

PRIVACY-PRESERVING FACE REDACTION USING CROWDSOURCING

by

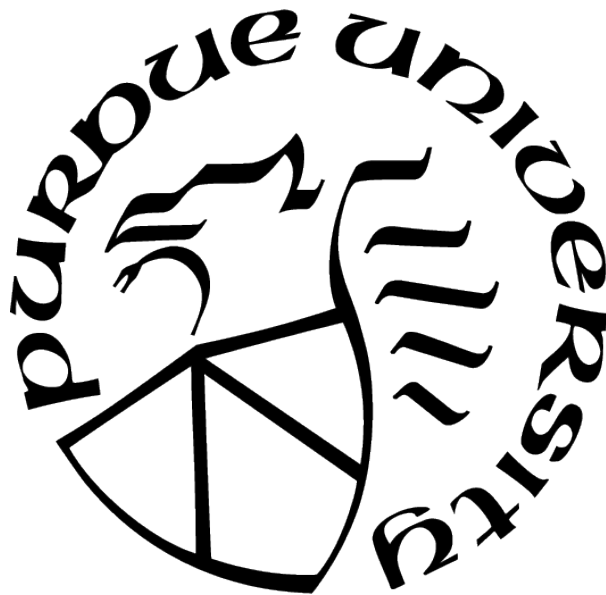
Abdullah Bader Alshaibani

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



School of Electrical and Computer Engineering

West Lafayette, Indiana

August 2021

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Alexander J. Quinn, Chair

School of Electrical and Computer Engineering

Dr. Edward J. Delp

School of Electrical and Computer Engineering

Dr. David S. Ebert

School of Electrical and Computer Engineering

Dr. Amy R. Reibman

School of Electrical and Computer Engineering

Approved by:

Dr. Dimitrios Peroulis

I dedicate this dissertation to my wife Fanan and our daughters Fay and Farah, and everyone who stood by us during this journey. This degree would not have been possible without them.

ACKNOWLEDGMENTS

I want to thank my family, who supported me during my entire academic endeavor. My older brother for constantly pushing me to be the best. My younger brother for continually pushing me back up. My sisters for their never-ending support. My parents never gave up and kept me going forward. And my wife for being my rock and voice of reason.

I would also like to thank my research advisor Alex Quinn for his guidance and support. Finally, I would like to thank Kuwait University for funding my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES	11
LIST OF FIGURES	12
ABBREVIATIONS	19
GLOSSARY	20
ABSTRACT	21
1 INTRODUCTION	22
1.1 Thesis Statement	24
1.2 Research Questions	24
1.3 Contributions	24
2 RELATED WORK	26
2.1 Motivating applications	26
2.2 Privacy-preserving crowdsourcing	26
2.3 Design and technical foundations	28
2.4 Human Perception of Faces	28
2.5 Face Redaction	29
2.6 Automated Face Detection	29
3 THE INTOFOCUS METHOD	30
3.1 Parameters	31
3.1.1 Filter Method	33

3.2	System Precautions	34
3.2.1	Attention Check	35
3.2.2	Collusion Prevention	35
3.3	Experiments Setup	36
3.3.1	Treatment Condition (IntoFocus)	36
3.3.2	Control Conditions (naïve)	36
3.3.3	Attention Check	37
3.3.4	Face Selection	37
3.3.5	Participants	37
3.4	Experiment 1: Actors	37
3.4.1	Dataset	38
3.4.2	Bonus	39
3.4.3	Image Presentation	39
3.4.4	Experiment Task	40
3.4.5	Evaluation	40
3.4.6	IntoFocus vs. Control	41
3.4.7	Attention Check	41
3.4.8	Results	41
3.5	Experiment 2: Random People	43
3.5.1	Dataset	44

3.5.2	Bonus	44
3.5.3	Image Presentation	44
3.5.4	Experiment Task	45
3.5.5	Evaluation	45
3.5.6	IntoFocus vs. Control	46
3.5.7	Results	47
3.6	Discussion	47
3.6.1	Future work	48
3.7	Conclusion	49
4	HUMAN PERCEPTION OF MEDIAN FILTERED FACES	50
4.1	Face Perception Studies	51
4.1.1	Rationale Behind the Experiment	52
4.1.2	Study Setup	52
4.1.3	Study Part 1: Detection	52
4.1.4	Study Part 2: Identification	53
4.1.5	Dataset	54
4.1.6	Evaluation	54
4.1.7	Results	56
4.2	Evaluating The Filter Levels	57
4.2.1	Experiments Setup	57

4.2.2	Treatment (IntoFocus)	57
4.2.3	Attention Check	58
4.2.4	Face Selection	58
4.2.5	Participants	59
4.2.6	Dataset	59
4.2.7	Image Presentation	60
4.2.8	Experiment Task	60
4.2.9	Evaluation	60
4.2.10	Results	60
4.3	Discussion and Future Work	61
5	THE PTERODACTYL SYSTEM	67
5.1	Introduction	67
5.2	Related Work	68
5.3	The Pterodactyl System	68
5.4	AdaptiveFocus Median Filter	69
5.4.1	Image Requirements	70
5.4.2	Finding The Face Locations	71
5.4.3	Selecting The Filter Levels	71
5.5	Experiment Setup	75
5.5.1	AdaptiveFocus Filters	75

5.5.2	Control Conditions	76
5.5.3	Face Identification	77
5.5.4	Face Detection	77
5.5.5	Participants	77
5.5.6	Attention Check	78
5.5.7	Dataset	78
5.5.8	Experiment Task	78
5.5.9	Evaluation	79
5.6	Results	79
5.6.1	Detection	80
5.6.2	Identification	81
5.6.3	Time	81
5.6.4	Cost	82
5.6.5	System Analysis	83
	Detection	83
	Identification	84
5.7	Discussion	85
5.8	Conclusion	87
6	CONCLUSION	88
6.1	IntoFocus	88

6.2	Human Perception of Median Filtered Faces	89
6.3	The AdaptiveFocus Filter And The Pterodactyl System	90
7	FUTURE WORK	92
	REFERENCES	95
A	EXPERIMENT IMAGES	106
A.1	IntoFocus Experiment 1 images	106
A.2	IntoFocus Experiment 2 images	130
A.3	The Human Perception Experiment Images	143
A.4	The Improved IntoFocus Experiment Images	151
A.5	The Pterodactyl Experiment Images	167
B	FACES	202
C	INTERFACES	214
C.1	IntoFocus Experiment Interfaces	214
C.2	Human Perception Study Experiment Interfaces	216
C.3	Pterodactyl System Evaluation Interface	220
D	CONSENT FORMS	221
	VITA	227

LIST OF TABLES

3.1	The number of times the conditions succeeded in performing the tasks of redaction and identity preservation. In the faces redacted row, higher is better. In the faces identified row, lower is better.	48
-----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

LIST OF FIGURES

3.1	The diagram shows the entire flow of the system with five stages and a minimum of 3 workers for a single stage. The top left image is what the requester sent to be redacted. The green blobs are the worker's highlights. The image at the top right is the resulting image from the redaction process.	32
3.2	This figure shows the interface used for experiment 1 with actors. The workers were required to select a face and answer if they could find a face in the image. The highlighting is optional because of the possibility of them not being able to see any faces. They were also required to select how they were able to identify the faces.	38
3.3	This figure shows the results of the experiment. IntoFocus was significantly better than the others because it combines the flexibility of low $ksize$ value with the preservation of higher $ksize$ value. With faces detected at each redacted at a higher filter level, the IntoFocus method can have an identification rate lower than that of the highest filter because faces detected at each stage are redacted at a higher filter level than that stage.	42
3.4	This figure shows the reasons the workers provided when they were able to correctly identify a face for all the conditions in experiment 1. A trend can be seen that as the $ksize$ value is reduced, the number of workers that were able to identify a face because it was visible increases. In the IntoFocus method, the face being visible was not a concern. Since these are the percentages, we can see that the highest identification rate happened for the IntoFocus method is when the workers were familiar with the scene where the image was taken.	43
3.5	This figure shows the interface used for experiment 2 with random people. The workers were required to select a face and answer if they were able to find a face in the image. The highlighting is optional because of the possibility of them not being able to see any faces.	45
3.6	This figure shows the percentages of faces redacted in experiment 2. IntoFocus was significantly better than the others because it combines the flexibility of low $ksize$ value with the preservation of higher $ksize$ value.	46
4.1	Participants were required to redact the entire face to the top of the hairline, including ears, facial hair, hats, and hair covers. This is an example of the expected face redactions.	51

4.2	Face detection study. Participants annotated each face they detected with an ellipse. They could select the ‘No face found’ button if they are unable to detect any faces. The study was performed on AMT, with each HIT containing five images. One of the images was used for the attention check. That image would have a reduced filter level to make for easy detection. The attention check image was changed to a different image in the set at each leaf node in the search tree, to find the appropriate filter levels for all the images in the task.	63
4.3	Face Identification study. Participants selected the reference face (right) matching the person in the main image (left) who was marked with a red ellipse. Multiple faces could be selected if they could eliminate some reference faces from consideration but could not identify the subject face (left) as definitely matching <i>one</i> of the reference faces (right).	64
4.4	Study results. The blue line and points are the second percentile of the detect values (the point right before the detection rate reaches 100%). The red line and points are the 98th percentile of the identify values (the point right before the identification rate reaches 0%). The green line is the staircase model that was used to select the filter levels for the IntoFocus method.	65
4.5	This figure shows the IntoFocus task interface displaying an image in the process of redaction. A subtle difference in the filter level can be seen covering the face of the person on the left. Here, the workers must perform two tasks on five different images. First, they add ellipses on all of the faces in the image. Next, they attempt to select the correct face that matches a person in the image. If they cannot perform the detection task, they must click the <i>No face Found</i> button. If they cannot identify the person in the image, they must select the <i>Any of these</i> or the <i>None of these</i> button.	66
5.1	This figure shows the process an image takes in the AdaptiveFocus filter. Starting with the clear image, it is resized, made square and each tile is separated. Next, a window of size $n \times n$ centered at each tile slides across the image and assigns a filter level to each of the tiles. Finally, an obfuscated image is created.	70
5.2	This figure shows an example of a filter map. The left image shows the image to be redacted where the white lines separate the tiles. The image on the right is the ground truth filter map for that image. The black region means that there are no faces in that region. The brighter regions specify the filter levels required for that face to be detectable but not identifiable. The darker the region, the lower the filter required, and the brighter the region, the higher the filter required.	72
5.3	This figure shows the convolutional neural network used in the AdaptiveFocus filter. It inputs images of size $112 \times 112 \times 3$ and outputs the probability of each of the filter levels being correct based on the content of the image.	73

5.4	This figure shows the different AdaptiveFocus methods. The top left is built based on the AdaptiveFocus (inner face) method, the bottom left is based on the AdaptiveFocus (full-face method), and the right are the two steps for the AdaptiveFocus (two-step) method.	74
5.5	This figure shows the difference between the inner face and full-face images. In the inner face, only the inside of the face is used. In the full-face, the full face is covered including hair, beard, and hats.	75
5.6	This figure shows the interface used to evaluate the Pterodactyl system and the AdaptiveFocus filter. Participants were tasked with detecting all the faces in five filtered images and identifying a face in each of the images.	76
5.7	This figure shows the detection results of all the methods.	79
5.8	This figure shows the identification results of all the methods. The results show that the AdaptiveFocus methods had the lowest face identification rates and the IntoFocus method had a significantly high disclosure rate.	80
5.9	This figure shows the time it takes to redact an image using IntoFocus and AdaptiveFocus. The results show that the methods using the AdaptiveFocus filter are seven times faster than methods using IntoFocus.	81
5.10	This figure shows the cost to redact an image using the IntoFocus method and the AdaptiveFocus methods.	82
5.11	This figure shows the detection results of the crowd-based methods using the Pterodactyl system and the IntoFocus system. The detection rate in the IntoFocus method does not change when the method changes. In contrast, the AdaptiveFocus methods observe a significant increase in detection performance.	83
5.12	This figure shows the identification results of the crowd-based methods using the Pterodactyl system and the IntoFocus system. All the methods observe a significant decrease in disclosure rates.	84
7.1	This image shows a face in the top right behind the main face that was not detectable by the crowd workers in most of the AdaptiveFocus methods.	93
A.1	Grid 1 of the images used in the experiments in chapter 3	107
A.2	Grid 2 of the images used in the experiments in chapter 3	108
A.3	Grid 3 of the images used in the experiments in chapter 3	109
A.4	Grid 4 of the images used in the experiments in chapter 3	110
A.5	Grid 5 of the images used in the experiments in chapter 3	111
A.6	Grid 6 of the images used in the experiments in chapter 3	112
A.7	Grid 7 of the images used in the experiments in chapter 3	113
A.8	Grid 8 of the images used in the experiments in chapter 3	114

A.9	Grid 9 of the images used in the experiments in chapter 3	115
A.10	Grid 10 of the images used in the experiments in chapter 3	116
A.11	Grid 11 of the images used in the experiments in chapter 3	117
A.12	Grid 12 of the images used in the experiments in chapter 3	118
A.13	Grid 13 of the images used in the experiments in chapter 3	119
A.14	Grid 14 of the images used in the experiments in chapter 3	120
A.15	Grid 15 of the images used in the experiments in chapter 3	121
A.16	Grid 16 of the images used in the experiments in chapter 3	122
A.17	Grid 17 of the images used in the experiments in chapter 3	123
A.18	Grid 18 of the images used in the experiments in chapter 3	124
A.19	Grid 19 of the images used in the experiments in chapter 3	125
A.20	Grid 20 of the images used in the experiments in chapter 3	126
A.21	Grid 21 of the images used in the experiments in chapter 3	127
A.22	Grid 22 of the images used in the experiments in chapter 3	128
A.23	Grid 23 of the images used in the experiments in chapter 3	129
A.24	Grid 1 of the images used in the experiments in chapter 3	130
A.25	Grid 2 of the images used in the experiments in chapter 3	131
A.26	Grid 3 of the images used in the experiments in chapter 3	132
A.27	Grid 4 of the images used in the experiments in chapter 3	133
A.28	Grid 5 of the images used in the experiments in chapter 3	134
A.29	Grid 6 of the images used in the experiments in chapter 3	135
A.30	Grid 7 of the images used in the experiments in chapter 3	136
A.31	Grid 8 of the images used in the experiments in chapter 3	137
A.32	Grid 9 of the images used in the experiments in chapter 3	138
A.33	Grid 10 of the images used in the experiments in chapter 3	139
A.34	Grid 11 of the images used in the experiments in chapter 3	140
A.35	Grid 12 of the images used in the experiments in chapter 3	141
A.36	Grid 13 of the images used in the experiments in chapter 3	142
A.37	Grid 1 of the images used in the experiments in chapter 4	143
A.38	Grid 2 of the images used in the experiments in chapter 4	144

A.39 Grid 3 of the images used in the experiments in chapter 4	145
A.40 Grid 4 of the images used in the experiments in chapter 4	146
A.41 Grid 5 of the images used in the experiments in chapter 4	147
A.42 Grid 6 of the images used in the experiments in chapter 4	148
A.43 Grid 7 of the images used in the experiments in chapter 4	149
A.44 Grid 8 of the images used in the experiments in chapter 4	150
A.45 Grid 1 of the images used in the experiments in chapter 4	151
A.46 Grid 2 of the images used in the experiments in chapter 4	152
A.47 Grid 3 of the images used in the experiments in chapter 4	153
A.48 Grid 4 of the images used in the experiments in chapter 4	154
A.49 Grid 5 of the images used in the experiments in chapter 4	155
A.50 Grid 6 of the images used in the experiments in chapter 4	156
A.51 Grid 7 of the images used in the experiments in chapter 4	157
A.52 Grid 8 of the images used in the experiments in chapter 4	158
A.53 Grid 9 of the images used in the experiments in chapter 4	159
A.54 Grid 10 of the images used in the experiments in chapter 4	160
A.55 Grid 11 of the images used in the experiments in chapter 4	161
A.56 Grid 12 of the images used in the experiments in chapter 4	162
A.57 Grid 13 of the images used in the experiments in chapter 4	163
A.58 Grid 14 of the images used in the experiments in chapter 4	164
A.59 Grid 15 of the images used in the experiments in chapter 4	165
A.60 Grid 16 of the images used in the experiments in chapter 4	166
A.61 Grid 1 of the images used in the experiments in chapter 5	167
A.62 Grid 2 of the images used in the experiments in chapter 5	168
A.63 Grid 3 of the images used in the experiments in chapter 5	169
A.64 Grid 4 of the images used in the experiments in chapter 5	170
A.65 Grid 5 of the images used in the experiments in chapter 5	171
A.66 Grid 6 of the images used in the experiments in chapter 5	172
A.67 Grid 7 of the images used in the experiments in chapter 5	173
A.68 Grid 8 of the images used in the experiments in chapter 5	174

A.69	Grid 9 of the images used in the experiments in chapter 5	175
A.70	Grid 10 of the images used in the experiments in chapter 5	176
A.71	Grid 11 of the images used in the experiments in chapter 5	177
A.72	Grid 12 of the images used in the experiments in chapter 5	178
A.73	Grid 13 of the images used in the experiments in chapter 5	179
A.74	Grid 14 of the images used in the experiments in chapter 5	180
A.75	Grid 15 of the images used in the experiments in chapter 5	181
A.76	Grid 16 of the images used in the experiments in chapter 5	182
A.77	Grid 17 of the images used in the experiments in chapter 5	183
A.78	Grid 18 of the images used in the experiments in chapter 5	184
A.79	Grid 19 of the images used in the experiments in chapter 5	185
A.80	Grid 20 of the images used in the experiments in chapter 5	186
A.81	Grid 21 of the images used in the experiments in chapter 5	187
A.82	Grid 22 of the images used in the experiments in chapter 5	188
A.83	Grid 23 of the images used in the experiments in chapter 5	189
A.84	Grid 24 of the images used in the experiments in chapter 5	190
A.85	Grid 25 of the images used in the experiments in chapter 5	191
A.86	Grid 26 of the images used in the experiments in chapter 5	192
A.87	Grid 27 of the images used in the experiments in chapter 5	193
A.88	Grid 28 of the images used in the experiments in chapter 5	194
A.89	Grid 29 of the images used in the experiments in chapter 5	195
A.90	Grid 30 of the images used in the experiments in chapter 5	196
A.91	Grid 31 of the images used in the experiments in chapter 5	197
A.92	Grid 32 of the images used in the experiments in chapter 5	198
A.93	Grid 33 of the images used in the experiments in chapter 5	199
A.94	Grid 34 of the images used in the experiments in chapter 5	200
A.95	Grid 35 of the images used in the experiments in chapter 5	201
B.1	Grid 1 of the faces used in the experiments.	203
B.2	Grid 1 of the faces used in the experiments.	204
B.3	Grid 2 of the faces used in the experiments.	205

B.4	Grid 3 of the faces used in the experiments.	206
B.5	Grid 4 of the faces used in the experiments.	207
B.6	Grid 5 of the faces used in the experiments.	208
B.7	Grid 6 of the faces used in the experiments.	209
B.8	Grid 7 of the faces used in the experiments.	210
B.9	Grid 8 of the faces used in the experiments.	211
B.10	Grid 9 of the faces used in the experiments.	212
B.11	Grid 10 of the faces used in the experiments.	213
C.1	Interface of the IntoFocus experiment one in chapter 3	214
C.2	Interface of the IntoFocus experiment two in chapter 3	215
C.3	Interface of the first perception study in chapter 4	216
C.4	Interface of the second perception study in chapter 4	217
C.5	Interface of the detection aspect of the perception study in chapter 4	218
C.6	Interface of the identification aspect of the perception study in chapter 4	219
C.7	Interface of the experiment to evaluate the Pterodactyl system in chapter 5	220
D.1	The consent form used in the IntoFocus experiments in chapter 3.	222
D.2	The first part of the consent form used in the in lab portion of the perception study experiments in chapter 4.	223
D.3	The second part of the consent form used in the in lab portion of the perception study experiments experiments in chapter 4.	224
D.4	The first part of the consent form used in the online portion of the perception study experiments in chapter 4 and the online experiments in the evaluation in chapter 5.	225
D.5	The second part of the consent form used in the online portion of the perception study experiments experiments in chapter 4 and the online experiments in the evaluation in chapter 5.	226

ABBREVIATIONS

AMT	Amazon Mechanical Turk
HIT	Human Intelligence Task
CNN	Convolutional Neural Network

GLOSSARY

Detection	The ability to find a face in a filtered image
Identification	The ability to correctly match a face in a set of 8 faces

ABSTRACT

Face redaction is used to deidentify images of people. Most approaches depend on face detection, but automated algorithms are still not adequate for sensitive applications in which even one unredacted face could lead to irreversible harm. Human annotators can potentially provide the most accurate detection, but only trusted annotators should be allowed to see the faces of privacy-sensitive applications. Redacting more images than trusted annotators could accommodate requires a new approach. This dissertation leverages the characteristics of human perception of faces in median-filtered images in a human computation algorithm to engage crowd workers to redact faces—without revealing the identities. IntoFocus, a system I developed, permits robust face redaction with probabilistic privacy guarantees. The system’s design builds on an experiment that measured the filter levels and conditions where participants could detect and identify faces. Pterodactyl is a system that focuses on increasing the productivity of crowd-based face redaction systems. It uses the AdaptiveFocus filter, a filter that combines human perception of faces in median filtered images with a convolutional neural network to estimate a median filter level for each region of the image to allow the faces to be detected and prevent them from being identified.

