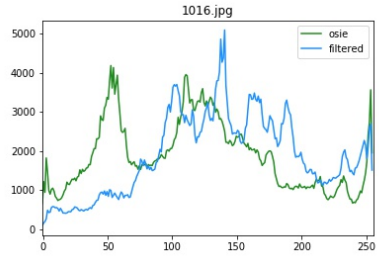

1011.jpg
osie
filtered

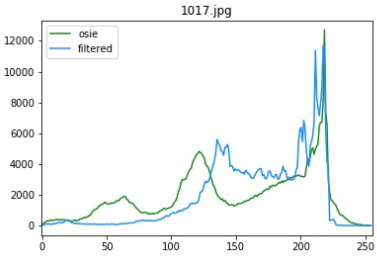

1012.jpg
osie
filtered

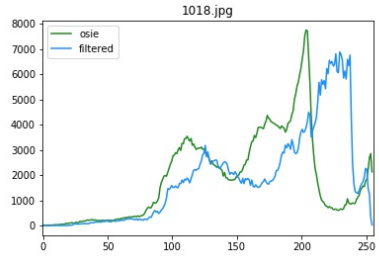

1013.jpg
osie
filtered

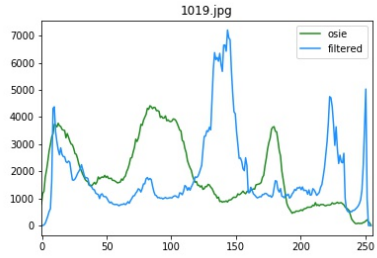

1014.jpg
osie
filtered

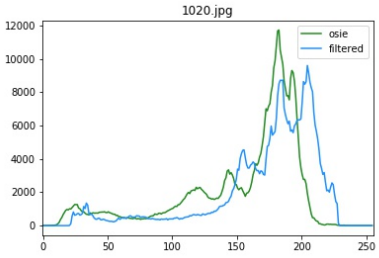

1015.jpg
osie
filtered

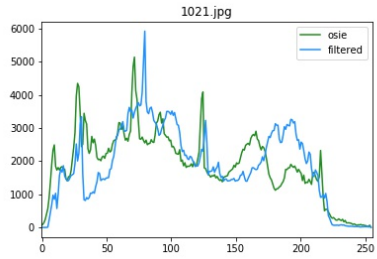

1016.jpg
osie
filtered

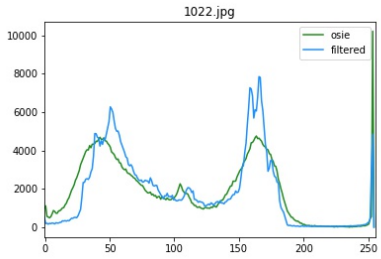

1017.jpg
osie
filtered

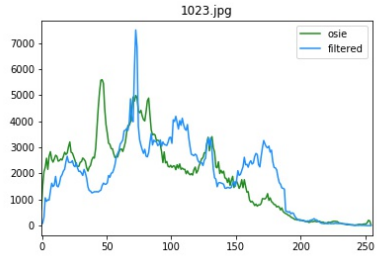

1018.jpg
osie
filtered

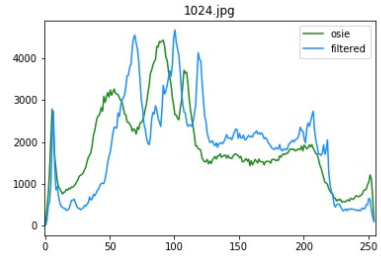

1019.jpg
osie
filtered

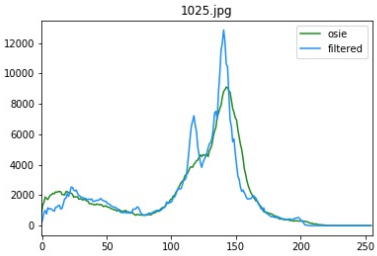

1020.jpg
osie
filtered

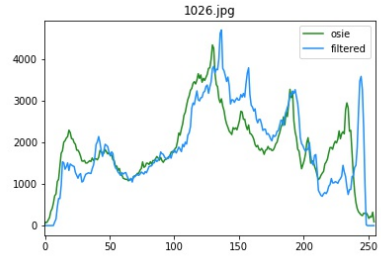

1021.jpg
osie
filtered

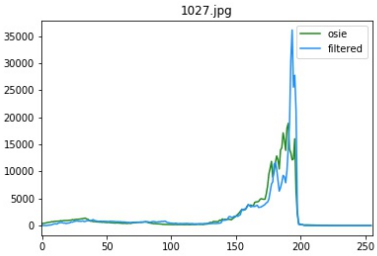

1022.jpg
osie
filtered

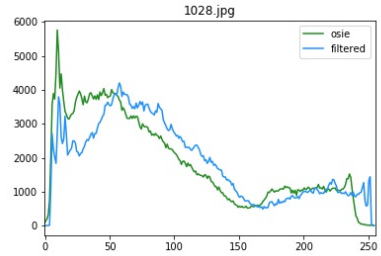

1023.jpg
osie
filtered

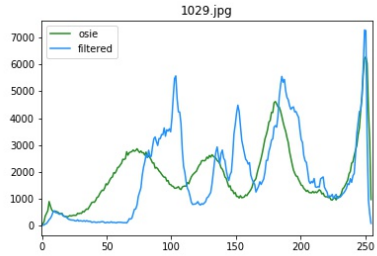

1024.jpg
osie
filtered

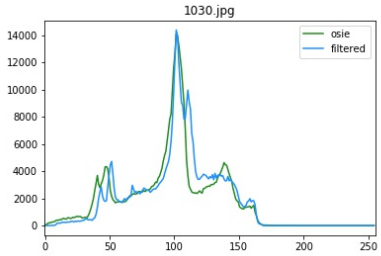

1025.jpg
osie
filtered

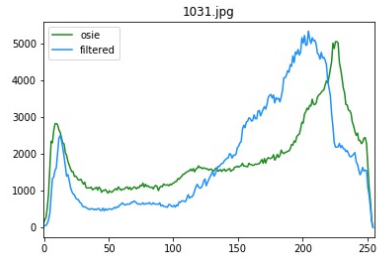

1026.jpg
osie
filtered

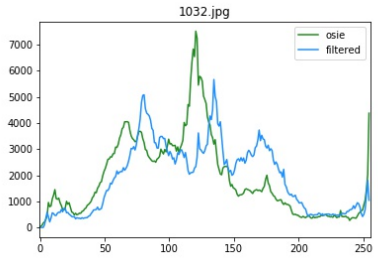

1027.jpg
osie
filtered

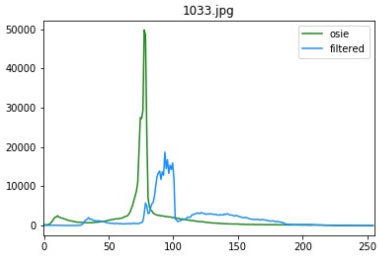

1028.jpg
osie
filtered

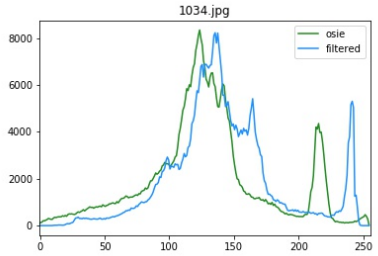

1029.jpg
osie
filtered

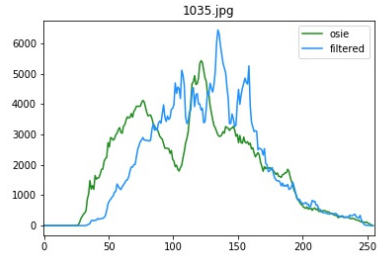

1030.jpg
osie
filtered

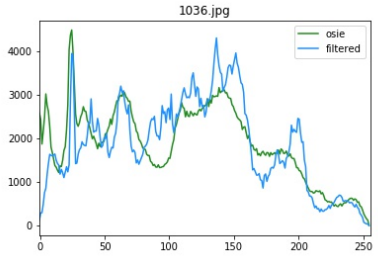

1031.jpg
osie
filtered

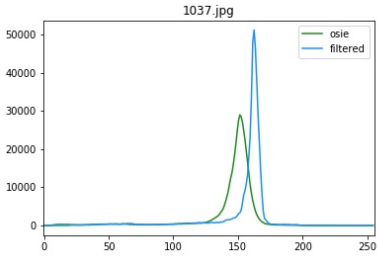

1032.jpg
osie
filtered