1. INTRODUCTION

With the increase of data privacy, people are becoming more aware of what is happening to their data—especially images—when they leave their devices. These images go through many phases of processing and might reach a point somewhere along the line of falling under the eyes of a person that should not have access to those images. Instead of keeping the images in their pure form, which is humanly recognizable, the images could be left in a state where if humans ever laid eyes on them, they would not harm those depicted in the images. Instead of relying on services that keep their data after processing, people search to keep their data private and make sure they are erased as soon as the required task is finished.

The nature of crowdsourcing entails an “open call” [1], which usually involves sharing the contents of the task openly. For tasks that involve sensitive information, this can create a risk of disclosure. If the sensitive information could be safely redacted, this risk would be mitigated, and workers could proceed with whatever labeling, annotation, or other required work.

Robust preservation of privacy is one of the critical steps toward realizing the potential of crowdsourcing [2] and ultimately reducing the barriers to transferring digital work more smoothly and confidently. Consider the following potential applications:

1. Search images from private (or semi-private) social media accounts for evidence of violence or bullying.
2. Redact a large number of images used in legal proceedings to comply with public disclosure laws.
3. Create derivative works from photos taken by children on a class field trip.

Each of these cases demands that the images not be disclosed publicly.

Face recognition algorithms have been very successful to the point that they have surpassed an average human’s ability to recognize faces [3]. Given the location of a face, the current algorithms have high accuracy, but detecting faces in images is a different problem completely. Automated face detection remains an active research problem with no truly robust solution.

Challenges include occlusions, pose, illumination (low or high), atypical skin tones, skin-colored backgrounds, and weather (rain, snow, haze) [4]. In each of the past six calendar years (2012-2018), over 400 published articles have been published with the term “face detection” in the title (based on searches with Google Scholar). Despite thousands of incremental improvements, even a goal of 95% recall (proportion of faces detected) with 95% precision (proportion of matches that are faces) remains beyond the reach of any current algorithm that we are aware of [4]–[6].

For applications where actual harm could result from accidental disclosure, a 10%—or even 5% risk—would be unacceptable. Since humans and machines have complementary strengths concerning this problem, we envision future hybrid approaches. This paper focuses solely on strategies for engaging humans.

Crowd workers could perform redaction, but the redaction task would also disclose sensitive information. From this apparent conflict comes our research objective is to *engage crowd workers to redact facial identities—without exposing those workers to the sensitive information they are redacting*.

Chapter 3, presents *IntoFocus*, a method and system that engages crowd workers to redact faces from still images. It starts by showing workers a heavily filtered form of the image and asking them to highlight regions containing a specified type of information (e.g., human face). Successive iterations present slightly less filtered images while blocking regions marked as potentially sensitive in prior iterations.

In chapter 4, I take the previously defined *IntoFocus* system and rebuild it using the foundations of a human perception study to guarantee the system’s success within the thresholds of the study. The study showed that to maintain the probabilistic privacy guarantees, the initial number of stages proposed was not sufficient for all the possible face conditions. It also showed that the selected filter levels need to be updated based on the new derived filter levels to maintain those probabilistic privacy guarantees.

Chapter 5 presents improvements to the system as a whole with the addition of new rules to improve the accuracy of the crowd workers. It also presents the *AdaptiveFocus* filter, which uses the data from the perception study to select the appropriate filter level for each of the regions of the images. The filter is based on a Convolutional Neural Network (CNN)

that uses the infliction point when each face’s probability of detection starts to recede from 100%. Using that information, the CNN estimates the appropriate filter to use to allow the face to be detected and at the same time prevent the facial identity from being revealed.

1.1 Thesis Statement

A median filter with variable window size enables faces in still images to be detected—for redaction—by untrusted crowd workers with higher accuracy than current machine learning face detection systems.

1.2 Research Questions

This dissertation addresses the following research questions:

- RQ1 Is it feasible to use crowd workers to redact faces in thoroughly obfuscated images and achieve higher face detection rates than automated methods?
- RQ2 Does the human ability for face detection in filtered images differ from one person to another? Is there a filter level where that difference no longer exists?
- RQ3 Can the human’s ability for face detection in filtered images compares with automated machine learning-based methods?
- RQ4 Can the people’s identities in the images be preserved when using a crowd-based system combined with basic protections?

1.3 Contributions

The primary contributions of this dissertation are as follows:

1. A human perception study to quantify the human’s ability to detect and identify faces in median filtered images. The model enables us to estimate the filter levels required for a face to be detectable and unidentifiable.

2. The IntoFocus system and method allow consistent privacy-preserving redaction of images by crowd workers. The system uses the filter levels and extracts a set that reduces identification but increases the detection of faces.
3. Our system demonstrates how to implement the IntoFocus method and one possible interface design.
4. Our implementation’s two experiments validate that the IntoFocus method enables consistent privacy-preserving redaction of images by crowd workers.
5. An experiment to quantify the difference between face detection and face identification in humans.
6. The Pterodactyl system focuses on improving the quality of the detections of crowd workers and reducing the probability of facial identification.
7. The AdaptiveFocus filter reduces the necessary crowd work of the IntoFocus method by 86%.

2. RELATED WORK

Portions of this chapter were copied verbatim from a paper published in the 8th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2020) [7].

The foundations of this work can be understood in terms of (1) motivating applications, (2) privacy-preserving crowdsourcing, (3) design and technical foundations, (4) human perception of faces and face redaction.

2.1 Motivating applications

Crowdsourcing and human computation began to gain prominence in 2005 in research [8] and with the founding of influential commercial services, such as Mechanical Turk. Initial applications were limited to data that could be shared publicly. One of the first published references to the need for privacy—for the requester’s data—was in 2010, regarding to document processing workflows [9].

The risk to humans became palpable with VizWiz, a mobile application that allows blind people to get help with everyday situations by sending a photo and a spoken question to workers on Amazon Mechanical Turk [10]. VizWiz holds the risk of sharing sensitive information inadvertently included in the picture. A similar dilemma exists when crowd workers assist robots. Sorokin used such an approach to enable robots to grasp unfamiliar objects [11]. The robot sends images of the object to workers who draw contours to help the robot grasp it. The IntoFocus *method* could someday be integrated into such systems to enable robust redaction of sensitive content before presenting the image to the workers who will render the assistance.

2.2 Privacy-preserving crowdsourcing

WearMail [12] introduced the use of a system that allows workers to search through a person’s email to answer a specific question that they have. They implemented privacy mechanisms that allow the requester to blacklist specific words and hide them from the

workers. Work in fine-grained categorization [13] used blurred images of birds where workers are allowed to reveal small regions of the image that would help them accurately categorizing the type of bird they are seeing without revealing the entire image.

Lasecki and collaborators have been the most active in developing methods for privacy-preserving crowdsourcing that we are aware of. One application engages crowd workers for behavioral coding of video (e.g., social science research) [14]. Their CrowdMask system is the most similar to IntoFocus that we are aware of, with respect to the purpose.

CrowdMask [15], [16] segments a single image into smaller segments and asks the workers to annotate segments containing sensitive information or adjacent to sensitive information. It uses a pyramid workflow, which is adequate for tasks where judgment about a particular segment can be made based on local information. However, because workers do not see the full photo, they might not be able to judge if a specific region contains private information (e.g., because it is cut in half or is otherwise taken out of context) and it does not account for the risk of having all the information in a single segment. In contrast, IntoFocus shows the entire image but uses gradual revelation to ensure that sensitive regions are not disclosed. In a follow-up, Lasecki et al. used Gaussian blur in a single layer, documented that behaviors can be identified even when a video is blurred sufficiently to hide identities [17].

Lasecki et al. [18] have demonstrated the risks of completely trusting crowd workers with sensitive information. They showed that when some workers were given some incentives would sabotage a task. They also showed that there are other workers that would not let such things happen, who went out of their way to report what was happening. A few recent efforts have proposed methods for addressing this challenge for image-oriented tasks.

One of the first involved a protocol, for instance-privacy based on clipping regions. It used a clipping function based on additional feedback provided by the requester [19]. The need for requester involvement was a limitation, and its “instance-clipping protocol” was not a comprehensive solution.

2.3 Design and technical foundations

Peekaboom [20] introduced the combination of crowd workers and object detection and identification. They used two workers; one tasked to reveal portions of an image and another tasked with identifying what is in the image. The work shows that even with limited revelation, humans are still able to identify objects. With the revelation of specific regions, humans can find or identify the objects in the images. We use this information to build a system that, given a highly filtered image, slowly reveals safe regions to help humans in finding the regions we are trying to hide.

Das et al. [21] showed that humans focus on different regions in images when finding specific objects than deep networks. They found that when they ask a human to search for an object in an image, they search in different regions than what deep networks search in. They also found that when deep networks are programmed to search in the regions that the humans focus on The deep networks had better performance. Their research shows that humans have a better understanding of images and the physical world than deep networks and that if these algorithms are not programmed to check those regions, they would not outperform humans.

Efforts to enhance image segmentation have included strategies that ask human workers to annotate objects in the foreground via various interactions [22]. Our focus is to redact the faces in an image before submitting it to crowdsourcing platform to solve the task, whether the face is in the foreground or the background. So instead of the workers having a clear image of the object/subject being segmented, they would have a redacted image.

2.4 Human Perception of Faces

Lewis et al. [23] experimented to find "what affects a human's perception of faces". Their experiments demonstrated that when people look at an image containing a face, they take less time to locate the face when they can see the body. This shows that people search the entire scene when looking for faces.

While making sense of an image had the effect of enhancing people's face detection abilities, identifying people of different races had the opposite effect [24], [25]. Results show

that recognition memory is better for faces of the same race as the participant. Showing that skin tone is a factor that needs to be considered in the face perception experiments.

These observations show that humans do not focus on facial features alone to detect faces like face detection algorithms that use Eigenfaces (Principal Component Analysis based) [26]–[28], support-vector machines [29], [30], facial skin color-based detection [31], neural network-based detection [32], [33]. This does not show that these methods are flawed or wrong. It shows that there are other factors in face detection that these methods did not use.

2.5 Face Redaction

Some methods that apply several people’s features on top of each other to hide a person’s facial features [34]–[36]. Jourabloo et al. [37] advanced their work by applying weights to specific images to influence the direction of the change. These methods relied on facial images to apply the de-identification method.

Other approaches used segmentation or contouring to detect people in images and then redacted the people [38]–[40]. These methods have the flaw that if the system could not correctly separate the human from the background, the method would not work correctly.

2.6 Automated Face Detection

Face detection has been an ongoing research problem for many years [32], [41], [42]. However, real progress has not been achieved until Viola-Jones[43]. The method combines simpler classifiers to increase the speeds of detection.

There has been significant progress in face detection in recent years[33], [44]–[49]. With the help of face detection datasets [5], [50]–[52], and many others, face detection is reaching a 90% success rate. Nevertheless, even with the availability of face images and large datasets, face detection in natural settings has not been solved yet.

3. THE INTOFOCUS METHOD

To safely leverage crowd work for face redaction, IntoFocus uses an algorithm based on progressive image clarification (i.e., presenting the image to workers with decreasing filter levels).

The input is an image containing any number of faces at any scale, assuming that the scale of possible faces is unknown and that machine detection might fail for some of them. (Machine face detection was not integrated to enable precise measurements of using this technique). The output will be the same image with all faces redacted.

The process proceeds iteratively:

Stage 1: In this first step, giant faces are redacted. The image is filtered using a median filter with a $ksize$ (kernel size) adequate to render typical faces of any size unrecognizable to humans. At this level, only giant faces (i.e., occupying the entire image) will be perceivable as faces. For 640×480 images, starting with a kernel size of 41×41 . That kernel size is referred to as a *filter level* of $k = 41$. These filter levels were selected using a mini-experiment on one of the researchers. The next chapter will discuss further details.

The heavily filtered image is presented to workers, who are asked to annotate all regions of the image that contain any part of a face. They use a brush interface to paint over all portions of the image that contain any part of a face.

To reduce the chance of disclosure, multiple judgments are collected from independent workers. Potential sources of variation include differences in individual perception abilities, inattention, laziness, and malicious subterfuge. Combining the judgments using a union, if any worker identifies a pixel as belonging to a face, it is then recorded as such. In the trials, three judgments were collected per $ksize$ value, but this could be configured to suit a given application’s security and affordability requirements.

Some over-redaction (false positives) is possible. This is considered acceptable based on the premise for IntoFocus. In the target applications, the protection of human identities is a higher priority than the preservation of other content. The implementation does not actively defend against deliberate over-reaction by malicious workers. Still, it can be addressed using

heuristics based on worker behavior or low-level image characteristics (e.g., if a worker flagged a texture such as grass or sand is extremely unlikely to contain a face).

Stage 2: The original clear image is filtered using a median filter with a lower *ksize* value. For 640×640 images, a $ksize = 27$ is used. Regions identified by *any* worker in stage 1 are redacted by filtering with a *higher* *ksize* value. In this implementation, $ksize = 81$ is used for regions marked as faces in stage 1. The second cohort of workers is presented with this image and asked to mark any perceptible faces at this *ksize* value. The interface is the same as before.

Stage $i+1$: With each successive stage, the *ksize* value is decreased, allowing smaller faces to be detected and redacted by the workers. Regions marked as faces in stage $i + 1$ are redacted by filtering with the *ksize* value from stage $i - 1$.

Stage n : The final stage uses a small *ksize* value to deidentify the smallest faces that could otherwise be recognizable to a worker who was familiar with the depicted person. This implementation uses a *ksize* value of 7 in the final stage. Other stages could be added for added protection against the disclosure of tiny—but still recognizable—faces.

Figure 3.1 shows the progress of an image as it goes through the IntoFocus method. The green blobs are the regions that the workers at each stage highlighted. By the end of the fifth stage, the system would release an image with the regions selected by the workers redacted and all the other regions still visible. Now the image can be safely uploaded into a crowdsourcing platform without compromising the people in the image.

Preserving people’s privacy in an image does not end with hiding information in photos regarding a single worker. It needs to hide the information from all the workers. Varshney [53] explored what affects a worker’s reliability and that some workers would collaborate with others to extract the information they needed. Workers can target a task and try to extract information from it. The goal is to hinder their progression and stop those attempts.

3.1 Parameters

This section describes some preliminary explorations that led to our choice of the median filter for the “filter” operation and selecting filter kernel size (*ksize*) used in the IntoFocus

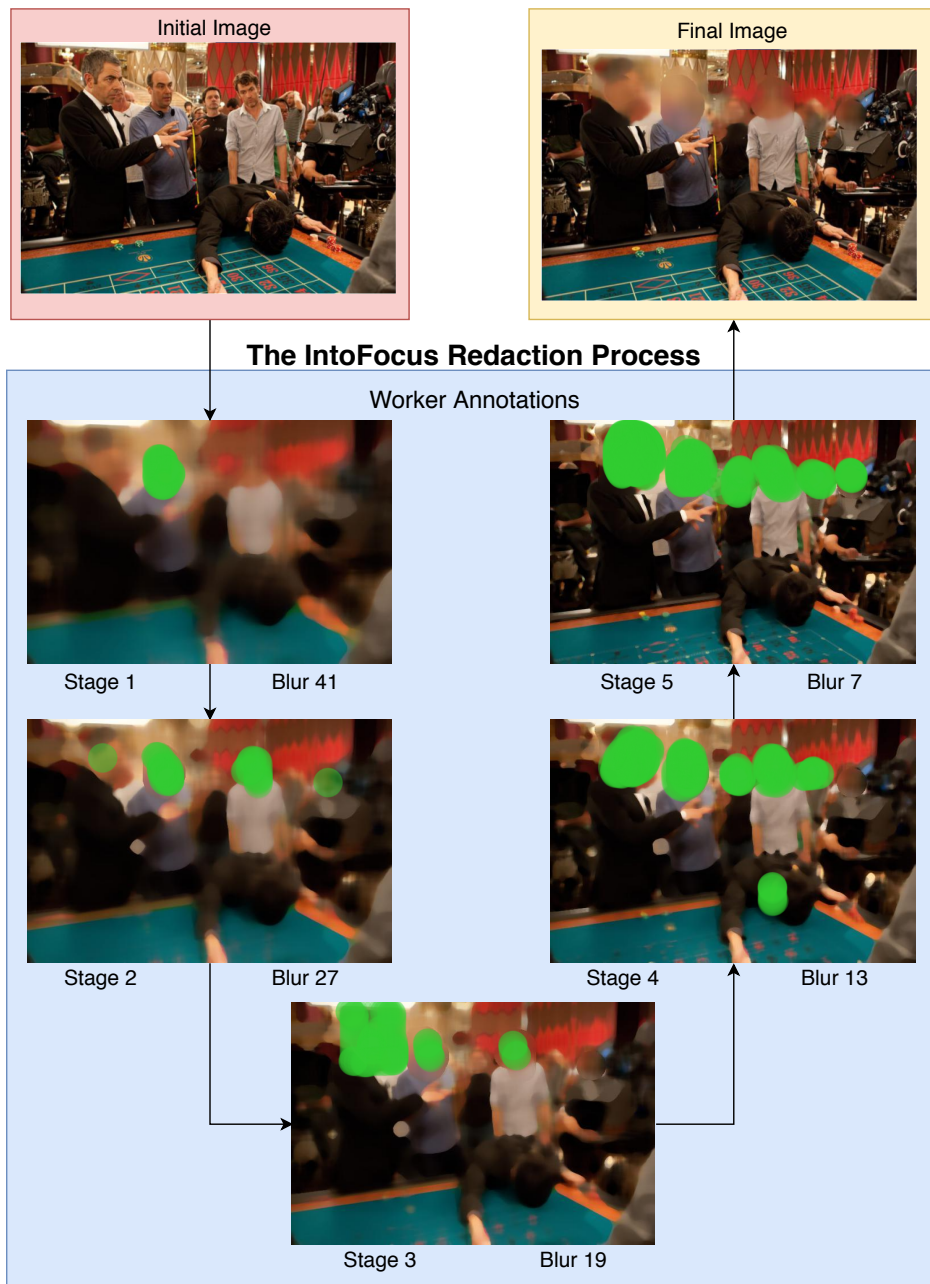


Figure 3.1. The diagram shows the entire flow of the system with five stages and a minimum of 3 workers for a single stage. The top left image is what the requester sent to be redacted. The green blobs are the worker’s highlights. The image at the top right is the resulting image from the redaction process.

algorithm. These values are not “optimal”, but the process for choosing them is explained so that others can understand our design rationale and consider future improvements.

3.1.1 Filter Method

IntoFocus requires a filter operation that reveals enough fidelity to discern a face’s outer contours while concealing smaller features that could be used to recognize the depicted person’s identity (e.g., nose, eyes). Five (5) filters were considered: Gaussian, scatter (sometimes known as “frosted glass”), square pixelation, unfocus, and median. Gaussian and pixelation were eliminated due to known attacks that identify text or faces from obfuscated images [35], [54], [55]. Unfocus was eliminated because it results in qualitatively similar images to gaussian blur, and thus we suspect it may be vulnerable to those attacks.

Scatter—a filter that displaces each pixel by a random distance and in a random direction—results in more significant destruction of information due to its stochasticity. However, it was found that when those images are subsequently processed with a Gaussian blur, the shapes become more perceptible. In other words, some of the obfuscation effects are reversible. Consequently, more iterations of the IntoFocus algorithm would be required with the scatter filter.

Decision: The median filter was chosen because it affords fewer opportunities for attack, and a median-filtered image cannot be further clarified. While we do not claim this choice to be optimal, our experience—including ad hoc exploration—indicates that it will reduce the number of iterations required to effectively redact an image without disclosing the facial identities of persons depicted.

The median filter depends on a value called the *ksize*. The filter creates a window of size *ksize* (*width*) \times *ksize* (*height*) centered at each pixel and computes the median of intensities of all pixels (for each color channel). The resulting median value for each channel becomes the new intensity for that pixel. In our implementation, we used the Python Pillow library’s [56] median function, and that implementation requires the *ksize* to be an odd number. The filter also needs the dimension of all the images to be in the same range.

Decision: For the evaluation of IntoFocus, the largest dimension in images was equal to 640 pixels ($height \leq 640$ and $width \leq 640$ and $max(height, width) = 640pixels$).

Decision: For these images (largest dimension = 640 pixels), we chose to use 5 iterations, with the following $ksize$ values: 41, 27, 19, 13, 7. These thresholds are preliminary and were used only to drive the development of the system. They were determined by two collaborators and myself, using a systematic process, but are not regarded as adequate for production use. The evaluation of these values cannot be treated as generalizable. See Chapter 4 for a new set of filter levels based on a perceptual study. These numbers represent the $ksize$ —i.e., height and width, in pixels, of the window—used for the median filter. Thus, in the first iteration, the image is filtered with $ksize = 41$. In the fifth (and last) iteration, the image is filtered with $ksize = 7$.

Initially, it was planned to display the faces discovered by workers at the same $ksize$ value with which they were found. For example, faces discovered in stage 1 ($ksize = 41$) would be shown with $ksize = 41$ in all subsequent iterations. However, when shown in the context of an image that was filtered at a lower level (e.g., $ksize = 19$), we found that the faces were easier to recognize. We considered concealing them entirely (i.e., solid black), but that might impede the discovery of other faces in future interactions. Therefore, we opted to filter the faces at a higher $ksize$ value.

Decision: After each iteration, the faces discovered by workers are further filtered with the following filtered at the following $ksize$ values: 81, 41, 27, 19, 13

Example: In stage 1, the image is filtered with $ksize = 41$. Workers annotate the faces of Alice and Bob. In stage 2, Alice and Bob are filtered with $ksize = 81$ while the rest of the image is filtered with $ksize = 27$. Workers annotate the face of Charlie. In stage 3, Alice and Bob are filtered with $ksize = 81$, Charlie is filtered with $ksize = 41$, and the rest is filtered with $ksize = 19$. This proceeds accordingly with the filter levels given above.

3.2 System Precautions

To ensure that the people’s privacy in the image is not compromised, the IntoFocus method uses the following methods to increase the accuracy of the output [57], [58].

3.2.1 Attention Check

To ensure that workers perform the task correctly, the system contains a mechanism where the system knows the locations of the faces in specific images in each set. If the worker did not highlight any clearly visible faces, the system would flag them and replace their work with another worker. This process is based on filtering outputs using ground truth comparisons [59], [60]. The reasoning behind the mechanism is identifying all workers that are not performing the task correctly. The workers are presented with the attention check image, and they are required to highlight easily detectable faces for their work to be accepted. The attention check images were displayed in random order, so the workers would not know which images evaluate the accuracy of their work. Due to the potential impacts of failure to prevent privacy disclosures, the system applies these mechanisms to uphold the promise of preserving the person’s privacy. IntoFocus evaluates the correctness of results for the attention check. If workers’ entries do not meet a baseline standard (described in 3.3.3), then an additional worker will be asked to perform the same task to compensate for the possibly missed highlights.

3.2.2 Collusion Prevention

In related work, Kaur et al. [16] demonstrated a system that segments an image and asks workers to choose regions that contain sensitive information. It is effective at reducing the risk of disclosure to individuals but is vulnerable to coordinated group attacks. Our system uses anti-collusion protection to prevent workers from working together and redacting the entire picture together. The first protection is that no worker can work on the same image more than once, such as seeing the same image at different stages. The second protection that helps prevent such an attack is that workers see images with all the previous highlights filtered out. Finally, to ensure that no one worker purposefully skips a face, each image is presented to three workers. All their highlights are combined and filtered out (the process of aggregating outputs[61]), so if one worker did not highlight a specific face, another worker would have the chance to do so, which increases the privacy of the people in the image. For workers to work together to extract all the information, they would need at least 15 different

workers, and all 15 need to accept the same task that contains the image they are trying to release unredacted.

3.3 Experiments Setup

The experiments (Figure 3.5) are designed to accommodate all the previous mechanisms and test each of them. We hypothesize that the IntoFocus method yields a balanced combination of identity preservation and face redaction. At the same time, each of the Control methods would focus only on one of them. The following experiment specifications hold for both experiments.

3.3.1 Treatment Condition (IntoFocus)

This is the method we propose in this paper, which was discussed in detail in the previous sections. The IntoFocus method will use five different stages, each with a different *ksize* value, and each stage will be presented to at least three different workers. At the beginning of each stage, the highlights of all the previous workers are redacted from the image before the image is presented to the new set of workers. Workers who previously worked on an image will not be allowed to work on that image again.

3.3.2 Control Conditions (naïve)

For the experiments, the control is defined as a fixed *ksize* value on all images. This method was chosen because it is one of the common methods used in practice for redaction and obfuscation. The *ksize* value is commonly static on all the images (similar to how it is used in these experiments). Giving us a total of five different control conditions where each of them is different from the others. There are 5 control conditions. The *ksize* values for the control conditions are the same as the stage *ksize* values and are thus called control X (where X is a number between 1 and 5), and the *ksize* values used are 41, 27, 19, 13, and 7, respectively.

3.3.3 Attention Check

To make sure the workers perform the task correctly, filtered images where the subject in the image can be easily found but not identifiable are used with each image set. A different attention check image is presented for every $ksize$ value of each set. The images were handpicked to make sure that the faces were visible. The test is convenient in that after the completion of each stage, the system decides if the worker passed the requirements or if the system needs to stop that worker from performing any additional tasks and replace their work. The method was evaluated by taking the location of each highlight and checking if it intersects with any of the faces in that image. A single intersection (does not cover the entire face) would pass the check, and the highlights would be accepted.

3.3.4 Face Selection

To test that the system can hide a person’s identity in an image, A set of 8 face photographs and an "I don’t know" image are presented next to each image. The workers are asked to select a face that they believe is in the image. There is always only one correct face from these photographs, and all the other faces are similar (hair color, features, face shape, etc.) in a way to the faces in the image or the correct face.


3.3.5 Participants

Participants are hired through Amazon Mechanical Turk. To avoid some of the issues faced in the preliminary studies, workers need to have a 90% success rate to participate. Each Human Intelligence Task (HIT) was rewarded \$0.75, and there was a total of 691 unique workers. A total of \$1898.57 for the first experiment (including bonuses) and \$593.25 for the second experiment.

3.4 Experiment 1: Actors

This experiment (figure 3.2) was performed using images of actors because of the increased image quality and the availability of multiple images per actor/actress in different conditions,

1. Highlight any of the following that you find in the image below: faces, tattoos, phone numbers. Use the smallest highlight possible to cover all the personal information. From the faces on the right select an actor/actress that can be seen in the image.

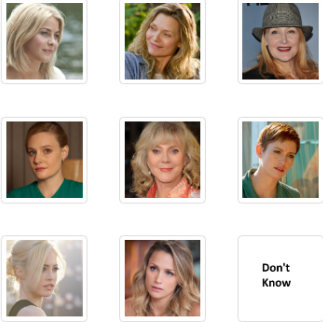


2. Can you see where any faces are (even if you can't recognize them)?

☐ Yes

☐ No

3. Try to find one of these actors in the image (at left). It might be difficult to find. **BONUS:** \$0.16 if you select the correct actor/actress; or \$0.02 if you select *Don't know*.



4. What allowed you to identify the actor?

☐ The scene ☐ The face is visible

☐ The clothes ☐ Tattoo

☐ Remember the movie ☐ Was not able to identify ☐ Other

5. Other comments on this image

Figure 3.2. This figure shows the interface used for experiment 1 with actors. The workers were required to select a face and answer if they could find a face in the image. The highlighting is optional because of the possibility of them not being able to see any faces. They were also required to select how they were able to identify the faces.

hairstyles, and locations. Giving the experiment a large area to test the system. The focus of this experiment was to measure the method’s ability to redact faces in images that contain faces in all colors, shapes, and sizes in the same image (people in the foreground and background at the same time). The dataset also provided information about how workers react to scenes and actors they are familiar with.

3.4.1 Dataset

The images being presented to the workers are from the IMDB dataset [62], [63]. This dataset was chosen because it provided many images from awards ceremonies, behind-the-scenes shots, and portrait images of all the actors. There were people in both the foreground and the background of the images. In some images, the people were camouflaged or hidden

in the corner of the image. The largest face was 600 pixels in height and 400 pixels in width. There was a total of 340 faces across the 80 test images.

In the experiment, there were 52 different actors, 27 male, and 25 female were chosen to be identified by the workers. The tested images contained at least one actor from the selected set, and other people present in the image were not limited to a specific race or skin color. To effectively measure whether workers recognized the face based primarily on facial features—as opposed to skin color, hair shape, or hair color—we had to narrow the scope of physical features. Skin color is essential because it is apparent even when filtered to a level that the facial features would not be identifiable. If we showed a filtered face and eight comparison faces representing four skin colors, the worker would have a 50% chance of guessing based only on the skin color in the filtered photo. We used Caucasian faces because their disproportionate prevalence in the entertainment industry made it easy to find celebrities that workers might have a chance of knowing.

3.4.2 Bonus

In this experiment, to encourage the workers to perform the experiment correctly and as accurately as possible, the method discussed as double or nothing [64] was adapted to add some incentive for the workers. Workers were given a bonus of \$0.02 if they selected that they did not know the actor in the image, and they were given \$0.16 if they selected the correct face. The reasoning behind this is to reduce random guessing and encourage the workers to only answer when they are sure of their response.

3.4.3 Image Presentation

There were 20 Human Intelligence Tasks (HITs). Each HIT contained five different images. Out of the five images, two were Treatment images, two were Control images, and one was an Attention Check image. Each of the main four images was tested for all of the six conditions. The attention check image was not used as part of the evaluation process because it was chosen to fail the identity preservation test. The attention check images were randomly ordered among the treatment and control conditions. There were a total of six different

conditions; treatment, control 1 (with k size value of 41), control 2 (k size value of 27), control 3 (k size value of 19), control 4 (k size value of 13), and control 5 (k size value of 7). The experiment required 15 workers each for each condition and a minimum of 1800 assignments.

3.4.4 Experiment Task

Before crowd workers start working on the task, they need to answer a questionnaire about actors and actresses they are familiar with. They are presented with several face images and are asked to select all the actors they know. After that point, the task begins. Workers are required to perform four tasks on each image. The first task is to highlight all the faces that can be detected in the image. The second task is to answer a simple question: "did you find a face in the image?". The third task is to select the face of an actor or actress they think is present in the image from the set of faces on the side of the image. The fourth task was to answer a question about how they could identify the face in the image. The fourth task was to correctly analyze the system and understand if it failed and what the reason was. During a pilot study, the second question proved helpful in identifying workers who did not perform the task correctly, which guided us towards focusing more on this issue.

3.4.5 Evaluation

The evaluation for the method and the system were performed using Amazon Mechanical Turk (AMT). A total of 20 different HITs were posted, covering 180 images (100 attention check images + 80 test images), containing 52 subjects (27 male and 25 female). The six methods were tested with the same images and were presented in 1800 assignments (an assignment in AMT refers to the agreement between the person requesting the task and the person performing the task). The effective hourly rate was \$5.67 per hour, which was less than our target of \$9 per hour. Individual workers earned between \$1.50 and \$34.62 per hour, but on average, workers took longer than we anticipated to complete the task.

The evaluation starts by showing the feasibility of the system when compared to the control. Then the results are explored to see how much familiarity increases a person's ability to identify someone they know.

3.4.6 IntoFocus vs. Control

The first step to verify that a system is feasible is to prove that it is significantly different from the control method. To compare the two categorical methods, we need to show that they are significantly different. To accomplish this, a chi-square test between the IntoFocus and each of the control methods was taken. The analysis shows a significant association between the method used and the results gained from the redaction experiment with $p < .001$ with control 1, control 2, control 3. This shows that the odds ratio of the IntoFocus method succeeding in redaction is 3.42 times higher than the chances of the control method succeeding.

3.4.7 Attention Check

The attention check images that checks which workers were performing their tasks correctly. In crowdsourcing rejecting a worker’s submission resulted in not paying that worker and reducing workers’ ability to accept tasks. That is why we contacted the workers who did not perform the task correctly instead of rejecting their work.

When a worker did not perform the task requested from him, the system allowed us to increase the number of assignments and gather more highlights to prevent the system from failing because of the inaccurate answers. The protection was able to catch 252 out of the 1819 assignments, where the workers failed to highlight any face in the attention check image.

3.4.8 Results

Out of the 340 faces in the images, only four faces were not successfully redacted by the IntoFocus method. Unlike the control, where each image was redacted under a single $ksize$ value, the IntoFocus method had undergone a full five-stage, five $ksize$ value iterative redaction process. Figure 3.3 shows the number of faces redacted over all the cases of treatment and control. The total number of faces was 340 faces throughout the 80 test images. The IntoFocus method successfully redacts 336 faces, showing that the method is better suited for redaction than a single $ksize$ value with several different-sized faces.

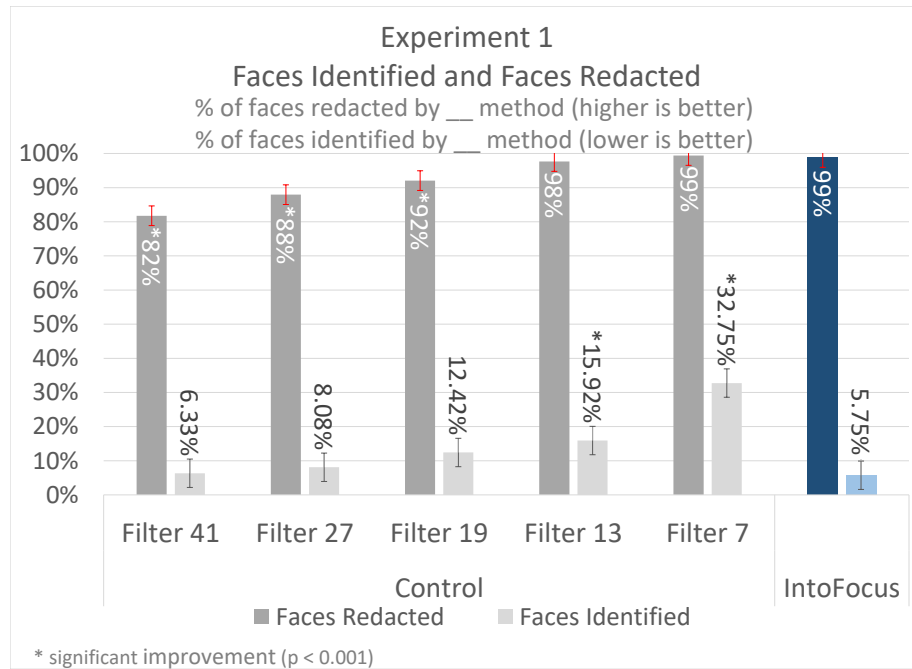


Figure 3.3. This figure shows the results of the experiment. IntoFocus was significantly better than the others because it combines the flexibility of low k size value with the preservation of higher k size value. With faces detected at each redacted at a higher filter level, the IntoFocus method can have an identification rate lower than that of the highest filter because faces detected at each stage are redacted at a higher filter level than that stage.

On the other hand, all the conditions had non-significant differences (except for control 4 and control 5) in face identification. Figure 3.4 shows that, unlike the control methods, The IntoFocus method had the lowest percentage of failure because a face was visible. The highest reason for the IntoFocus method's failure was when the workers were familiar with different aspects of the movie (recognized the clothes, knew the scene, and knew the movie). In the lowest control, some workers provided the full list of names of the actors present in an image, and in some cases, the name of the movie where that scene originated. These results show that having a personal interaction with the image or actors affects the worker's ability to remembering the scene.