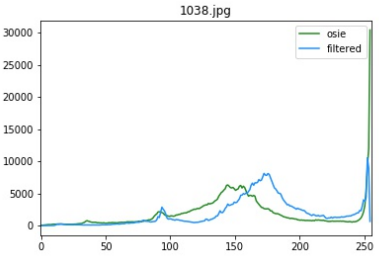

1033.jpg
osie
filtered

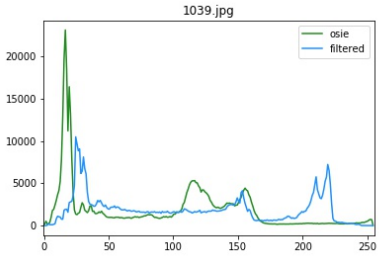

1034.jpg
osie
filtered

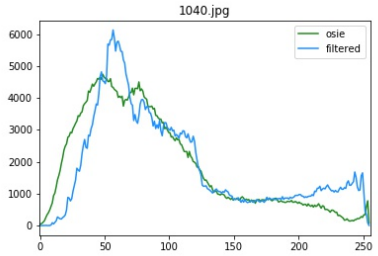

1035.jpg
osie
filtered

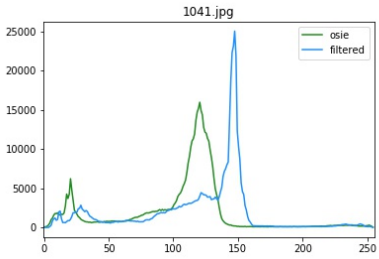

1036.jpg
osie
filtered

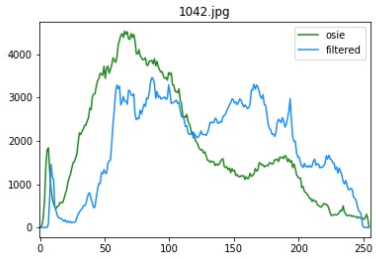

1037.jpg
osie
filtered

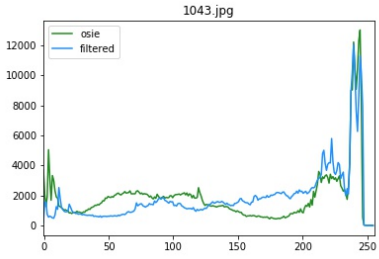

1038.jpg
osie
filtered

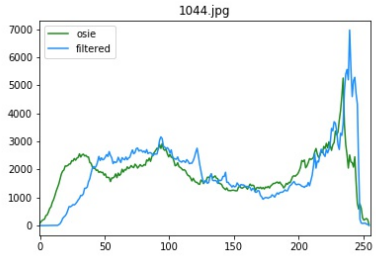

1039.jpg
osie
filtered

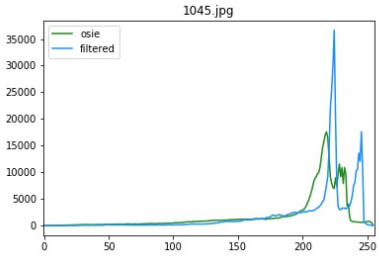

1040.jpg
osie
filtered

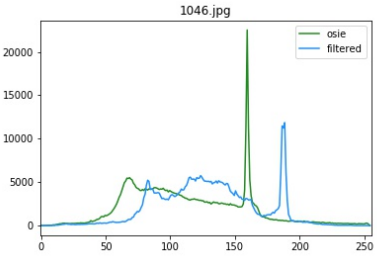

1041.jpg
osie
filtered

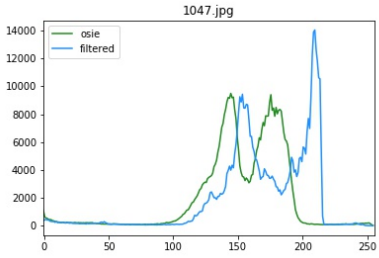

1042.jpg
osie
filtered

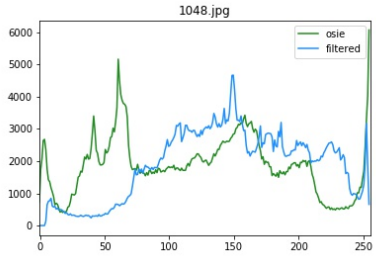

1043.jpg
osie
filtered

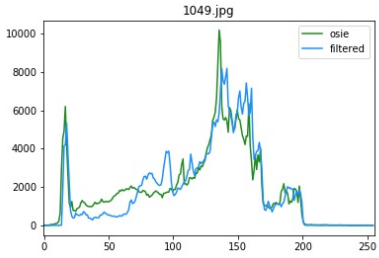

1044.jpg
osie
filtered

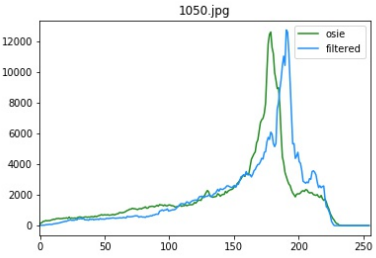

1045.jpg
osie
filtered


1046.jpg
osie
filtered


1047.jpg
osie
filtered


1048.jpg
osie
filtered

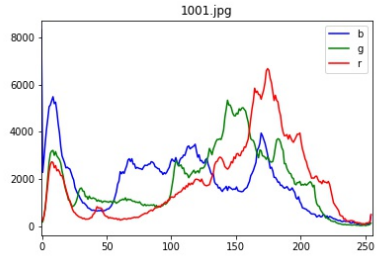

1049.jpg
osie
filtered

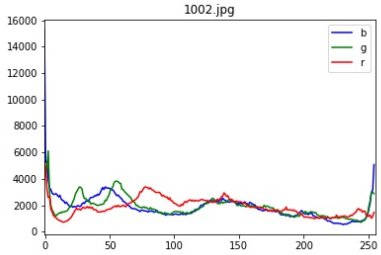

1050.jpg
osie
filtered

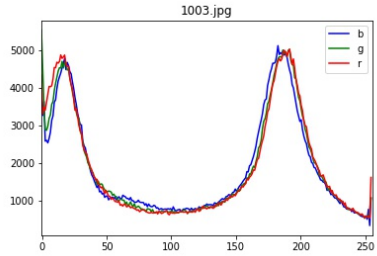
A.6 Image Histogram in Separate Channels: Blue, Green, Red

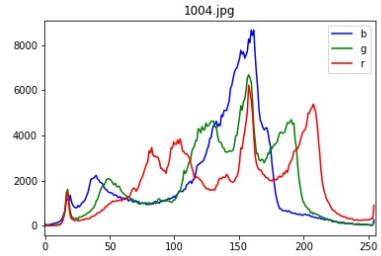
There are two groups of image histograms which is separated into the three channels: blue, green, and red. The first group is for baseline/raw image. The second group is for high-fidelity filtered images. The images in each group follow the same order in Table A.1.