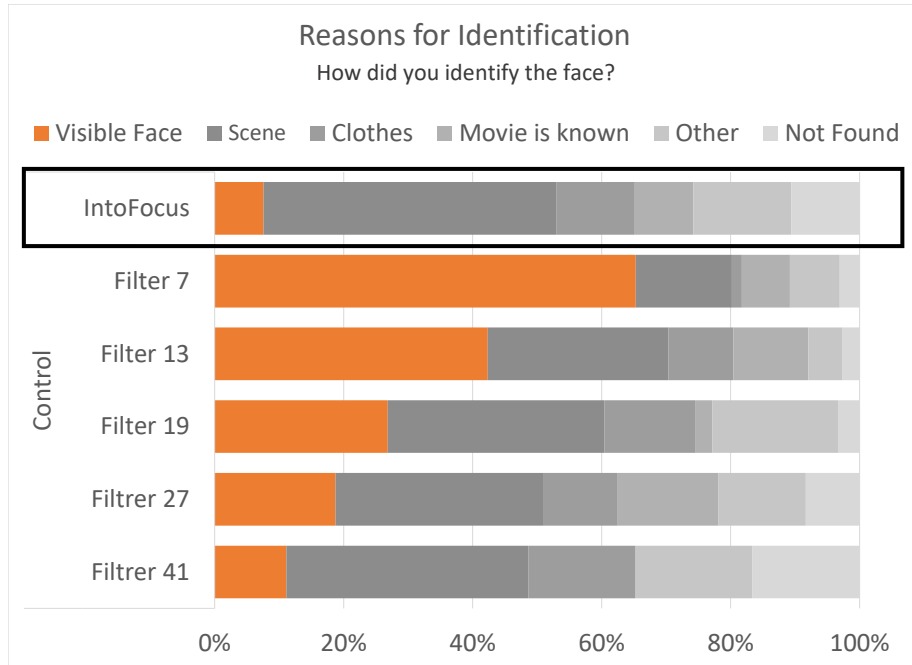


Figure 3.4. This figure shows the reasons the workers provided when they were able to correctly identify a face for all the conditions in experiment 1. A trend can be seen that as the k size value is reduced, the number of workers that were able to identify a face because it was visible increases. In the IntoFocus method, the face being visible was not a concern. Since these are the percentages, we can see that the highest identification rate happened for the IntoFocus method is when the workers were familiar with the scene where the image was taken.

3.5 Experiment 2: Random People

For this experiment (Figure 3.5) precautions have been taken to reduce the issue that increased the workers' ability to identify someone in an image. Unlike the previous experiment, a dataset of random people was collected. This experiment aims to test the system with real-life images and see how that affects the worker's ability to identify someone in the image.

3.5.1 Dataset

The images used in this experiment were taken of random people posing for the camera, talking amongst themselves, eating, working on computers, cleaning, advertising, and other activities. The images contained people in the foreground and background of the image. The smallest face in the dataset was 3 pixels in height and width, and the largest face was 200 pixels. There was a total of 89 faces across the 24 test images.

The reason behind building a dataset from scratch was to truly test how the system performs when the workers do not have prior knowledge of the images or when they were taken. This experiment was done to test how workers identify people they do not know. These images were taken using a Google Pixel cellphone camera with HDR (high dynamic range). The dataset contained faces of mixed ethnic backgrounds, and all were between the age of 18 and 35.

The test sample contained a total of 30 images, 6 of which were attention check images. 60 different people were participating in the images, 34 males and 26 females.

3.5.2 Bonus

This experiment did not provide any bonuses because of the increased number of workers who failed the previous experiment’s attention check and only selected faces.

3.5.3 Image Presentation

There were six HITs, each with five different images. Each HIT contained four images for a single specific condition and one attention check image. The order of the images was randomized such that the order was different for each worker.

There were a total of six different conditions; treatment, control 1 (k size value of 41), control 2 (k size value of 27), control 3 (k size value of 19), control 4 (k size value of 13), and control 5 (k size value of 7). So six image sets, six conditions, 15 workers each, for a minimum of 540 assignments.

Image 1 Image 2 Image 3 Image 4 Image 5

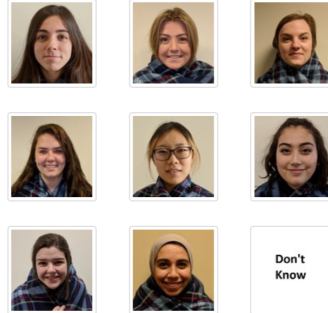
1. Highlight the entire face, hair, hat (if any). Important:
Read instructions at top before you begin.



2. Can you see where any faces are (even if you can't recognize them)?

- ☐ Yes
- ☐ No

3. Try to find one of these people in the image. It might be difficult to find.



4. Comments on this image

Figure 3.5. This figure shows the interface used for experiment 2 with random people. The workers were required to select a face and answer if they were able to find a face in the image. The highlighting is optional because of the possibility of them not being able to see any faces.

3.5.4 Experiment Task

This experiment starts by showing the crowd workers the filtered images directly and asks them to add highlights to any face they see in the image and select a face they think is present.

3.5.5 Evaluation

The system was submitted to AMT for a total of six different HITS where each HIT would have a minimum of 90 assignments (90 different workers). For each condition, each batch of five images was assigned to 15 different workers. With six conditions in total, the number of assignments was 565 (90 assignments per HIT = 540 assignments and 25 assignments for

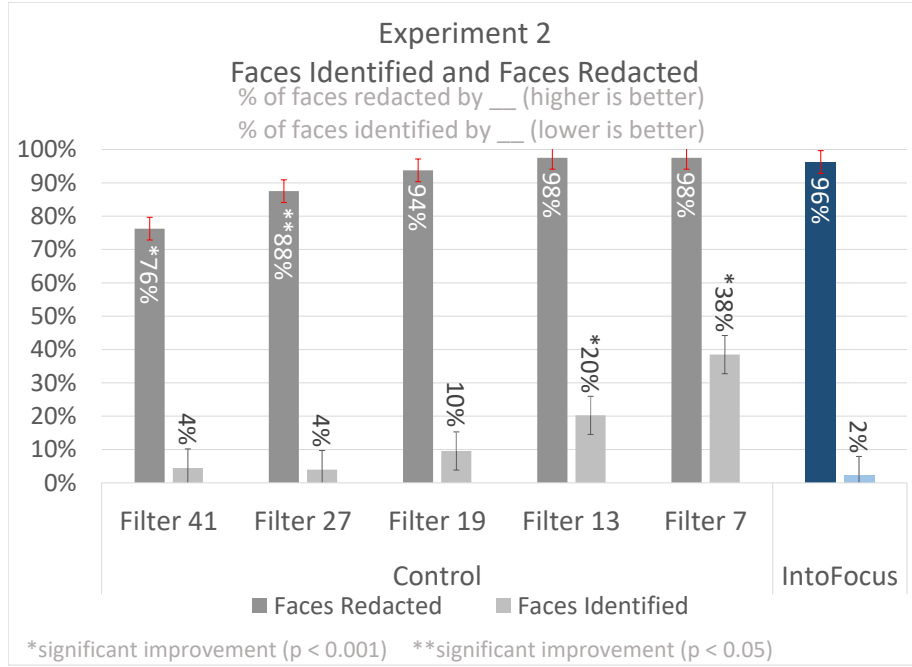


Figure 3.6. This figure shows the percentages of faces redacted in experiment 2. IntoFocus was significantly better than the others because it combines the flexibility of low k size value with the preservation of higher k size value.

workers that failed the attention check). Each worker was paid \$0.75 per HIT for an hourly rate of \$8.17, and we had 273 unique workers.

3.5.6 IntoFocus vs. Control

For this experiment, there were six different conditions. The IntoFocus method showed a significant improvement in redaction against control 1 and control 2 and significant improvement in identity preservation against control 4 and control 5. There was no significant difference between the rest of the methods.

3.5.7 Results

Figure 3.6 shows that the IntoFocus method achieved missing only three faces out of the total of 89 faces. The IntoFocus method had improved results compared with experiment 1 regarding the worker’s ability to identify the person in the image. When looking at both redaction and identity preservation, it can be seen that overall the IntoFocus method was the most balanced solution. It did not miss many faces like the higher *ksize* value methods and did not allow many workers to identify the faces like the lower *ksize* value methods.

3.6 Discussion

Many face detection algorithms for images in the wild could be used as a substitute instead of IntoFocus [65]. The motivation to pursue IntoFocus even though these methods exist was because of the lower success rates of these methods. The survey paper for face detection in the wild [65] reports that the best face detection algorithm has a success rate of 78.8%. One of the current state-of-the-art algorithms has the highest score of 88.9% [66]. These results are not good enough when speaking of a person’s privacy.

To design a system that preserves people’s privacy, we need to understand how knowing the person in the image affects the worker’s ability to identify them. The actor dataset used in the first experiment was perfect for the study. The dataset gave us a chance to present workers with people they are familiar with. It also gave us images that had people in different sizes and different regions of the image. This experiment shows that if the workers know the actors or the movie, they can identify the person they are redacting no matter the *ksize* value.

The second experiment showed that even if the workers do not know the subjects, they can still guess correctly. Several of the workers explicitly told us that they could identify the person because of their hairstyle, hair color, and skin tone. Some workers pointed out that they could correctly identify the faces because of the expected body mass they found from looking at the faces. The findings show that giving a set of face images does not necessarily mean that the method does not work correctly. The faces on the side of the image provide guesses for the workers. They can select the one that closely matches the filtered figure.

Table 3.1. The number of times the conditions succeeded in performing the tasks of redaction and identity preservation. In the faces redacted row, higher is better. In the faces identified row, lower is better.

Experiment 1	IntoFocus method	Control 1	Control 2	Control 3	Control 4	Control 5	Amazon Rekognition Face Detector
Faces Redacted	98.82%	81.76%	87.94%	92.05%	97.64%	99.41%	92.94%
Faces Identified	5.75%	6.33%	8.08%	12.41%	15.91%	32.75%	–
Experiment 2	IntoFocus method	Control 1	control 2	control 3	control 4	control 5	Amazon Rekognition Face Detector
Faces Redacted	96.62%	78.65%	88.76%	94.38%	97.75%	97.75%	77.52%
Faces Identified	2.19%	4.39%	3.95%	9.45%	19.12%	38.46%	–

The two experiments were different in incentives (bonuses) given to the workers. The first experiment gave an amount to workers who could find the correct face in the image. The second experiment did not give out any amount for any answer. The incentives were removed from the second experiment mainly because of the high number of submissions that failed the attention check (workers did not add any redactions). The incentives used made the workers focus on face selection because they were rewarded for that aspect only. The problem is that the main task was to highlight and identify faces, not just one of them. The removal of the incentives showed a significant increase in the number of workers that passed the attention check.

3.6.1 Future work

The method still has some obstacles to pass. The addition of machine learning and computer vision techniques would reduce the load on the workers and increase the productivity of the method. These techniques can automate the attention check image selection process instead of manually building the ground truth and would make workers focus on faces that are not detectable by face detection algorithms.

The method would benefit immensely from using face de-identification [34], [37] methods, where another face would replace a highlighted face before submitting the image to the task. The issue with such algorithms is that they are not ready for "in the wild" images and are purely restricted to images taken from the front of a person. This would also solve problems relating to segmentation tasks, where they would need to see the outline of a human to perform the task correctly.

The Anti-Collusion methods are evaluated by forming two teams and asking one team to perform the task. Another team to try to make the system release an image without proper redaction of all the faces.

The current system focuses on faces because of the lack of a filter that withstands different image data types (e.g., faces, texts, medical records, credit cards). A possible solution is an image filter that works with both text and images, one that would hide the small details that help people with identification and still allows them to find the location of the text.

3.7 Conclusion

This paper presented IntoFocus, a method that given an image, uses crowdsourcing to redact facial information that allows a person to be identified. IntoFocus adds to the knowledge of accurately redacting images while minimizing identity revelation to the worker and maintaining consistent results for different-sized faces. IntoFocus provides a method for crowdsourcing applications [10], [11], [67] to reduce exposure of facial information.

Through the results of the method, IntoFocus found and redacted 336 faces out of 340 in the first experiment and 86 faces out of 89 in the second experiment. The system comes with a cost of \$0.9375 per image, and the image’s content is not shown to any person.

4. HUMAN PERCEPTION OF MEDIAN FILTERED FACES

Portions of this chapter were copied verbatim from a paper published in the 8th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2020) [7].

Humans are adept at detecting faces. They have been trained since birth to detect and identify people and their faces. For a human to identify someone, they look for features that do not focus on a face. Some focus on the gait as a person walks, while others focus on hairstyle and color. That difference shows that there is more than one way of identifying people.

When it comes to detecting people in filtered images, people tend to look for the shape of a body with a semi-circle on top. They use any information they can find in an image to improve their detection ability [23].

This chapter explores and builds upon how people perceive faces in median filtered images. It starts with a study that asks participants to detect and identify faces in different images and scenes. The motivation for this exploration was to extract the filter levels that would guarantee a high rate of detection and a low rate of identification.

To evaluate the human perception of faces in median filtered images, an online study is performed. Because of the online nature of the study, a person is tasked with either detecting or identifying a face. The purpose was to find the range of filter levels between when a person can detect a face and when a person can identify that face. The reason for finding the range for different images is to derive the set of filter levels that allow faces—of different features—to be detected but not identified.

The definition of detection in our experiment is a person’s ability to add an ellipse covering a face from the jawline or the lower edge of a beard until the top of the hairline, including any hats or hair covers (figure 4.1). On the horizontal axis, the ellipse needs to cover both ears and any hair or hat covering the sides of the head.

The definition of identification in our experiment is a person’s ability to specify which category a face belongs to and select the correct face from a list of 8 faces of similar gender, skin tone, and hair color.



Figure 4.1. Participants were required to redact the entire face to the top of the hairline, including ears, facial hair, hats, and hair covers. This is an example of the expected face redactions.

4.1 Face Perception Studies

To calculate thresholds, we conducted a two-part study of (a) face detection (figure 4.2) and (b) face identification (figure 4.3) by humans.

The ultimate goal was to find optimal filter levels that would ensure that IntoFocus can detect each face with some probability (e.g., $P(\text{any worker detects}) \geq 0.99$) while limiting the risk that any worker identifies a face to some low probability (e.g., $P(\text{any worker identifies}) \leq 0.02$). To do this, we needed to answer a key question:

If $N\%$ of people can detect a face at blur level k_{detect} (or lower), what is the minimum blur level at which no more than $M\%$ of people can identify the face, supposing they knew the person or had some reference photo available.

We took $N = 98\%$ and $M = 2\%$.

The studies were performed on Amazon Mechanical Turk (AMT) [68]. Each participant was only allowed to perform either detection or identification tasks. Each participant performed the required task on five different images. There were 2844 participants each was paid \$0.75 for their time. Each Human Intelligence Task (HIT) took an average of 3:14 minutes.

4.1.1 Rationale Behind the Experiment

The IntoFocus system relies heavily on selecting filter levels that allow detection but prevent the identification of faces. Without mapping the people’s performance using the selected filter method, these filter levels are only random guesses. Thus, the first task is to show a gap in filter levels between detection and identification. Then, using that information, we apply a model that extracts the filter levels to use.

4.1.2 Study Setup

The study was set up to find the filter levels for all possible faces in an image of fixed size, with the largest dimension being equal to 640 pixels ($height \leq 640$ and $width \leq 640$ and $max(height, width) = 640 \text{ pixels}$).

For this experiment, a median filter was applied to all the images. Each image is filtered with the levels ($ksize$) from 1 up to 189.

4.1.3 Study Part 1: Detection

The task is to add an ellipse on each of the faces seen in the image. Ellipses can be modified and/or removed after they are added. In each HIT, one image served as an attention check. The image had a reduced blur level, and all the faces could be easily detectable. These images were added to ensure the validity of the data and that each participant was performing the task correctly. The study revolves around finding the $P(detect_f)$ where f is a specific face in the image. A binary search is used to find the filter level where at most, one person does not find the face. That point represents the point where the filter level is high

enough that people will start to be unable to detect the face and low enough to ensure a large enough gap from when the face can be identified.

4.1.4 Study Part 2: Identification

Participants were presented with the *main* image with one face outlined with a red ellipse (We applied the ellipse). They then attempted to match the depicted person to one of eight (8) reference faces.

Reference faces were selected with the same gender, hair color, and skin tone of the person in the ellipse. This was to minimize the chance that participants might guess correctly based on characteristics other than the facial features. The use of people of similar features also ensured that the faces provided k -anonymity [36] to the face in question, where $k = 8$. (The scope of this research is limited to *facial* identities.)

Each image had only one red ellipse, even if other faces were present—this ensured consistency in our study design.

This part of the study focuses on searching for the point where participants start to identify the face in the image. Unlike the detection study, people in this study have the ability to guess the correct face. Because an attention check would not apply in this case, all the participants' inputs were considered in the evaluation. The search was focused on searching for two filter levels. The first when $P(\text{identify}_f) = 0$ and the second when $0 > P(\text{identify}_f) \leq 4\%$. The filter level when $P(\text{identify}_f) = 0$ is considered the point where the face was too filtered for anyone even to guess the correct face (when we surpassed k -anonymity of 8). The filter level when becomes $0 > P(\text{identify}_f) \leq 4\%$, this is the point when at least one person has enough information to guess the correct facial identity, while all the others were not able to extract that information. If the two cases followed each other, with the first case being at a higher filter level, then that is the point being searched for. Otherwise, the search continues at a higher filter level.

4.1.5 Dataset

The dataset is a combination of the IMDB dataset [62] and the images of random people. There were a total of 60 images used in the experiment. The datasets covered faces of all colors, shapes, and sizes. There were a total of 157 different faces to select from in this study. There was a total of 336 faces in the images.

4.1.6 Evaluation

The study was initially planned to be performed in a lab, but due to the COVID-19 pandemic, we were forced to move it online. To control the cost of the study, we used a binary search to find the threshold (filter level) where a face becomes detectable or identifiable.

In the detection study, the search tree was used to find the filter level for each image where only one worker out of 25 cannot detect the face. That filter level being searched for represents the point at which the image starts becoming too filtered for people to detect faces. The starting point and boundaries used in the binary search algorithm were obtained from a pilot study on the same images in an in-lab study. The starting point was the median filter level of all the participants in the pilot study. The upper and lower bounds were the highest detection and the lowest detection filter levels, respectively.

In the identification study, the search tree was used to find the filter level where only one worker could correctly identify the face. That filter level being searched for represents filter level where the image becomes too filtered for people to identify the faces correctly. The upper bound for the identification study was the lowest filter level that allowed detection. The lower bound was the filter level that allowed all the pilot study participants to identify the person correctly. If the face requires a filter level higher than the upper boundary, the filter level would be incremented by a value of 2 (the nearest odd number) until the modified success level is found.

When the identification filter level is higher than the filter level where all the participants can detect, then we have the probability that a face is identified, given that the face is detected (in the identification study, the locations of the faces to be identified are given). Thus, the probability required is $P(\text{identify} \cap \text{detection})$. The new

probability of identification value can be calculated using the Kolmogorov axiom [69], $P(\text{identify} \cap \text{detection}) = P(\text{identify} | \text{detect})P(\text{detect})$.

The experiment was designed to produce the filter levels at which each face in the image is detectable and identifiable. Using that information, we use the following model to extract the appropriate filter levels for each face in the images. To evaluate, we need to model the detection and identification separately. Starting with the detection model, it needs to guarantee that 98% of the population can detect the face. 98% of the population would be able to detect at the point of the two percentile. Using the probability density function $P(a < X \leq b)$ where a and b are filter levels and setting the probability to equal the 2% required for detection:

$$P(a < X \leq b) = 0.02$$

Setting the lower end of the interval a to 1, which is a clear image.

$$P(1 < X \leq b) = 0.02$$

Since the probability density function is the difference between the cumulative distribution function of a subtracted from the cumulative distribution function of b

$$P(1 < X \leq b) = F_X(b) - F_X(1) = 0.02$$

If a face cannot be detected when an image is filtered, we must assume that the face cannot be identified when the image is clear. Under that assumption, the cumulative distribution function of 1 is 0.

$$P(1 < X \leq b) = F_X(b) = 0.02$$

$$F_X(b) = \sum_{b_i \leq b} p(b_i)$$

Thus, we are searching for b, where the cumulative distribution is equal to 0.02. This method is applied to all the different faces, extracting the 2% detection values for all the possible faces. The same is done for identification, with 98% of the population not being able to identify.

For each node in the search tree, 25 workers on AMT were hired to perform the given task (detection or identification). HITs that did not pass the attention check were replaced. No workers were rejected in this study. After finding the filter levels for each image, they are ordered by the face size (width and height). Polynomial regression was performed to estimate the filter levels for detecting and identifying different-sized faces (Figure 4.4).

The results were evaluated by taking the blur level when each face was detected and identified. Using the 98th percentile for the identification and the 2nd percentile for the detection, we get a region where all faces $\pm 2\%$ are detectable and none of the faces $\pm 2\%$ are identifiable. Now that the boundary is set, starting from the lowest identification data point, a vertical line is drawn from the beginning of the identification line until it intersects with the detection line and at the intersection point, taking a horizontal line until it intersects the identification line. That horizontal line represents the lowest filter level used. The process is repeated, creating a staircase, and each horizontal line found is a stage to be used in the IntoFocus method. Each task was offered for \$0.75, with a total of 2844 assignments. The hourly rate was \$13.99. The total cost for mapping people’s performance on 60 median filtered images was \$2,986.20.

4.1.7 Results

The results show a gap in the filter levels between the lowest detect and the highest identify (Figure 4.4). The model used starts with a face width and height of 27 pixels, projects horizontally to the identification line, then projects vertically to the detection line. The horizontal projections were the filter levels used in the IntoFocus method. The model stops when the horizontal line no longer intersects with the identification line. The resulting filter levels were (85, 53, 35, 25, 17, 13, 9). The results confirm our first hypothesis: a gap exists between when participants can detect and identify a face. The plot in Figure 4.4 uses only the second percentile for detection (the point where almost every person can detect the faces) and the 98th percentile for identification (the point where at most one person can identify). Even though the values are the extremes in both cases, the two are separable.

For these images (longest dimension = 640 pixels), our model gave us 7 iterations, with the following *ksize* values: (85, 53, 35, 25, 17, 13, 9). These numbers represent the *ksize*—i.e., height and width, in pixels, of the window—used for the median filter. Thus, in the first iteration, the image is filtered with *ksize* = 85. In the seventh (and last) iteration, the image is filtered with *ksize* = 9.

After each iteration, the locations of the faces discovered by workers are further filtered with the following *ksize* values: 113, 85, 53, 35, 25, 17, 13.

Example: In stage 1, the image is filtered with *ksize* = 85. Workers annotate the faces of Alice and Bob. In stage 2, Alice and Bob are filtered with *ksize* = 113 while the rest of the image is filtered with *ksize* = 53. Workers annotate the face of Charlie. In stage 3, Alice and Bob are filtered with *ksize* = 113, Charlie is filtered with *ksize* = 85, and the rest is filtered with *ksize* = 35. This proceeds accordingly with the values given above.

4.2 Evaluating The Filter Levels

Now that the new filter levels were extracted from the results of the previous experiment. A new experiment that tests the hypothesis (1) The people performing the redaction task will have $\pm 2\%$ success rate in correctly identifying any of the test faces (excluding random chance).

4.2.1 Experiments Setup

The experiment (Figure 3.5) presents crowd workers with an image filtered with one of the previously found filter levels. Like the previous experiment, it requires workers to add an ellipse on a face and correctly select the face from the given faces. The ellipses will count toward detection, while the selected faces will count towards identification.

4.2.2 Treatment (IntoFocus)

This is the IntoFocus filter method described in the previous chapter. It starts with the highest filter level and asks crowd workers to add an ellipse on all the faces they can find. Once done, these locations will be redacted at a higher filter level. The filter level on the rest of the image is reduced, and the task is repeated until the final filter level. Each filter level will require three crowd workers. Each crowd worker will only be allowed to work on an image once.

4.2.3 Attention Check

The system contains an attention check mechanism in each set of five images to ensure that workers perform the task correctly. An image is shown at a filter level where all the faces can be easily detected (based on the data gathered in the previous study). If the worker did not add an ellipse on all the faces, the system would flag them and replace them with another worker. The reasoning behind the mechanism is identifying all workers that are not performing the task correctly. The attention check images were displayed in random order, so the workers would not know which images evaluate the accuracy of their work. With the task being of moderate to high risk, the system needs to have such mechanisms to uphold the promise of preserving the person’s privacy. If the system flags a worker, their work will be disregarded, and an additional worker will be asked to perform the same task to compensate for the possibly missed faces.

4.2.4 Face Selection

As part of our evaluation, we validate that workers cannot identify the faces in the main photo. A set of 8 reference faces are presented next to each image. Workers are asked to select any faces that they believe are in the image. If they cannot match any of the reference faces to the main image, they can click a button labeled, “I don’t know.”

For any main image, we evaluate this for only one of the depicted faces. (The study design becomes intractable if we try to evaluate this for all faces.) Therefore, in each trial, only one of the reference faces is present in the main image.

Success is indicated when workers choose “I don’t know” or guess with random probability, based on the number of choices offered (i.e., 12.5% for eight reference images). We cannot judge success on any individual trial, but we can measure the rate of success for a group of trials.

Since this work is solely focused on *facial* identities, workers should not narrow the set of reference faces by any characteristics other than the facial identity. Therefore, the reference faces are selected to have the same non-facial characteristics: hair color, hair length, and skin tone. The faces were selected to test for 8-anonymity[36] in the images where the

quasi-identifiers are gender, skin tone, and hair color. The sensitive attribute is the face in the main image.

4.2.5 Participants

Participants are hired through AMT. To avoid some of the issues faced in the preliminary studies, workers need to have a 90% success rate to participate. Each HIT was rewarded \$0.75, and there was a total of 127 unique workers. A total of \$216 for the experiment and an average of \$10.29 per hour. The average time to finish a hit was 4:22 minutes.

4.2.6 Dataset

The images being presented to the workers are from the Internet Movie Data Base (IMDB) dataset [62] and a dataset that we collected to ensure the system is tested on real-life scenarios (chapter 3). The IMDB dataset was chosen because it provided many easily obtainable images from awards ceremonies, behind-the-scenes shots, and portrait images of actors. There were people in both the foreground and the background of the images. In some images, the people were camouflaged or hidden in the corner of the image. There was a total of 232 faces in the 50 test images. The largest face was 600 pixels in height and 400 pixels in width.

In the evaluation, a set of 186 different people were chosen to be identified by the workers. The tested images contained at least one (1) person from the selected set, and other people present in the image were not limited to a specific race or skin tone. To effectively measure whether workers recognized the face based primarily on facial features—as opposed to skin tone, hair shape, or hair color—we had to narrow the scope of physical features. Skin tone is critical because it is apparent even when filtered to a level that the facial features would not be identifiable. If we showed a filtered face and eight comparison faces representing four (4) skin tones, the worker would have a 50% chance of guessing based only on the skin tone in the filtered photo.

4.2.7 Image Presentation

There were ten (10) HITs, each containing five (5) different images. Out of the five (5) images, four (4) were treatment images, and one (1) was an attention check. The attention check images were randomly ordered among the treatment and control conditions. The attention check image was not used as part of the evaluation process because it was chosen to fail the identity preservation test. There was one (1) condition; treatment. There were ten (10) image sets, one (1) condition, twenty-one (21) workers each, for a minimum of 210 assignments.

4.2.8 Experiment Task

Workers are required to perform two tasks on each image. The first task is to add ellipses on all the faces that can be detected in the image. The second task is to select the faces of people they think are present in the image from the set of faces on the side of the image. If they could not solve one of the two tasks, they were required to click on buttons that say that they could not perform that task.

4.2.9 Evaluation

The evaluation for the method and the system were done on AMT. A total of 10 different HITs were posted, covering 50 images (10 attention check images + 40 test images). The IntoFocus method was presented in 210 assignments (an assignment in AMT refers to the agreement between the person requesting the task and the person performing the task). The system will be evaluated based on the ability to maximize detection and minimize identification. The results will be compared with Microsoft Azure’s face detection system instead.

4.2.10 Results

Out of the 232 faces in the images, 229 (98.7%) faces were detected. Microsoft Azure’s [70] face detection system detected 203 out of 232 (87.5%) from the same dataset. Out of the

840 images x assignments, only seven faces were correctly identified. The results from the experiment yielded an identification rate of 0.83%.

The filter levels were selected based on an identification rate of 4%. These values were selected to calculate the filter level when participants would start to identify faces. Because of the number of participants used at each step of the perception study (25 participants at each node of the search tree), the requirement was that at most one participant was able to identify, and $1/25 = 0.04$. Based on that data, the acceptable identification rate of the method is up to 4%. Similarly, a detection rate of 96% was selected for the filter level calculation.

Unlike the previous chapter, the filter levels were calculated with the guessing in identification as part of the formula and were designed to lower guessing to the minimum of 4%.

4.3 Discussion and Future Work

The perception study of median filtered faces showed that people are genuinely different in detecting and identifying faces. Some participants could accurately pinpoint faces when their peers could not detect anything until several filter levels later. This shows that people have different abilities when it comes to face detection. Similar to automated methods, where a method might perform better under specific conditions.

The addition of machine learning and computer vision techniques would reduce the load on the workers and increase the productivity of the method. These techniques have the ability to automatically redact the easily visible faces and allow people to focus on the occluded faces. Another direction was to use a machine-learning algorithm to assign different filter levels to different image regions. The image would then be sent out to crowd workers, and they would be tasked with the redaction. This process reduces the number of stages used by the IntoFocus method, reduces cost, and reduces the time needed to redact an image fully.

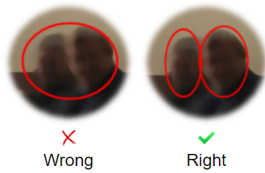
The polynomial curves estimated for detection and identification in the perception study only use one variable (face size) to estimate the filter levels. The results obtained from the perception study show that face size is the largest factor in estimating the filter level. Still, other factors exist, such as skin tone, illumination, brightness, the difference in contrast

between the face and the surrounding areas, which are all factors that affect the filter level. Further exploration into other factors could help strengthen the IntoFocus method by (1) increasing detection, (2) decreasing identification, and (3) reducing the number of filter levels (stages).

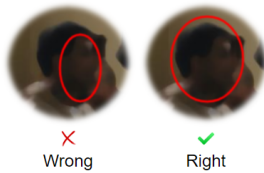
The IntoFocus method has a high cost to redact a single image. The minimum cost with the addition of the new stages is \$3.94 to redact a single image fully. Because of the large ratio (54%) of participants that fail the attention check images, the cost to redact a single image is above the minimum stated above. In the first iteration of the IntoFocus system (chapter 3), the cost of redacting a single image was \$2.81. For the system to be feasible for large sets of images, the cost must be significantly reduced.

Guidelines

Each ellipse covers only one face



Cover the entire face including hair and facial hair



If unsure if something is a face do not mark it



Shortcuts: Press "d" to start drawing an ellipse.

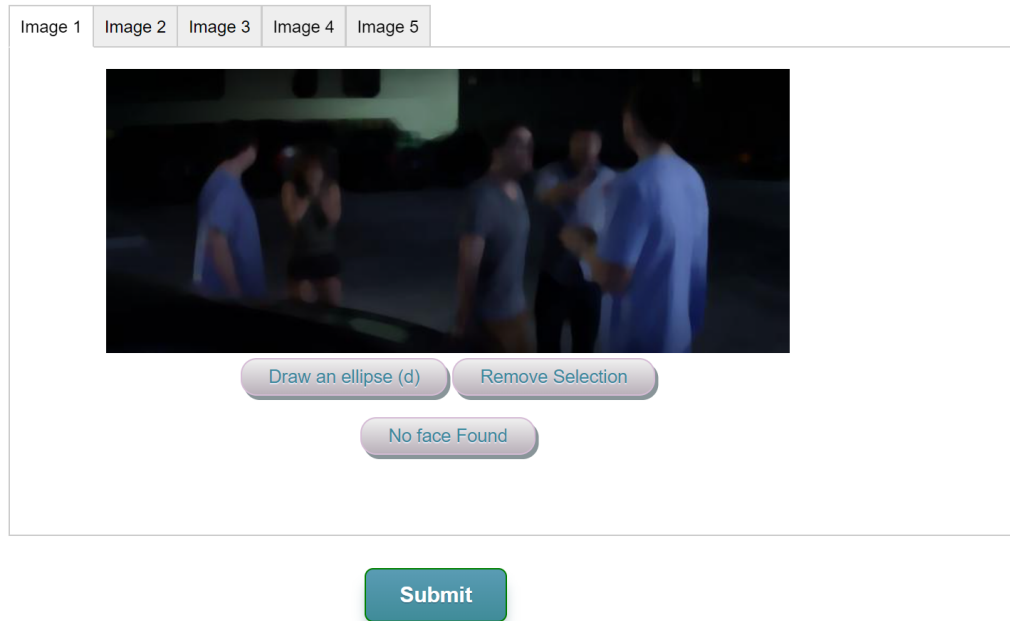











Figure 4.2. Face detection study. Participants annotated each face they detected with an ellipse. They could select the ‘No face found’ button if they are unable to detect any faces. The study was performed on AMT, with each HIT containing five images. One of the images was used for the attention check. That image would have a reduced filter level to make for easy detection. The attention check image was changed to a different image in the set at each leaf node in the search tree, to find the appropriate filter levels for all the images in the task.

The red ellipse in the image below indicates the location of a face in the image. Try to identify the person shown in the red ellipse by selecting the person(s) you think it might be from the smaller faces to the right. Select all of the smaller faces that might be the same person as the one in the red ellipse. If you are confident that it matches only one of the reference faces (at right), then select only that one reference face. This HIT will only work on Google Chrome. ?

Image 1 Image 2 Image 3 Image 4 Image 5



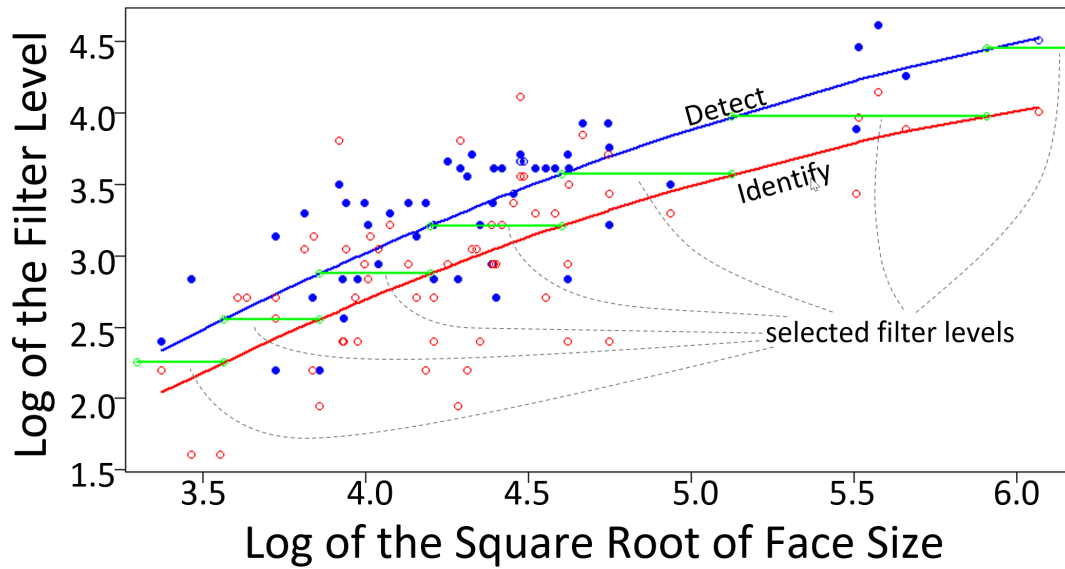


Any of these None of these

Submit

Figure 4.3. Face Identification study. Participants selected the reference face (right) matching the person in the main image (left) who was marked with a red ellipse. Multiple faces could be selected if they could eliminate some reference faces from consideration but could not identify the subject face (left) as definitely matching *one* of the reference faces (right).

Estimated lines for the 2nd Percentile of Detection and the 98th Percentile of Identification



- Detect data points — Detect Regression Line
- Identify data points — Identify Regression Line
- Selected Filter Levels with Model Intersection Points

Figure 4.4. Study results. The blue line and points are the second percentile of the detect values (the point right before the detection rate reaches 100%). The red line and points are the 98th percentile of the identify values (the point right before the identification rate reaches 0%). The green line is the staircase model that was used to select the filter levels for the IntoFocus method.

Add ellipses on the entire face, hair, and hat (if any). Be sure you have covered all parts of the eyes, nose, lips, eyebrows, forehead, cheeks, ears, chin. From the faces on the right, select the faces that can be seen in the image. **Questions?** aalshai@purdue.edu


Image 1

Image 2

Image 3

Image 4

Image 5



Draw an ellipse (d)

Remove Selection

No face Found

Any of these
"Don't Know"

None of these
"Not Listed"

Submit

Figure 4.5. This figure shows the IntoFocus task interface displaying an image in the process of redaction. A subtle difference in the filter level can be seen covering the face of the person on the left. Here, the workers must perform two tasks on five different images. First, they add ellipses on all of the faces in the image. Next, they attempt to select the correct face that matches a person in the image. If they cannot perform the detection task, they must click the *No face Found* button. If they cannot identify the person in the image, they must select the *Any of these* or the *None of these* button.

5. THE PTERODACTYL SYSTEM

Portions of this chapter were copied verbatim from a paper submitted for publication to the 9th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2021).

5.1 Introduction

Machine learning face detection is almost perfect, but there are still some gaps. For example, in face redaction, privacy is the primary concern. Still, some automated systems cannot provide such privacy and recommend not using private information on their systems due to possible security risks [71]. Thus, even though it is minimal, a risk still exists.

For images that require the highest possible privacy, the following approaches could solve the problem:

1. Building a face detector in-house (at the cost of lower than perfect accuracy).
2. Hiring a team to manually redact the images (at the expense of possible disclosure).
3. Adapt a crowd-based privacy-preserving redaction system for in-house use (at the expense of a potentially large team and high price).

This chapter presents the Pterodactyl system, a crowd-based system that uses the AdaptiveFocus filter to allow face detection and prevent identification. The goal of the proposed system and filter is to address the above issues and obstacles.