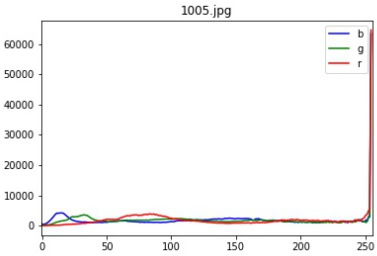
Baseline/Raw Images

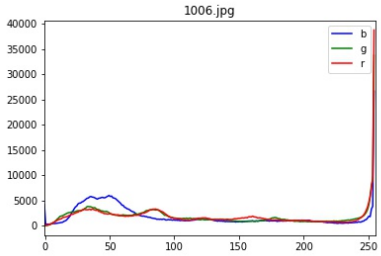


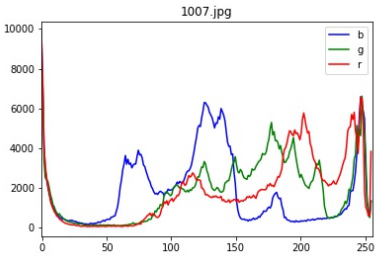


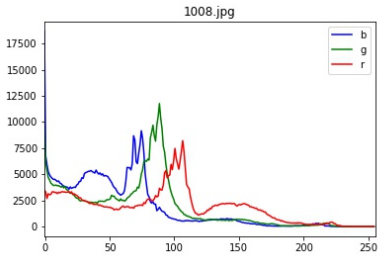


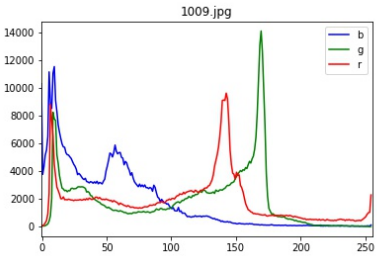


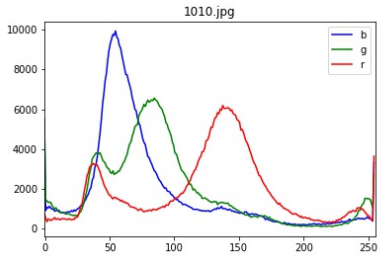


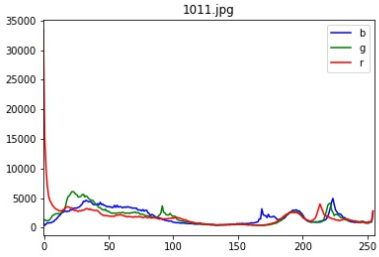


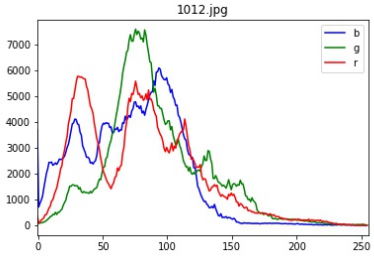


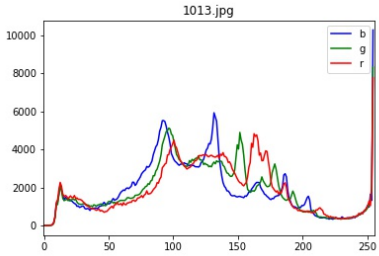


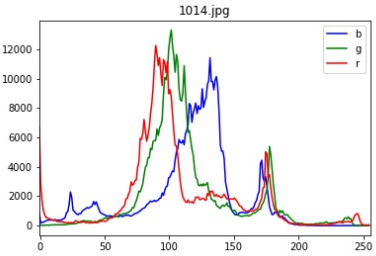


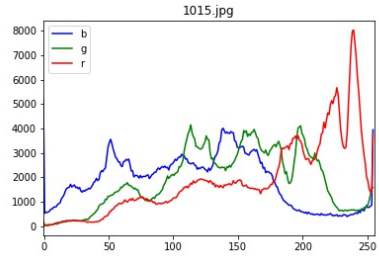


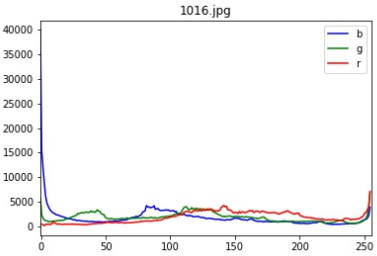

1011.jpg
b
g
r

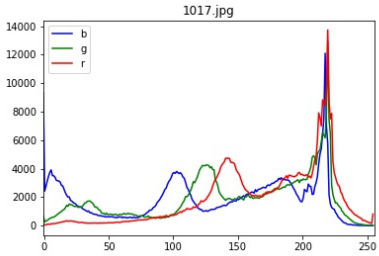

1012.jpg
b
g
r

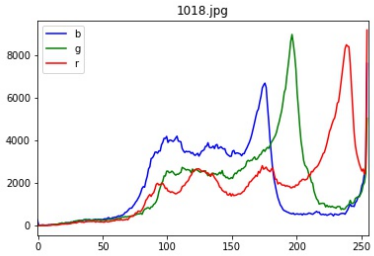

1013.jpg
b
g
r

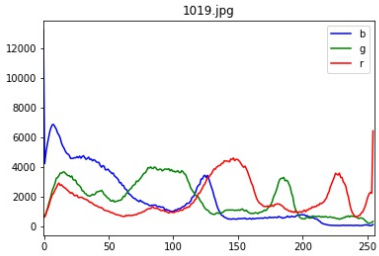

1014.jpg
b
g
r

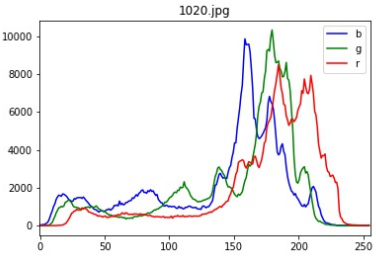

1015.jpg
b
g
r

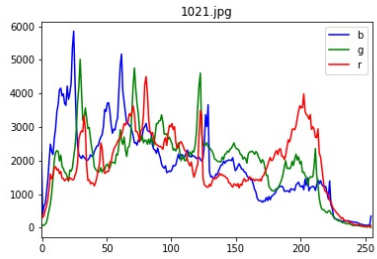

1016.jpg
b
g
r

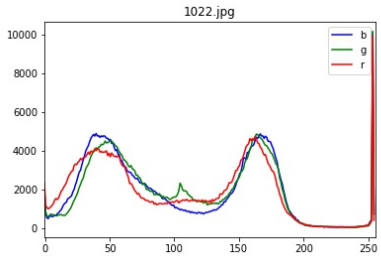

1017.jpg
b
g
r

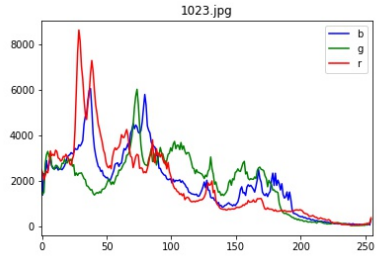

1018.jpg
b
g
r

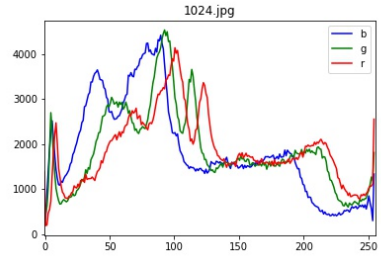

1019.jpg
b
g
r

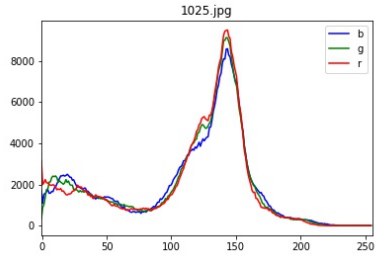

1020.jpg
b
g
r

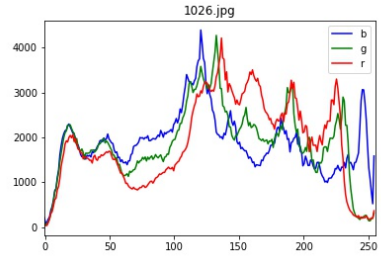

1021.jpg
b
g
r

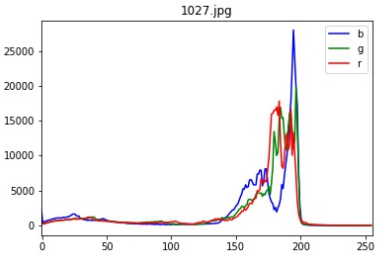

1022.jpg
b
g
r

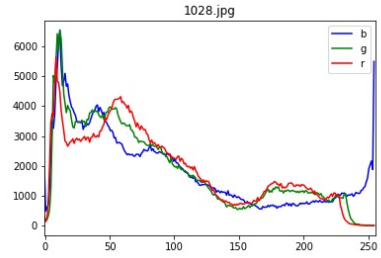

1023.jpg
b
g
r

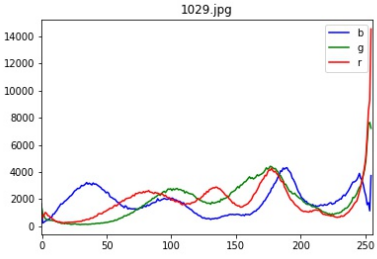

1024.jpg
b
g
r

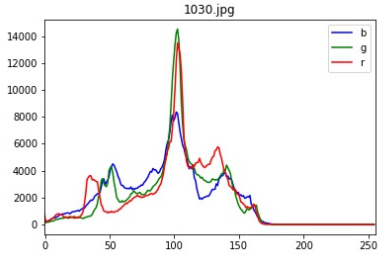

1025.jpg
b
g
r

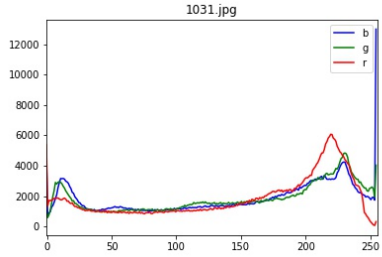

1026.jpg
b
g
r

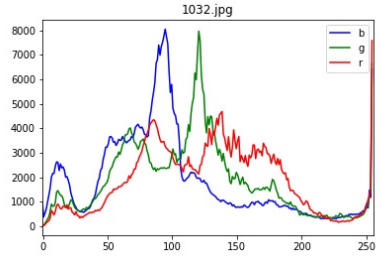

1027.jpg
b
g
r

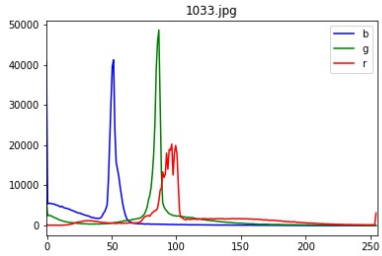

1028.jpg
b
g
r

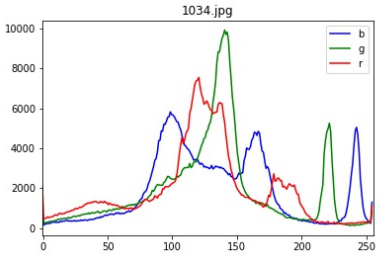

1029.jpg
b
g
r

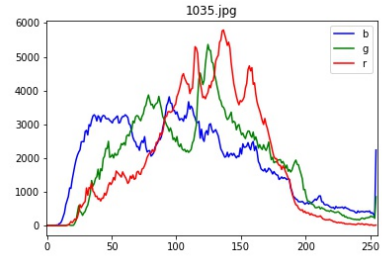

1030.jpg
b
g
r

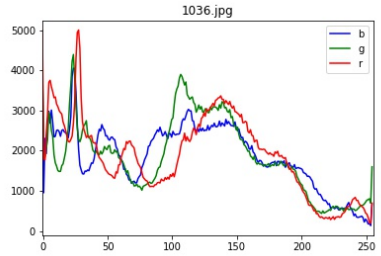

1031.jpg
b
g
r

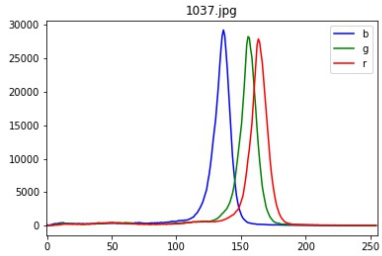

1032.jpg
b
g
r

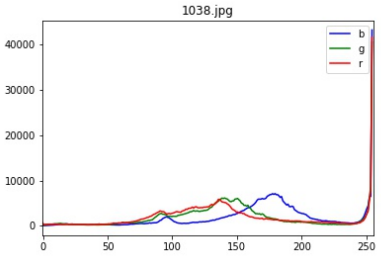

1033.jpg
b
g
r

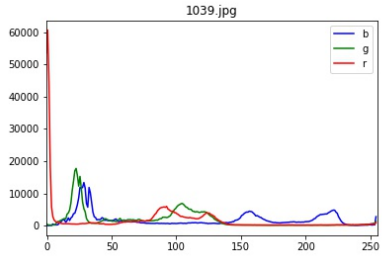

1034.jpg
b
g
r

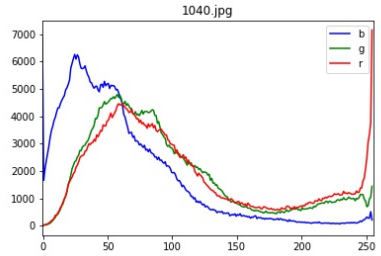

1035.jpg
b
g
r

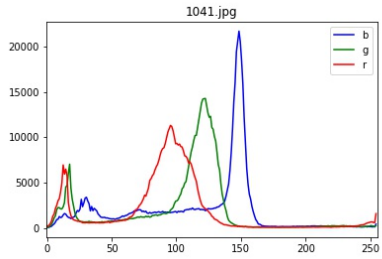

1036.jpg
b
g
r

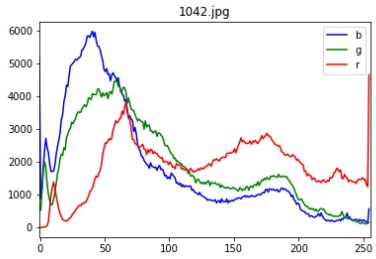

1037.jpg
b
g
r

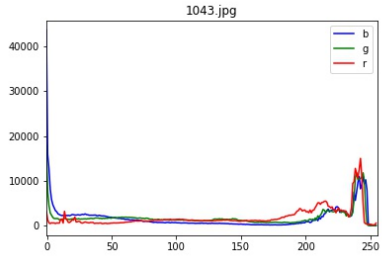

1038.jpg
b
g
r

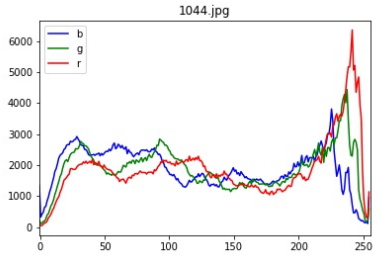

1039.jpg
b
g
r

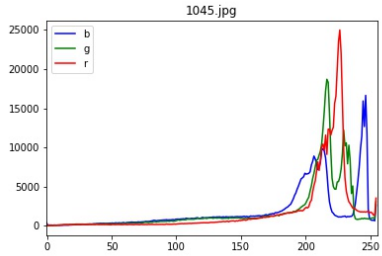

1040.jpg
b
g
r

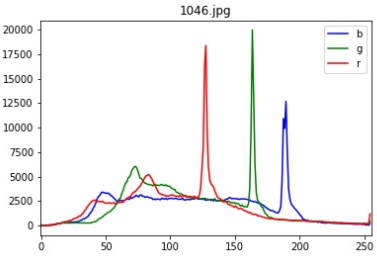

1041.jpg
b
g
r

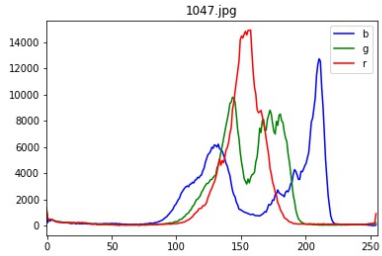

1042.jpg
b
g
r

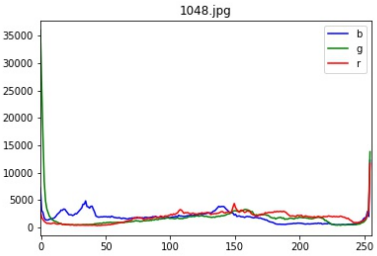

1043.jpg
b
g
r

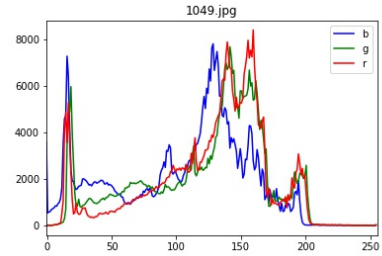

1044.jpg
b
g
r

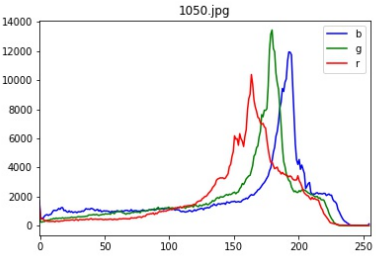

1045.jpg
b
g
r


1046.jpg
b
g
r

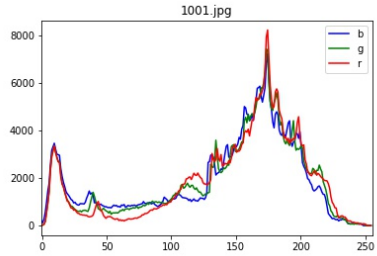

1047.jpg
b
g
r

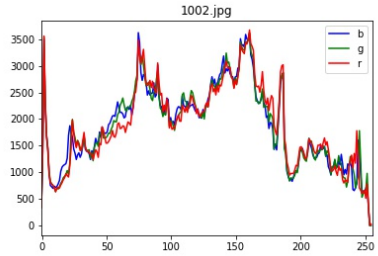

1048.jpg
b
g
r

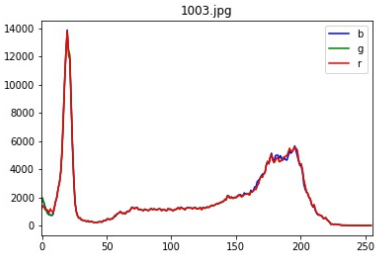

1049.jpg
b
g
r

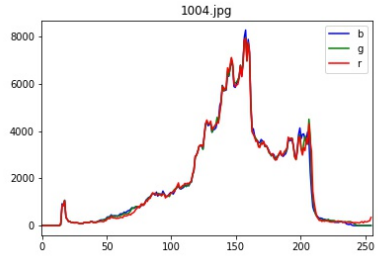

1050.jpg
b
g
r

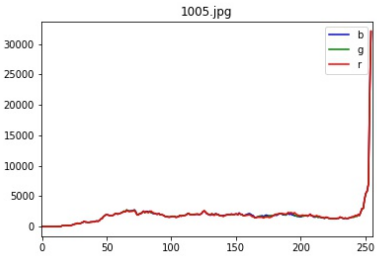
Filtered Images

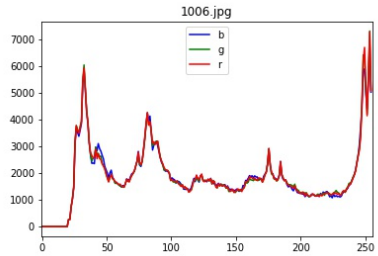


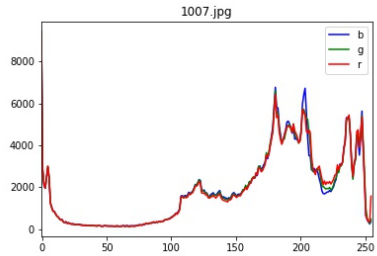


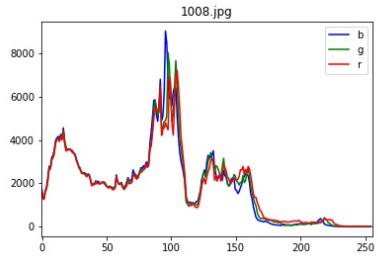


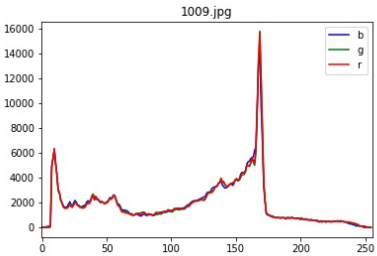


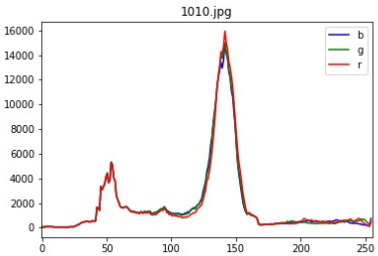


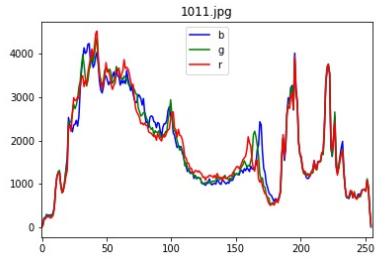


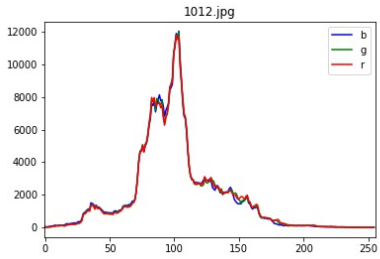


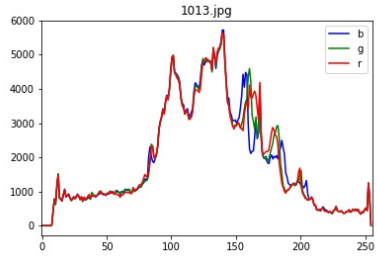


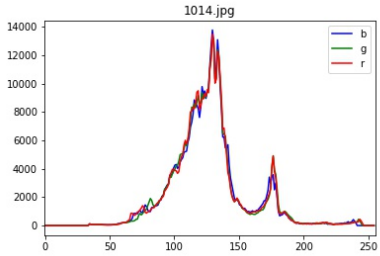


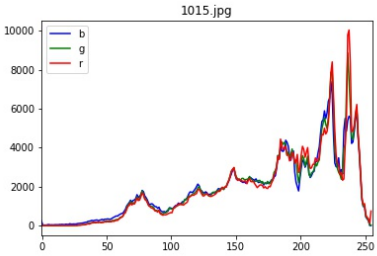


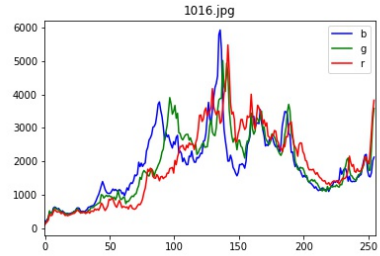

1011.jpg
b
g
r

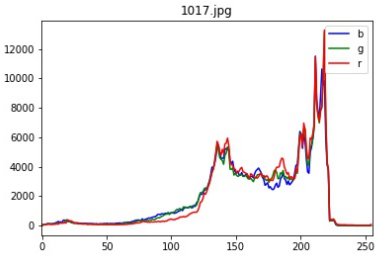

1012.jpg