The contributions of this paper are as follows.

1. We present Pterodactyl, a system for image redaction that applies a set of rules and restrictions to increase early face detection and prevent facial identity disclosure.
2. We present Pterodactyl, a system for image redaction that uses the AdaptiveFocus to combine machine learning and the median filter to allow privacy-preserving face redaction in images at a lower cost.
3. We evaluate the system and filter by comparing crowd-based alternative and automated face detection systems.

5.2 Related Work

Face detection is an active problem in computer vision [72]. In their work, they categorize face detectors into three categories. The first is called CNN Cascade Face detector; they start by creating image pyramids and use a sliding window as input to the CNN [44], [73]. The second is a Region-based Face detector; they propose a region for input to the CNN [74], [75]. Finally, Proposal Free Network, which does not require region proposals [76], [77]. These approaches are all possible in providing the regions to apply the AdaptiveFocus filter. But to rely on such methods would limit the AdaptiveFocus filter to the limitations of the proposed method.

There are several benchmark datasets where researchers are trying to achieve the highest possible success rate. The MogFace face detector [78] achieved the state-of-the-art performance on Wider Face [50], FDDB [5], Pascal Face [79], and AFW [80]. They achieved near-perfect results on FDDB, Pascal Face, and AFW. In the Wider face dataset, they achieved 97.7%, 96.9%, and 93.8% in the easy, medium, and hard image sets, respectively. These results show that machine face detection has not yet achieved perfection. State-of-the-art methods also require large amounts of processing power and training time to build.

Instead, the images used to train the AdaptiveFocus filter will apply the Sliding Window [81] approach, similar to Cascade methods, but for a classification [82] problem instead of a detection problem [83]. The filter would need to assign a filter level appropriate for all the regions, not just the ones that contain a face.

5.3 The Pterodactyl System

The Pterodactyl System is an improvement over the IntoFocus system (chapter 4). In the IntoFocus system, the focus was on achieving the highest possible detection rate with the lowest possible identification rate. The downside of the approach is that it requires time and a large team (at least 21 different people to redact a single image). Because of the required team size, it becomes hard to manage on crowdsourcing platforms. 57.4% (283 of the assignments) of the participants failed their attention check image. Based on the cost of the task, \$0.75, and the associated fees, the cost of the extra assignments would be \$297.15.

The Pterodactyl System focuses on reducing the cost while maintaining a non-significantly different performance in detection and identification.

The Pterodactyl system uses a combination of rules and requirements to ensure the quality of the results. The first requirement is to use at least three crowd workers. That ensures that people of differing detection abilities perform the detection task. The second requirement is assigning a qualification that blocks crowd workers from working on the same images again. This makes sure that crowd workers cannot see the same image more than one once. The third requirement is that each image set is seeded with an image that contains at least two faces. This detects if a crowd worker is not adding ellipses on all the visible faces in the images. It also informs us if the instructions need to be improved.

In addition to the above requirements, the following rules were also added when analyzing the attention check images. 1) All the faces need to be redacted. 2) the number of ellipses added is equal to the number of faces in the image. 3) A single ellipse does not intersect with three or more faces. 4) None of the ellipses goes beyond a 100% increase in the width and height of the face. These added protections are only applied to the attention check images where the ground truth information already exists. These rules were added to make sure that workers are following the task requirements.

5.4 AdaptiveFocus Median Filter

the task was to create an image filter that would allow people to detect but not identify faces in images using the data gathered in the perception study (chapter 4). The AdaptiveFocus filter 5.1 will take an image as input, assign appropriate median filters to obfuscate the faces, and return a thoroughly obfuscated image that allows face detection and prevents face identification. The AdaptiveFocus filter needs to solve the following problems: 1) image restrictions and requirements 2) How to find the locations of faces in the image 3) Which filter level to use.

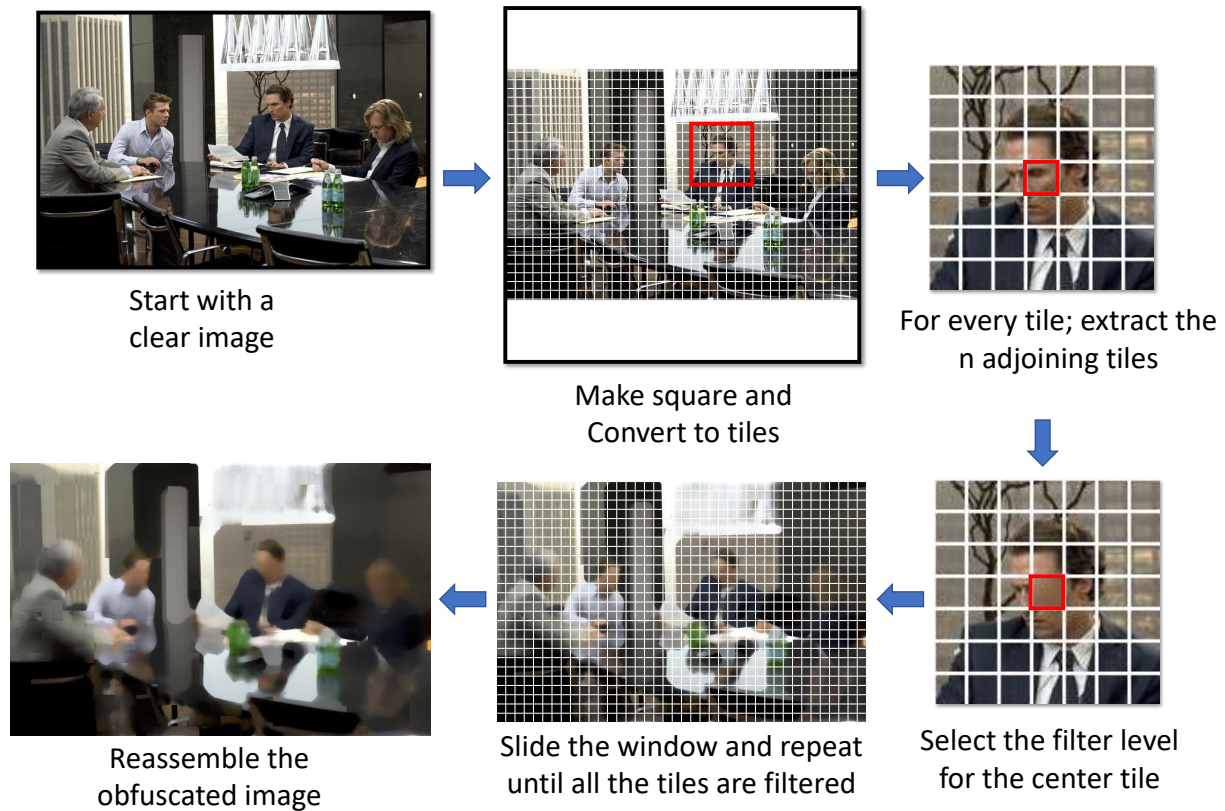


Figure 5.1. This figure shows the process an image takes in the AdaptiveFocus filter. Starting with the clear image, it is resized, made square and each tile is separated. Next, a window of size $n \times n$ centered at each tile slides across the image and assigns a filter level to each of the tiles. Finally, an obfuscated image is created.

5.4.1 Image Requirements

The face perception study performed in chapter 4 required that the image be 640×640 pixels, and the collected data is only valid under that image size. Because the AdaptiveFocus filter is based on the data collected in that study, the AdaptiveFocus filter undergoes the requirements. The AdaptiveFocus filter can be applied to larger image sizes in theory, but a size threshold exists where the AdaptiveFocus filter no longer works for an image.

5.4.2 Finding The Face Locations

The general approach in face detection/redaction systems is to extract regions with a high probability of containing faces and redacting those regions. However, the purpose of the AdaptiveFocus filter is to obfuscate the image so people can perform the detection task. Therefore, the filter does not need to detect the locations of the faces; it needs to assign the correct filter levels to all the regions of the image to allow detection and prevent identification. To solve that problem, the image is segmented into square tiles, and each tile is assigned a filter level based on its content and the content of the surrounding regions. Based on the data collected in the perception study (chapter 4), the smallest detectable face has a size of 209 pixels^2 , and the image has a size requirement of 640×640 . Therefore, we selected the size for tiles to be $16 \times 16 = 256$ pixels, and a small face will fit in one tile. To have a fixed tile space, the size of all the images is increased (in width or height), so the exact size will be 640×640 . Thus, each image will contain 40×40 tiles, and each of the tiles will be evaluated and assigned a filter level.

5.4.3 Selecting The Filter Levels

The filter level selection is based on the face perception study of median filtered faces (chapter 4). The study presented people with median filtered images containing people and tasked them with detecting and identifying the faces. The AdaptiveFocus filter utilizes the detection point of inflection where 100% of the study participants could detect the faces as the appropriate filter level for a specific face. For example, a face has been tested three times with median filters 25, 29, and 31 (ksize), with detection of 100%, 100%, and 92%, sequentially. The filter level selected for that face will be 29. A filter map is created for each image (figure 5.2) using the inflection points of all the faces in the images. The filter map shows the filter level to use for each face to be detectable and not identifiable. The color range for the filter map goes from black (no filter needed) to white (highest filter needed).

We train a Convolutional Neural Network (CNN) on the tiles containing faces as input. It outputs the filter level required for each tile. The neural network needs to answer the question:

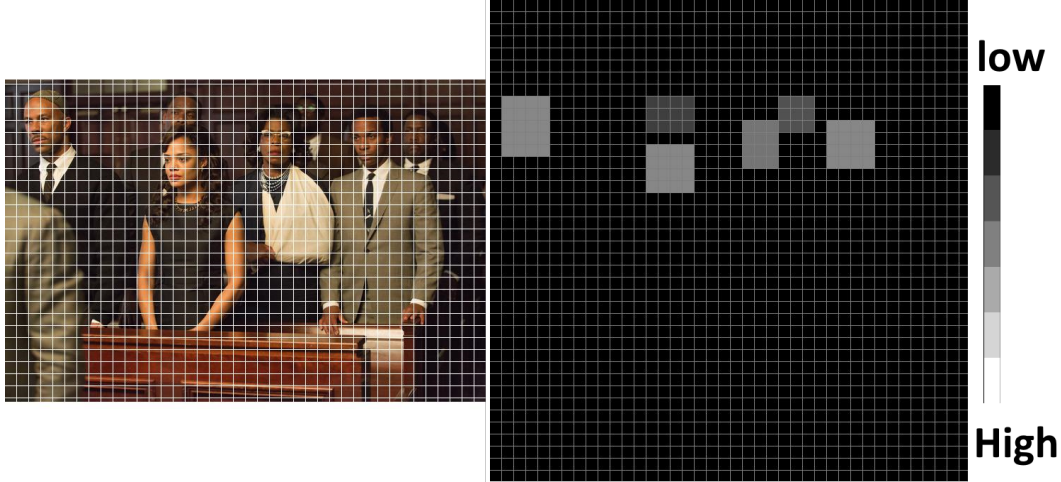


Figure 5.2. This figure shows an example of a filter map. The left image shows the image to be redacted where the white lines separate the tiles. The image on the right is the ground truth filter map for that image. The black region means that there are no faces in that region. The brighter regions specify the filter levels required for that face to be detectable but not identifiable. The darker the region, the lower the filter required, and the brighter the region, the higher the filter required.

If there was a face in this tile, what filter level is needed, to allow detection and prevent identification?

If the tile has a face, the proposed filter level will allow that face to be detectable but not identifiable. If the tile does not contain a face, it will still suggest a filter level that would allow a face to be detectable and not identifiable. We chose this approach because current face detection algorithms [78] are not yet perfect. With this approach, the filter will obfuscate the entire image and allow face detection on all the regions that might contain a face, and obscure the regions that do not have faces.

For the CNN to correctly select the correct filter level, it will need to analyze the tile to be classified and n adjacent tiles. We evaluated different values of n . We split the infliction point data into a training and validation set. We trained the neural networks starting from $n = 3$ to find the value of n where the classification accuracy no longer increases. Based on the results of the CNN training, it was found that before $n = 7$, there were significant accuracy increases, and after $n = 7$, the increase was less than 0.5%. Thus, the chosen value

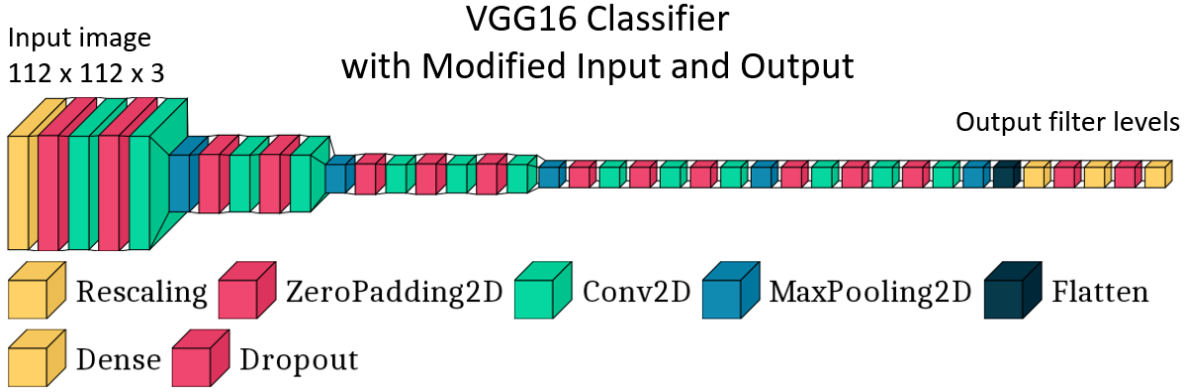


Figure 5.3. This figure shows the convolutional neural network used in the AdaptiveFocus filter. It inputs images of size $112 \times 112 \times 3$ and outputs the probability of each of the filter levels being correct based on the content of the image.

for n was 7. Since each tile was 16×16 pixels, the classification model will input square images with a width and height of $16 \times 7 = 112$ pixels.

The CNN model used for the classification task was based on the VGG16 classifier (figure 5.3) [84]. It was compiled using the categorical cross-entropy loss function and the stochastic gradient descent [85] optimizer. The training occurred over 100 epochs. The dataset used was the infliction point dataset, and it was split into 75% training and 25% validation, with a total of 9 different classes. The classes were faces detectable at 7, 13, 17, 23, 29, 35, 41, 53, and 85. The images were augmented using horizontal and vertical flip, rotations [86], Gaussian blur, noise injection [87], and applying multiple augmentations together. Photometric [88] were considered, but because lighting and color might affect the appropriate filter levels for a tile, they were avoided. The dataset started with 60 images containing 185 faces. With the augmentation, we generated a total of 219876 images across the nine classes. Training the CNN took 32.4 days (777.8 hours). The models were all trained on a CPU-only machine without a dedicated graphics adapter or specialized equipment for neural network processing. There was no class for no faces detected because the face detection and redaction are performed by crowd workers, not the neural network.

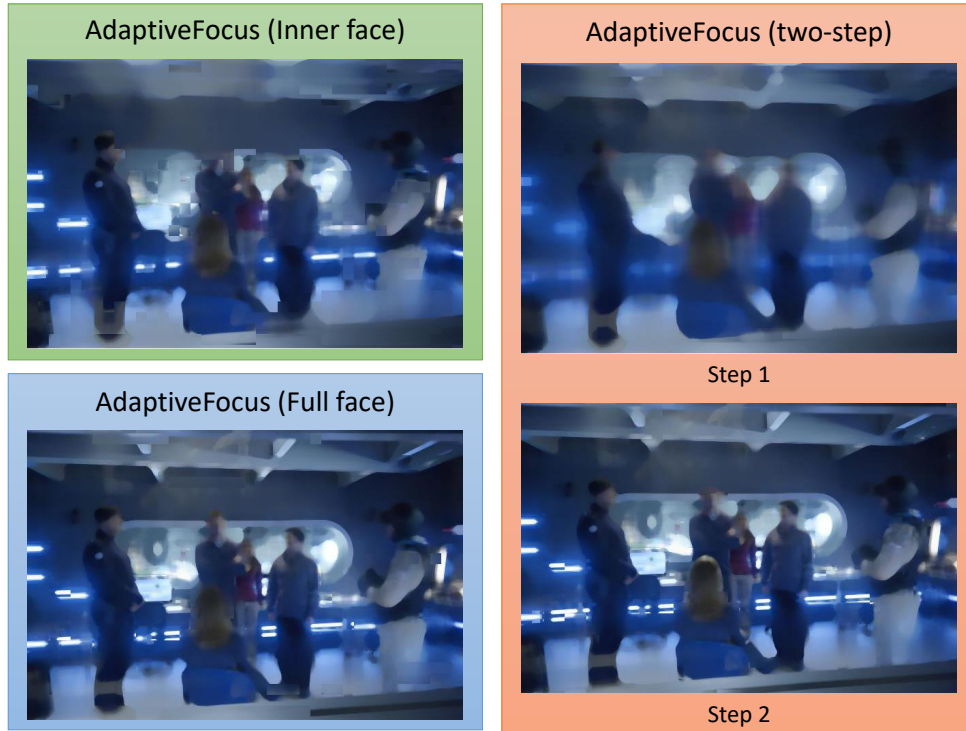


Figure 5.4. This figure shows the different AdaptiveFocus methods. The top left is built based on the AdaptiveFocus (inner face) method, the bottom left is based on the AdaptiveFocus (full-face method), and the right are the two steps for the AdaptiveFocus (two-step) method.

Moreover, there were three different variants of the AdaptiveFocus filter (figure 5.4). The first variant contained images of the full-face, including hair, beards, and hats (figure 5.5). The second variant contained only the inner face, excluding any hair, beards, or hats (figure 5.5). This final variant had two separate CNNs, one for the higher filter levels (29, 35, 41, and 85) and a second for the lower filter levels (7, 13, 17, and 23). This variant applies the AdaptiveFocus filter into the IntoFocus process (chapter 3) to increase the probability of face detection. The variants were created by applying transfer learning on the full-face variant for 30 epochs.

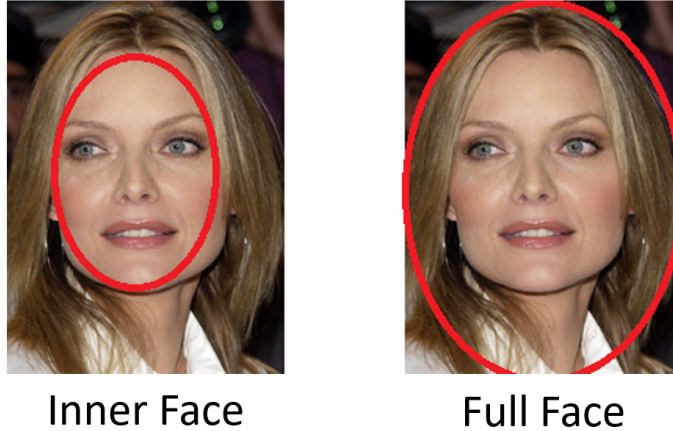


Figure 5.5. This figure shows the difference between the inner face and full-face images. In the inner face, only the inside of the face is used. In the full-face, the full face is covered including hair, beard, and hats.