b
g
r

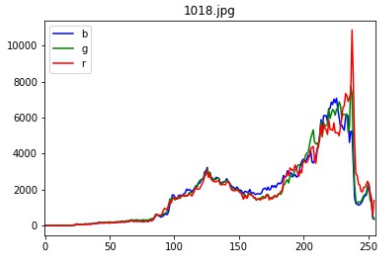

1013.jpg
b
g
r

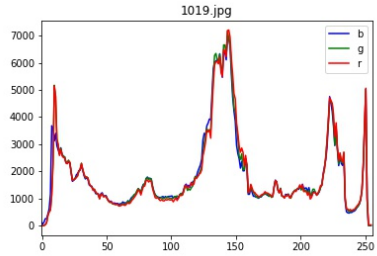

1014.jpg
b
g
r

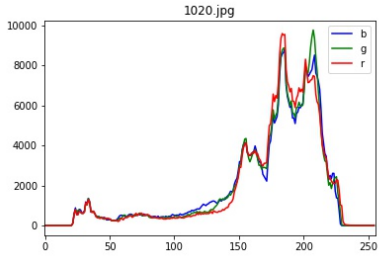

1015.jpg
b
g
r

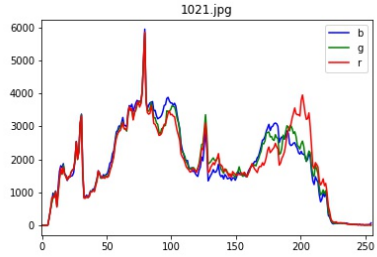

1016.jpg
b
g
r

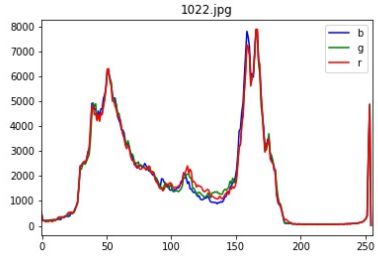

1017.jpg
b
g
r

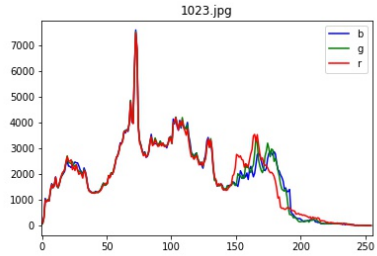

1018.jpg
b
g
r

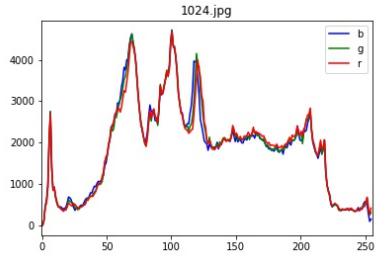

1019.jpg
b
g
r

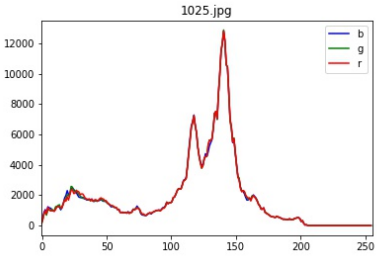

1020.jpg
b
g
r

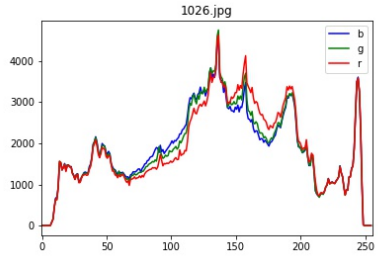

1021.jpg
b
g
r

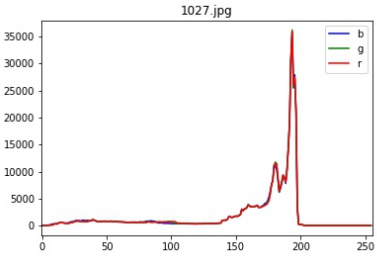

1022.jpg
b
g
r

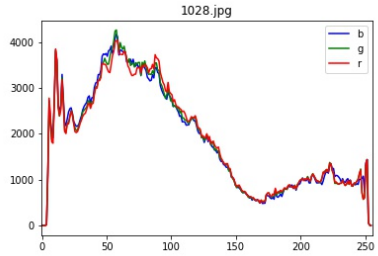

1023.jpg
b
g
r

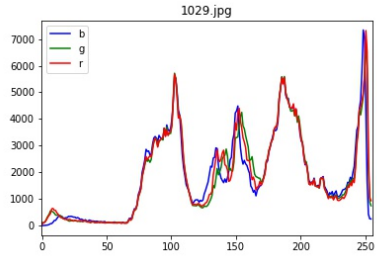

1024.jpg
b
g
r

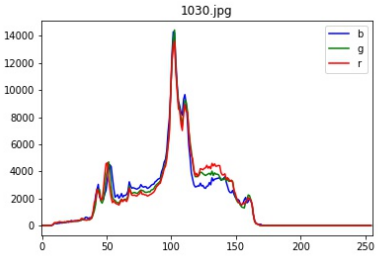

1025.jpg
b
g
r

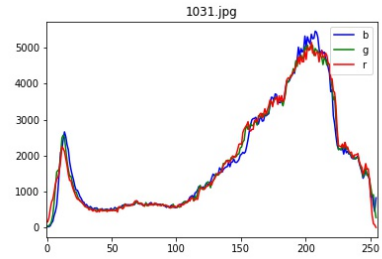

1026.jpg
b
g
r

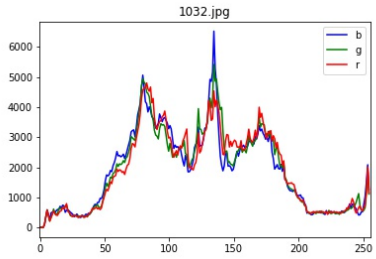

1027.jpg
b
g
r

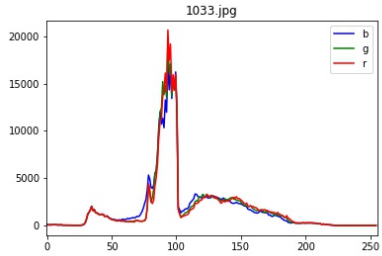

1028.jpg
b
g
r

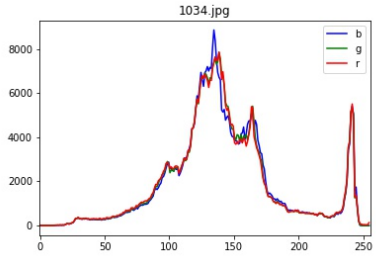

1029.jpg
b
g
r

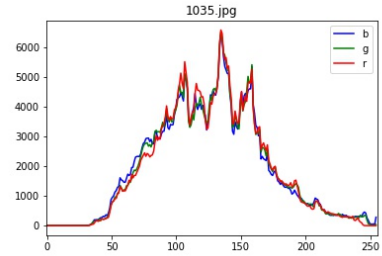

1030.jpg
b
g
r

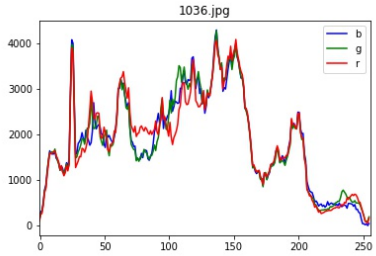

1031.jpg
b
g
r

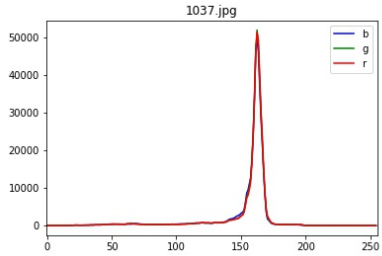

1032.jpg
b
g
r

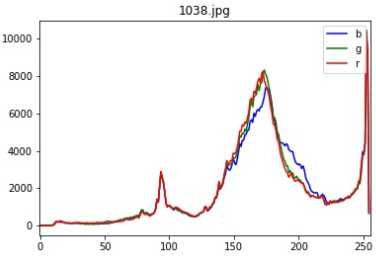

1033.jpg
b
g
r

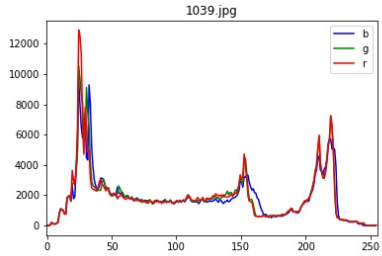

1034.jpg
b
g
r

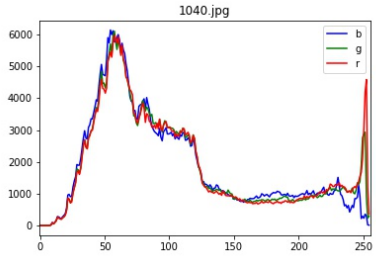

1035.jpg
b
g
r

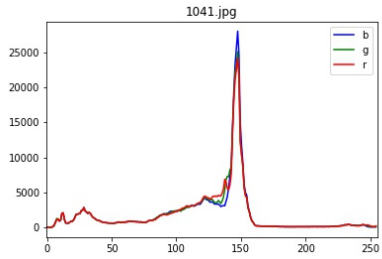

1036.jpg
b
g
r

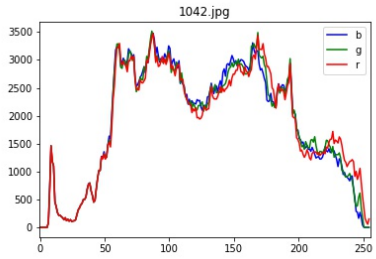

1037.jpg
b
g
r

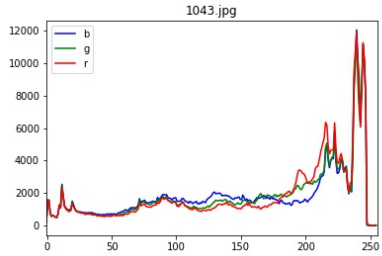

1038.jpg
b
g
r

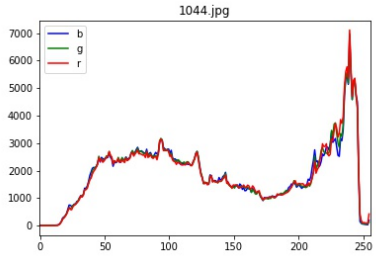

1039.jpg
b
g
r

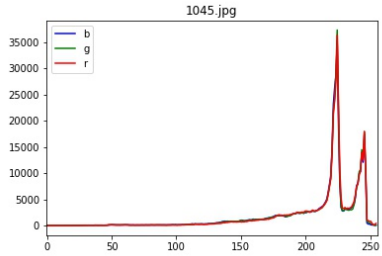

1040.jpg
b
g
r

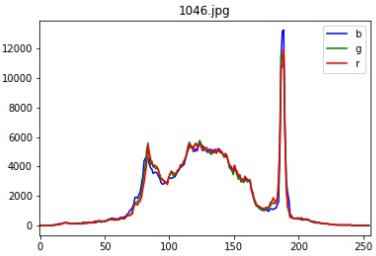

1041.jpg
b
g
r

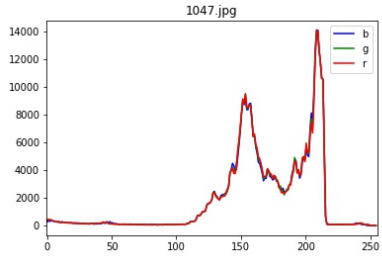

1042.jpg
b
g
r

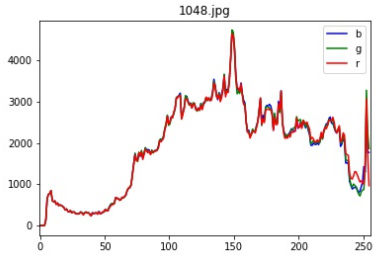

1043.jpg
b
g
r

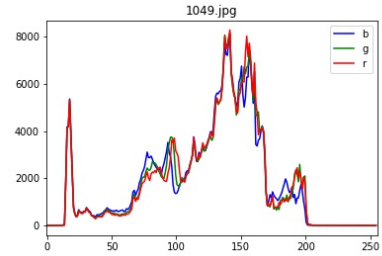

1044.jpg
b
g
r

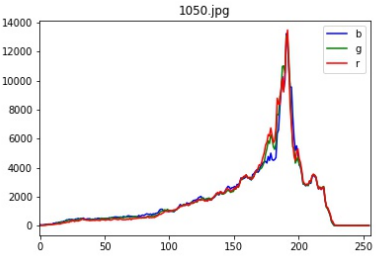

1045.jpg
b
g
r


1046.jpg
b
g
r


1047.jpg
b
g
r


1048.jpg
b
g
r


1049.jpg
b
g
r


1050.jpg
b
g
r

A.7 Spatial Frequency

Below is the table of spatial frequency values for the 50 baseline images and 50 high-fidelity filtered images. The table contains columns: image, condition (baseline or filtered), and spatial frequency.

Table A.3: The values of spatial frequency for each image.

Image	Condition	Spatial Frequency	Condition	Spatial Frequency
1001	Baseline	15.09762885	Filtered	5.085821272
1002	Baseline	24.39110785	Filtered	5.899361077
1003	Baseline	29.67346451	Filtered	8.840743197
1004	Baseline	20.85045588	Filtered	3.984710449
1005	Baseline	13.03251901	Filtered	4.356590102
1006	Baseline	26.46944582	Filtered	4.113808809
1007	Baseline	13.63919072	Filtered	6.184703596
1008	Baseline	11.50597016	Filtered	4.843984433
1009	Baseline	10.49489791	Filtered	2.923714224
1010	Baseline	30.97627048	Filtered	9.208264806
1011	Baseline	16.12293383	Filtered	6.603932856
1012	Baseline	13.37566004	Filtered	3.348118495
1013	Baseline	26.96556543	Filtered	6.54594264
1014	Baseline	11.23184952	Filtered	1.995500668
1015	Baseline	17.26870406	Filtered	7.204234432
1016	Baseline	13.58215804	Filtered	7.577510707
1017	Baseline	27.31088402	Filtered	5.724526848
1018	Baseline	20.75551771	Filtered	8.289646027
1019	Baseline	10.77108162	Filtered	3.20242985
1020	Baseline	9.805154576	Filtered	5.711363713

Image	Condition	Spatial Frequency	Condition	Spatial Frequency