5.5 Experiment Setup

The experiment (Figure 4.5) to evaluate the Pterodactyl system was designed to follow the experiment performed to evaluate the IntoFocus method (chapter 4). The experiment aims to evaluate the Pterodactyl system combined with the AdaptiveFocus filter and compare the results with the 7-stage IntoFocus method and system. We hypothesize that the Pterodactyl system will maintain a non-significantly different result in identification and detection with the highest method.

5.5.1 AdaptiveFocus Filters

In the experiment, we evaluate three different models of the AdaptiveFocus filters. The first is trained on detecting face regions, including hair, hats, and beards. The second model is trained on detecting the inner face regions, excluding hair. The third is trained to work on two stages, where the faces detected on the first stage are redacted before progressing to the second stage. Each of the filters is presented to three different crowd workers. Any crowd worker who worked on a specific image set could not work on that set again for all the available conditions.

Add ellipses on ALL faces, including hair. Be sure you have covered all parts of the eyes, nose, lips, eyebrows, forehead, cheeks, ears, chin. From the faces on the right, select the faces that can be seen in the image.

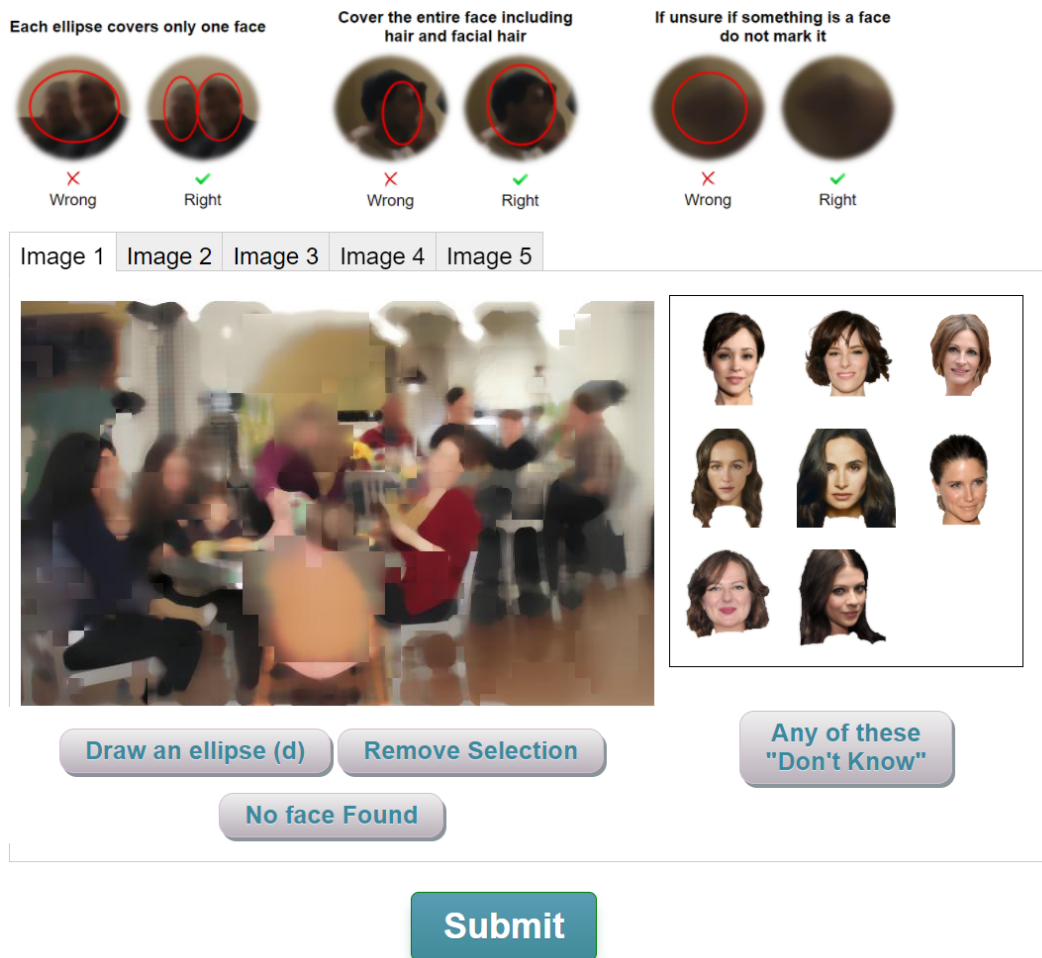


Figure 5.6. This figure shows the interface used to evaluate the Pterodactyl system and the AdaptiveFocus filter. Participants were tasked with detecting all the faces in five filtered images and identifying a face in each of the images.

5.5.2 Control Conditions

The control in this experiment is the IntoFocus method and system (chapter 4). The system proposes using a 7-stage redaction process, starting from the highest filter level, and iteratively reducing the filter level, while asking the crowd to redact faces in the images. Before the start of each stage, all the faces that crowd workers previously detected are redacted. Then, the process is repeated seven times until the entire image is redacted.

5.5.3 Face Identification

In each image, we present participants with a set of eight faces, one of which is a subject located in the obfuscated image. The goal of the methods being evaluated is to hide the identities of the faces from the participants. Out of the eight faces, only one face is located in each of the images. The participants are allowed to select multiple faces if they can narrow the list of possible faces. If the participants could not narrow down the list of faces, they were asked to choose the "don't know" button.

The eight faces have similar facial features (hair color, hair length, skin tone) to provide k-anonymity [36] for those faces. This ensures that if a participant identified a face, it was because the facial features were visible.

5.5.4 Face Detection

To measure face detection, the participants need to add ellipses that encompass the inner regions of each face. Covering part of the face did not count towards a detected face. This was to ensure that the participant was able to actually detect the face and not just adding ellipses to the images.

5.5.5 Participants

We divided the experimentation into two categories. The first category employed crowd workers on Amazon Mechanical Turk (AMT) [68] to perform the redaction task. The second category engaged in-person participants to complete the redaction task.

In the first category, crowd workers must have a 90% task success rate to participate. Each Human Intelligence Task (HIT) rewarded \$0.75, and the requirement was to redact only five images. There were 248 unique crowd workers. The cost of the experiment was \$621, with an average of \$9.32 per hour. The task took an average time of 4.8 minutes to complete.

In the second category, the task for each participant was to redact 50 images. The task took the participants an average of 21 minutes to complete. There were only three participants.

One for the AdaptiveFocus (full face) method and two for the AdaptiveFocus (two step) method.

5.5.6 Attention Check

The attention check proposed in chapter 3 was to ensure that participants perform the task correctly. In each HIT, an image—where the detection task was purposefully made trivial—was seeded [59], [60]. Participants who do not add ellipses on all the faces in the test image; have their work discarded and replaced by another participant. We made modifications to the requirements in the Pterodactyl system. The new requirement is that participants are required to add ellipses on all the faces in the image instead of adding a single ellipse. We also added rules for the ellipses, as stated in the previous section.

5.5.7 Dataset

In the experiment, we used the IMDB image dataset [62]. The experiment contained ten different HITs; each hit contained five images, four of which were being evaluated, and the other was to test for task correctness. Thus, making a total of 40 images for evaluation and ten images for task correctness.

5.5.8 Experiment Task

Participants are required to perform two tasks. First, they needed to add ellipses on all the visible faces. Second, they needed to attempt to try to identify one of the faces in the image. When a participant performs the task on a specific image set, they can not perform the task on that image set again, even if the filter method is changed; this is achieved by assigning a qualification on AMT that blocks them from accessing HITs containing the same images.

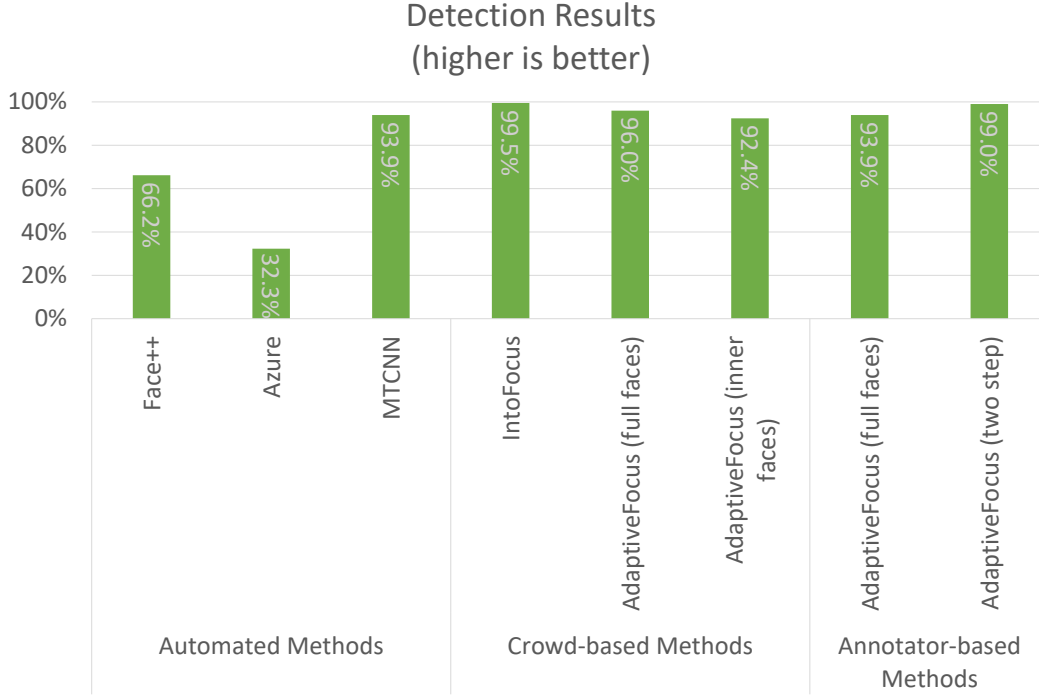


Figure 5.7. This figure shows the detection results of all the methods.

5.5.9 Evaluation

All the crowd-based methods will be evaluated using the Pterodactyl system, except the IntoFocus system was evaluated with both. The AdaptiveFocus filter was designed to perform with annotators and crowd workers. To evaluate, crowd workers were hired on Amazon Mechanical Turk, and annotators were hired for an in-person evaluation. The results were also compared with the IntoFocus method (chapter 4), Microsoft Azure’s face detector [70], Face++ [89], and the MTCNN pre-trained face detector [90].

5.6 Results

The evaluation is separated into the following sections: detection, identification, time, and cost. The first part compares all the methods, with the AdaptiveFocus methods using the Pterodactyl system and the IntoFocus method using the IntoFocus system. The second part compares the AdaptiveFocus method and the IntoFocus method (chapter 4) using the IntoFocus system and the Pterodactyl system.

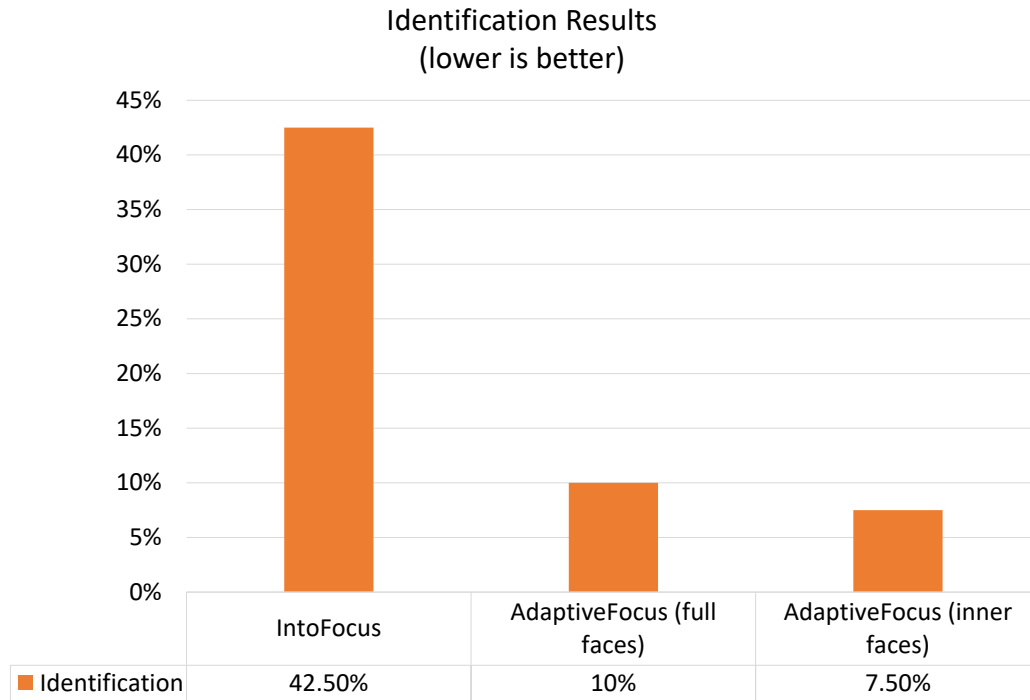


Figure 5.8. This figure shows the identification results of all the methods. The results show that the AdaptiveFocus methods had the lowest face identification rates and the IntoFocus method had a significantly high disclosure rate.

5.6.1 Detection

Detection in the evaluation is adding an ellipse or a bounding box covering the entire face (figure 5.7). The IntoFocus method had the highest detection rate of all the methods, missing only one face. Followed by the annotator-based AdaptiveFocus full-face method which only missed two faces. The crowd-based AdaptiveFocus full-face method was followed with 96% detection. Then the MTCNN face detector and the annotator-based AdaptiveFocus full-face method had the same detection rate. The crowd-based AdaptiveFocus inner faces method had the lowest detection among the AdaptiveFocus methods, and the remaining automated methods had the lowest detections. The annotator-based methods only used a single annotator, and that had the negative effect of not having multiple views on a single image. The two annotators in the AdaptiveFocus (two-step) method did not detect faces that were detected in the other AdaptiveFocus methods.

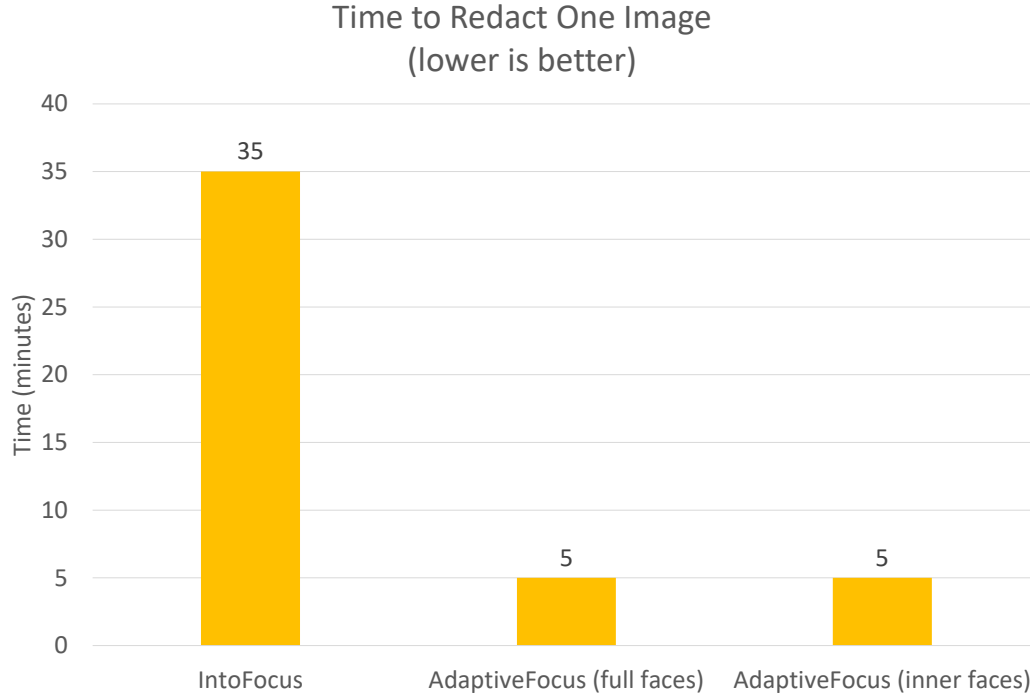


Figure 5.9. This figure shows the time it takes to redact an image using IntoFocus and AdaptiveFocus. The results show that the methods using the AdaptiveFocus filter are seven times faster than methods using IntoFocus.

5.6.2 Identification

Identification in the evaluation is the number of images where the faces were identified during the experiment (figure 5.8). The AdaptiveFocus-based methods had lower significantly lower identification rates than the IntoFocus method. The results show that the face were identified in 42.5% of the test images when using the IntoFocus method.

5.6.3 Time

The time is the evaluation of the time needed to detect/redact all the faces. Methods using the AdaptiveFocus filter generate the results seven times faster than methods using the IntoFocus process. The problem with the IntoFocus process is that each stage is reliant on the previous stages. Since crowd workers are expected to complete each stage in five minutes,

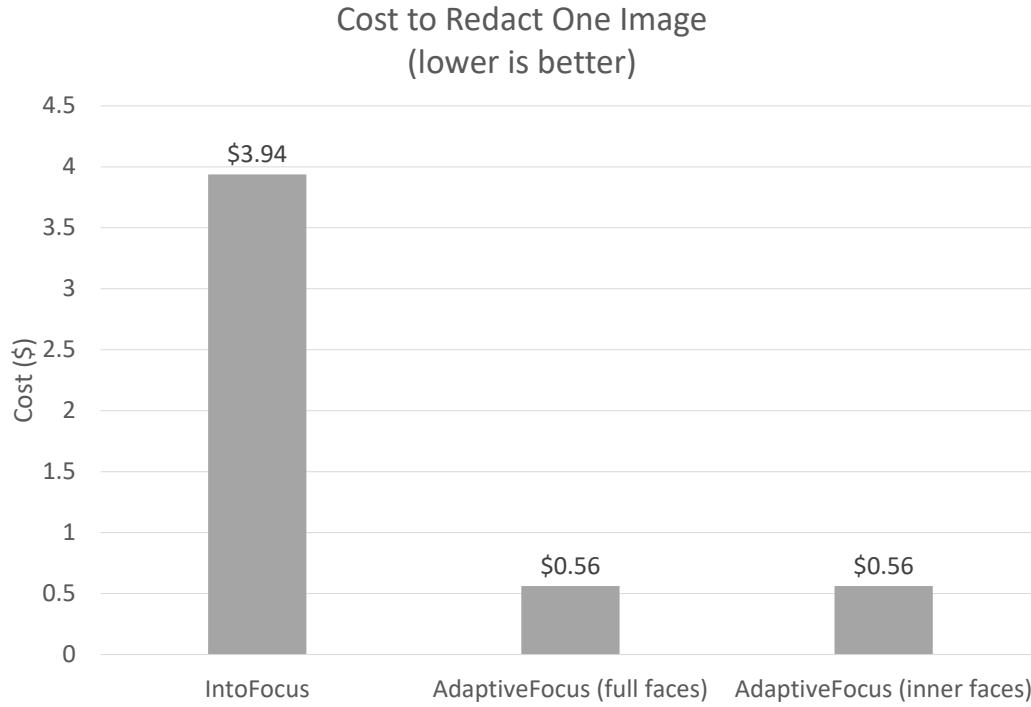


Figure 5.10. This figure shows the cost to redact an image using the IntoFocus method and the AdaptiveFocus methods.