1021	Baseline	13.66635117	Filtered	6.171241096
1022	Baseline	18.98557517	Filtered	6.437865829
1023	Baseline	15.65297568	Filtered	6.551064128
1024	Baseline	21.2404229	Filtered	6.431888061
1025	Baseline	11.07376624	Filtered	2.593826761
1026	Baseline	15.32367968	Filtered	5.7923719
1027	Baseline	15.78562532	Filtered	5.352051138
1028	Baseline	14.18594414	Filtered	3.705187836
1029	Baseline	18.63805984	Filtered	7.992023092
1030	Baseline	7.716034681	Filtered	3.745354468
1031	Baseline	45.07277978	Filtered	7.786516197
1032	Baseline	13.19376508	Filtered	4.846766868
1033	Baseline	28.11616871	Filtered	5.411309723
1034	Baseline	16.49264236	Filtered	4.257440736
1035	Baseline	13.32981774	Filtered	5.96700228
1036	Baseline	22.30018383	Filtered	10.44835698
1037	Baseline	9.92770815	Filtered	2.941110035
1038	Baseline	18.91220276	Filtered	6.377872755
1039	Baseline	13.58875619	Filtered	5.6492571
1040	Baseline	15.50628651	Filtered	4.993070545

Image	Condition	Spatial Frequency	Condition	Spatial Frequency
1041	Baseline	7.039302726	Filtered	3.6670906
1042	Baseline	18.76036648	Filtered	3.467272617
1043	Baseline	6.784134644	Filtered	3.387211755
1044	Baseline	19.85873144	Filtered	8.549660988
1045	Baseline	21.57023676	Filtered	7.997155339
1046	Baseline	23.79882037	Filtered	4.712705418
1047	Baseline	10.13458024	Filtered	4.979187703
1048	Baseline	14.10800957	Filtered	4.087902347
1049	Baseline	13.92318574	Filtered	5.770539644
1050	Baseline	15.84927733	Filtered	7.388253753

A.8 CNN Architecture

Below is the architecture of CNN model employed in this work. The visualization shows the details of each layer, and input and output sizes in each layer. This figure can be generated using Keras library.

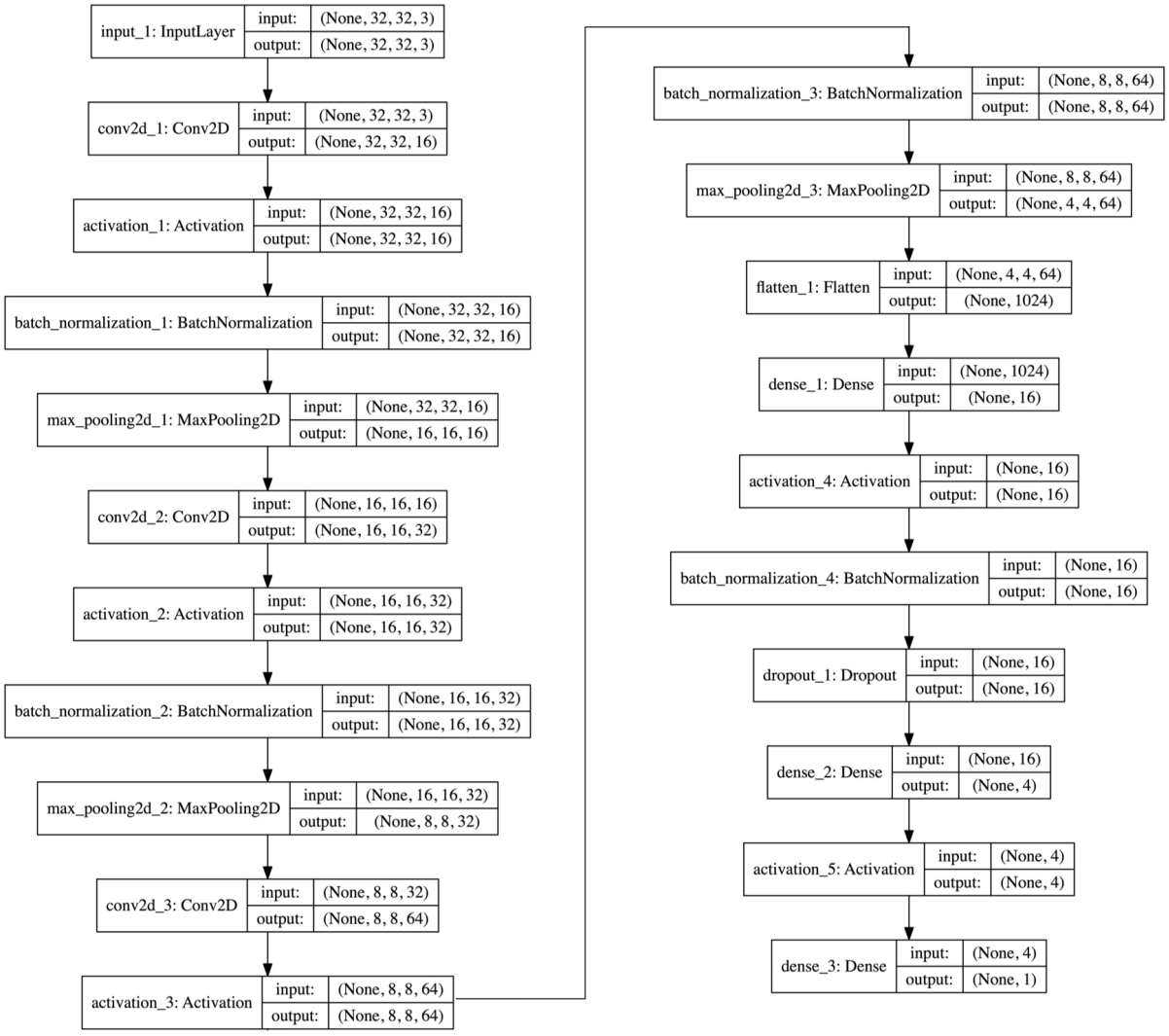


Figure A.1: The detailed CNN architecture that is employed in section 5.5.2.

B Sample Code

B.1 Luminance

```
import matplotlib.pyplot as plt
import cv2

def read_image_as_float32_hls(image_path):
    image = cv2.imread(image_path)
    hls_uint8 = cv2.cvtColor(image, cv2.COLOR_RGB2HLS)
    return ( hls_uint8 / 255.0 ).astype(np.float32)

def luminance_img_hist(image_path, output_file, color):
    h, l, s = cv2.split(read_image_as_float32_hls(image_path))
    plt.hist(l.flatten(), bins='auto', color=color)
    plt.title(output_file)
    plt.savefig(output_file)
    plt.show()

def luminance_two_img_hist_line(img, filtered_img, output_file, \
color, filtered_color):
    h, l, s = cv2.split(read_image_as_float32_hls(img))
    h_filtered, l_filtered, s_filtered = \
        cv2.split(read_image_as_float32_hls(filtered_img))
    plt.hist(l.flatten(), bins='auto', color='white', edgecolor=color)
    plt.hist(l_filtered.flatten(), bins='auto', color='white', \
        edgecolor=filtered_color)
```

```
    plt.title(output_file)
    plt.legend(['osie', 'filtered'])
    plt.savefig(output_file)
    plt.show()
```

B.2 Chroma

```
def avg_two_img_hist_line(img, filtered_img, output_file, \
color, filtered_color):
    im = cv2.imread(img)
    # calculate mean value from RGB channels and flatten to 1D array
    vals = im.mean(axis=2).flatten()
    counts, bins = np.histogram(vals, range(257))
    # plot histogram centered on values 0..255
    plt.plot(bins[:-1] - 0.5, counts, color=color)

    filtered_im = cv2.imread(filtered_img)
    # calculate mean value from RGB channels and flatten to 1D array
    vals = filtered_im.mean(axis=2).flatten()
    counts, bins = np.histogram(vals, range(257))
    # plot histogram centered on values 0..255
    plt.plot(bins[:-1] - 0.5, counts, color=filtered_color)

    plt.xlim([-0.5, 255.5])
    plt.title(output_file[output_file.rfind('/')+1:])
    plt.legend(['osie', 'filtered'])
    plt.savefig(output_file)
```

```
    plt.show()



def img_hist_line_channel_rgb(img, output_file):
    im = cv2.imread(img)
    colors = {0:'b', 1:'g', 2:'r'}
    for channel in range(3):
        vals = im[:,:,channel]
        counts, bins = np.histogram(vals, range(257))
        plt.plot(bins[:-1] - 0.5, counts, color=colors[channel])
    plt.xlim([-0.5, 255.5])
    plt.title(output_file[output_file.rfind('/')+1:])
    plt.legend(colors.values())
    plt.savefig(output_file)
    plt.show()
```

B.3 Spatial Frequency

```
def row_frequency(image):
    im = cv2.imread(image)
    M = im.shape[0]
    N = im.shape[1]
    rf_mean = 0.0
    rf_blue = 0.0
    rf_green = 0.0
    rf_red = 0.0
    for m in range(M):
```

```
        for n in range(1,N):
            rf_mean = rf_mean + (im[m,n,:].mean() - im[m,n-1,:].mean())**2
            rf_blue = rf_blue + (float(im[m,n,0]) - float(im[m,n-1,0]))**2
            rf_green = rf_green + (float(im[m,n,1]) - float(im[m,n-1,1]))**2
            rf_red = rf_red + (float(im[m,n,2]) - float(im[m,n-1,2]))**2
    rf_mean = math.sqrt(rf_mean/(M*N))
    rf_blue = math.sqrt(rf_blue/(M*N))
    rf_green = math.sqrt(rf_green/(M*N))
    rf_red = math.sqrt(rf_red/(M*N))
    return [rf_mean, rf_blue, rf_green, rf_red]

def column_frequency(image):
    im = cv2.imread(image)
    M = im.shape[0]
    N = im.shape[1]
    cf_mean = 0.0
    cf_blue = 0.0
    cf_green = 0.0
    cf_red = 0.0
    for n in range(N):
        for m in range(1,M):
            cf_mean = cf_mean + (im[m,n,:].mean() - im[m-1,n,:].mean())**2
            cf_blue = cf_blue + (float(im[m,n,0]) - float(im[m-1,n,0]))**2
            cf_green = cf_green + (float(im[m,n,1]) - float(im[m-1,n,1]))**2
            cf_red = cf_red + (float(im[m,n,2]) - float(im[m-1,n,2]))**2

    cf_mean = math.sqrt(cf_mean/(M*N))
```

```
    cf_blue = math.sqrt(cf_blue/(M*N))
    cf_green = math.sqrt(cf_green/(M*N))
    cf_red = math.sqrt(cf_red/(M*N))
    return [cf_mean, cf_blue, cf_green, cf_red]

def spatial_frequency(row_frequency, column_frequency):
    return math.sqrt(row_frequency**2 + col_frequency**2)
```

B.4 Regression Analysis

```
import statsmodels.api as sm
import pandas as pd
from sklearn import preprocessing

file = '../study_b/hit_count_luminance_rgb_r_g_b_spatial_frequency_extra.csv'
df = pd.read_csv(file)
cols = ['mean_luminance', 'mean_spatial_frequency']
X = df[cols]
standardized_X = preprocessing.scale(X)
standardized_X = pd.DataFrame.from_records(standardized_X, columns=cols)
y = df['hit_count']
X2 = sm.add_constant(standardized_X)
est = sm.OLS(y, X2)
est2 = est.fit()
print(est2.summary())
```