IntoFocus requires seven stages to be performed in series. Thus, increasing the time required for the system to generate a redacted image.

5.6.4 Cost

The cost in the evaluation is how much it costs to perform the detection/redaction of a single image. The AdaptiveFocus method is seven times more cost-effective than the IntoFocus method. The reduced cost is the number of crowd workers required to finish the process of redaction. AdaptiveFocus-based methods use three crowd workers to redact an image, while the IntoFocus method uses 21 crowd workers.

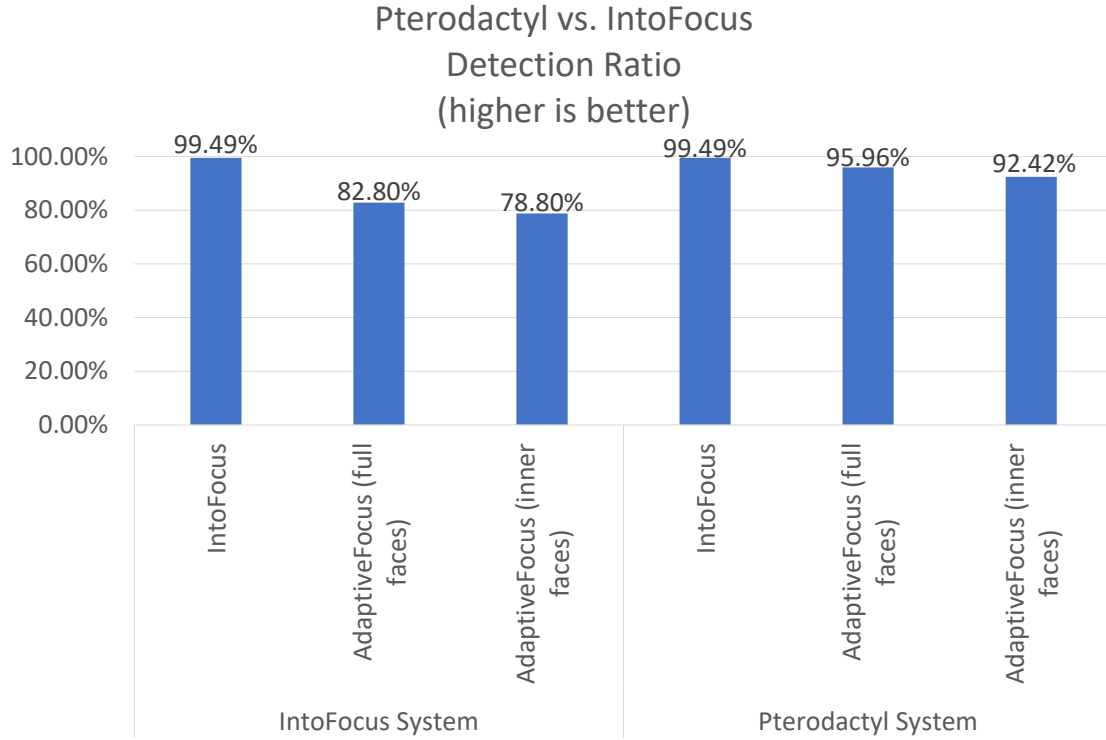


Figure 5.11. This figure shows the detection results of the crowd-based methods using the Pterodactyl system and the IntoFocus system. The detection rate in the IntoFocus method does not change when the method changes. In contrast, the AdaptiveFocus methods observe a significant increase in detection performance.

5.6.5 System Analysis

This section evaluates the comparison between the IntoFocus system (chapter 3) and the Pterodactyl system. The two systems are compared in terms of face detection and face identification.

Detection

The detection results of the IntoFocus system and the Pterodactyl system (figure 5.11) show that, even when the system is changed, the IntoFocus detection results do not change. In comparison, the AdaptiveFocus methods have a significant increase in detection for both

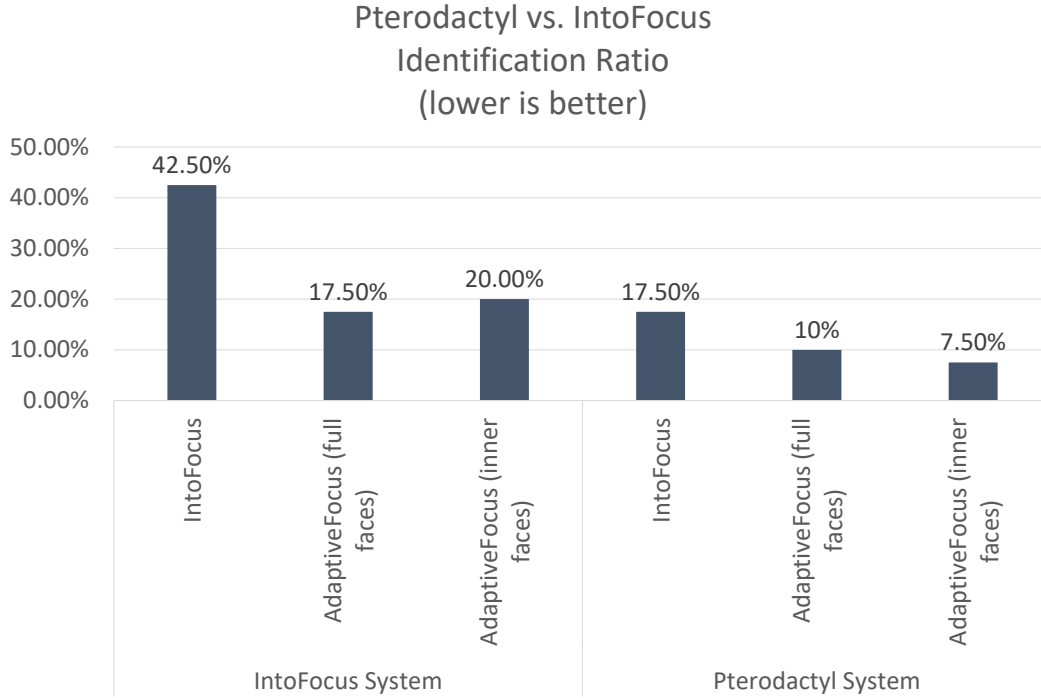


Figure 5.12. This figure shows the identification results of the crowd-based methods using the Pterodactyl system and the IntoFocus system. All the methods observe a significant decrease in disclosure rates.

methods. These results show the importance of the Pterodactyl system for the AdaptiveFocus system, but With the IntoFocus process, it does not affect the detection results.

Identification

The identification results (figure 5.12) show that the Pterodactyl system significantly improves the results of all the methods. With the IntoFocus method gaining the most significant decrease in terms of identification. However, even when using the Pterodactyl system, the IntoFocus method still has a higher disclosure rate than the AdaptiveFocus-based methods.

The above results show that with the addition of the above rules, the AdaptiveFocus-based methods have a significant increase in detection (figure 5.11) and a significant decrease in identification (figure 5.12). While the IntoFocus method did not improve in terms of detection, the identification results decreased by 25%. In the previous chapters, the images

went through several iterations of redaction. This process allowed the faces that were not adequately redacted in the first stage they were detectable to be redacted in the following stages at a lower filter level. However, this allows some of the faces to be identifiable to some of the crowd workers. This can be seen when comparing the identification and detection results for the IntoFocus method. It shows that crowd workers identified ten faces that would not have been identifiable if those faces were redacted at an earlier stage. Nevertheless, the number of faces detected remains the same. This shows that some faces can be detected at a later stage where they are identifiable, and the image is clearer because of the iterative process. On the other hand, the AdaptiveFocus methods are not iterative processes (except for the two-step method), so if a face is not detected in the initial run, it will not be detected at a later stage. That is why it can be seen that the ratio of faces detected decreases significantly with the added rules. This shows that with the AdaptiveFocus methods, a single crowd worker not following the task requirements can cause faces not to be redacted. However, with the addition of the rules proposed in the Pterodactyl system, the results improve significantly in both identification and detection.

5.7 Discussion

The results show that the Automated methods have the lowest time and cost, but detecting the main person in an image and their detection accuracy was not perfect. In the crowd-based methods, IntoFocus has the highest detection overall, but the number of faces identified, time, and cost per image were very high. On the other hand, the crowd-based AdaptiveFocus methods have significantly lower cost and identification rates than IntoFocus, but a higher detection rate than the Automated methods. Finally, the Annotator-based AdaptiveFocus methods had detections close to the IntoFocus method and cost much less because the annotators did not follow the first rule of Pterodactyl (at least three annotators). In the evaluation of the Annotator-based methods, only one person was hired for each of the methods. Adding an extra person would improve the odds of the faces being detected but increase the cost.

There was no increase in detection between the IntoFocus (IntoFocus system) and the IntoFocus (Pterodactyl system). But there was a significant increase in face identity preservation. This is because the strict rules have increased the number of faces redacted at earlier stages, thus decreasing the chances of disclosure.

In the AdaptiveFocus (full-faces), crowd workers had more face detections than the single annotator. Another observation was that the faces that the crowd workers did not detect were also not detected by the annotator. These results show that since a total of three crowd workers had their detections aggregated, they were able to achieve a higher detection rate than a single annotator. This confirms the observation in the previous chapter 4 that people have differing abilities when detecting faces in filtered images.

AdaptiveFocus (full-face) vs. AdaptiveFocus (inner face), the difference between the methods is that in the training set for the CNN, one used the full-face, including hair, beards, and hats. While the other only used the inner face without the hair, beards, and hats. Because the training set in the inner face method only includes faces and the lack of other features in the images, the filter could not accurately filter the surrounding regions, making the task harder for the crowd workers to accurately detect the faces.

In creating the AdaptiveFocus filter, the images were split into 40×40 tiles. This approach was the most viable because the data set was small (185 faces), and one category only had four faces. This was because of our condition of the point of infliction. If a face did not reach that point, it was not used in the training/validation.

In creating the filter, there were multiple possible approaches. The most viable was to use a face detector to redact the detectable faces and then submit the image to the crowd workers. This approach does work and is not a flawed approach. From the results of the AdaptiveFocus (full faces) method, none of the missed faces were detected by the MTCNN [90] face detector.

Another approach was to incorporate the MTCNN [90] face detector into the filter and select the regions for applying the AdaptiveFocus filter. With this approach, the remaining parts of the images would have a static predefined filter. However, when this approach was attempted, a single predefined filter could not obfuscate all the remaining faces and prevent identification.

5.8 Conclusion

This chapter presented the Pterodactyl system, a system that uses a set of rules to increase early detection of faces in multi-stage redaction methods and improve single-stage systems' identification and detection results. The experiment shows a significant increase in the number of faces detected and a decrease in the number of faces identified by the AdaptiveFocus-based methods. It also shows a significant decrease in the number of faces identified in the IntoFocus method.

This chapter also presented the AdaptiveFocus filter, a sliding window-based filter that selects the appropriate median filter level for each region of an image with the sole task of allowing faces to be detected and prevent those faces from being identified during the redact process. The AdaptiveFocus filter aims to determine the right filter level for each tile to maintain the guarantees of privacy and efficacy.

6. CONCLUSION

In this dissertation, I discussed the systems and methods I created to tackle the problem of utilizing crowd workers to redact faces from images without identifying said faces.

6.1 IntoFocus

In chapter 3, I described the IntoFocus system for crowd-powered face detection and redaction. The goal of the IntoFocus system was to present crowd workers with median filtered images and ask them to detect the locations of the faces. The image starts with a high filter level and progressively decreases the filter level. Before each stage, previously detected faces are redacted, and the filter level is reduced in the remaining parts of the image. At the end of the IntoFocus method, an image is generated with all the facial identities redacted.

The IntoFocus system implemented a set of requirements to increase the quality of the crowd workers by adding an attention check image to ensure that all the crowd workers correctly performed the task. Where crowd workers were only required to add an ellipse on the image, the ellipse did not need to cover a face and the use of three separate workers on each image to avoid unreliable or intentional sabotage. When combined, crowd workers were able to detect many of the faces in the images.

Two experiments were performed to evaluate. The evaluation compared the IntoFocus method with static median filter levels and Amazon Rekognition’s face detection system. The first experiment used a dataset of actors. The IntoFocus method had one of the highest results in detection and the lowest score in identification (meaning the least number of faces were identified). Crowd workers attributed the movie scene to be familiar to them because it allowed them to identify the faces. For that reason, the second experiment used photographs of people on the Purdue campus. The photographs allowed us to collect images where the gaps between the IntoFocus method and the static filters would be visible. Again, the IntoFocus method had one of the highest detection ratios and the lowest identification ratio. In both experiments, the IntoFocus method was able to detect a higher percentage of faces when compared with Amazon’s Rekognition. When comparing the results between IntoFocus and

the control methods, IntoFocus had a balance between identification and detection, where the control methods could only achieve one.

6.2 Human Perception of Median Filtered Faces

The second work started with the plan to improve the IntoFocus method by finding the exact filter levels that would allow faces to be detectable and not identifiable. This work began with a perception study of median filtered faces. This work aimed to find the gap in filter levels between when faces become detectable and identifiable. This was accomplished by performing a binary search to find the filter level where the number of people who can detect a face decreased from 100%. That filter level is where the face starts to become undetectable. The same was performed for identification, finding the filter level that prevented all the study participants from identifying a face.

After collecting that data, we improved the IntoFocus system using the data to select the filter levels that maximized detection and minimized identification. The filter level selection was performed using the 2nd percentile for detection and the 98th percentile for identification for all the faces based on the face size (width \times height). Applying a polynomial regression to estimate the lines representing detection and identification and using the lines to estimate the filter levels.

The extracted filter levels were used to create the improved IntoFocus method. The attention check requirements were modified, requiring crowd workers to redact one of the faces in the images correctly. Also, at this stage, faces presented for identification conform to the 8-anonymity rule [36], where the quasi-identifiers were gender, skin tone, and hair color. The improved IntoFocus method used seven stages to redact all the faces in an image. An experiment to verify the results of the modified IntoFocus method was performed, and the IntoFocus method only missed 1.3% of the faces in detection, and only 0.83% of the faces were correctly identified. The filters in this method considered that participants might guess because of the narrow sample given to them. That allowed the method to reduce the identification rate even further.

6.3 The AdaptiveFocus Filter And The Pterodactyl System

The AdaptiveFocus Filter takes the results of the perception study and uses it to create an image filter. The goal of the filter is to reduce the cost of the IntoFocus system and enhance the performance of the crowd workers. The first change was which data was used for detection and identification. Instead of using the percentiles (like the previous chapter) instead, the inflection points were used. The inflection points were calculated as the point right before the detection starts to decrease from 100%. For example, if a face was tested at three filter levels, 21, 23, and 25. With detection rates of 100%, 92%, and 87%. The filter level of 21 was used. Applying that method to all the faces in the perception study gives us a dataset that contains each face, its location in an image, and the filter level to allow that face to be detectable and prevent it from being identified.

Then each image is separated into 40×40 tiles, and each of the tiles is assigned a filter level based on a Convolutional Neural Network (CNN). Training the CNN on the tiles of the faces from the perception study and giving it a window of tiles instead of only the tile that needs to be classified. Where the tile being classified is centered in the window. This allowed the network to gain information about the surrounding area around the tile to increase the classification accuracy. Also, the CNN would only be training on tiles that contain faces, and the filter level associated with that tile will be the filter level needed for that face to be detectable and not identifiable because even the best available face detection methods do not meet our requirements for accuracy. If the filter is applied to only regions that a face detector provides, then the filter will undergo the same limitations as the face detector. Namely, if a face is not detectable by the detector, it will not be filtered. Another reason was that crowd workers, not the filter, would perform the face detection. With only training on face regions, the classifier will need to answer the following question:

If there was a face in this tile, what filter level is needed, to allow detection and prevent identification?

In the experiment, there were three variants of the AdaptiveFocus filter. The first was trained on the full faces, the second was trained on the inside of the faces only, and the third, using a two-step process, similar to IntoFocus, but with only two stages. The three filters

were compared with the IntoFocus method and automated methods. The IntoFocus method had the highest rate in terms of detection, and the two-step method followed. On the other hand, in terms of identification, the AdaptiveFocus had the lowest identification rates. The purpose of the AdaptiveFocus filter was to reduce the cost of the IntoFocus process while yielding similar results in terms of detection and identification.

The Pterodactyl system was developed to increase the quality of work by the crowd workers. During the experiments in chapter 4, 57.4% of workers had failed the attention check requirement of adding an ellipse on one of the faces. This highlighted a problem with the instructions and the attention check requirements. The Pterodactyl system adds additional rules for the crowd work to either be accepted or replaced. The first rule was for the attention check image (the image containing reduced filter levels and evaluating the crowd worker's understanding of the task). All the faces need to be redacted. The second rule, the number of ellipses added, must equal the number of faces in the image. The third, an ellipse must not intersect with more than two faces. The final rule is that none of the ellipses goes beyond 100% of the size of the face. If a crowd worker breaks one of these rules, their work is replaced by another crowd worker.

A comparison between the IntoFocus system and the Pterodactyl system was performed (figures 5.11 and 5.12). The results showed that for the IntoFocus method, the Pterodactyl system reduced the face identification rate by 25%. For the AdaptiveFocus (full faces) and AdaptiveFocus (inner faces), it increased the face detection rate by 13.2% and 13.6% respectively, and decreased the face identification rate by 7.5% and 12.5% respectively.

7. FUTURE WORK

The latest improvements in the Pterodactyl system dramatically increases the quality of the data gathered from crowd workers. However, still, many crowd workers have not been performing the task correctly. Although others replace their work, they pose a risk to the system. The issue arises from the crowd workers not following or understanding the instructions. The most direct approach is to train the crowd workers on how the task is expected to be performed and warn them that their task will be rejected if these instructions are not followed. Nevertheless, rejection is akin to an F grade in a transcript, and it stays with the crowd worker, and it can affect the work they can accept. That was why this approach was avoided. An alternative is to block the crowd workers that are not following the instructions from performing any of our tasks. Each of these has consequences and needs to be explored thoroughly.

An issue with the AdaptiveFocus filter was that applying the filter on one image took approximately 200 seconds¹. The current filter uses a tile size of 16×16 if that size is to be reduced to 8×8 , that would mean the number of computations needed for a single image would double, meaning it would need approximately 400 seconds to finish applying the filter to a single image (this is assuming we keep the input image size the same). Reducing the size and complexity of the convolutional neural network (CNN) will help in reducing the time needed for an image to be filtered. A new model recently released [91] can help address this problem because it reduces the training time and network size and still achieves near state-of-the-art classification.

The current model has only been tested on images. By performing the modifications above, it is possible to apply the same AdaptiveFocus filter to videos. However, the latency problem described above needs to be solved first for the filter to be feasible for videos.

Another issue with the AdaptiveFocus filter was that it could not match the IntoFocus system in face detection. In some cases, the faces were detectable (to us as examiners), and they could not be detected in others. An approach to solve the issue of the detectable faces to us is to increase the number of crowd workers that perform the redaction. But such solutions

¹[↑](#)The model is processed on a CPU-only machine without equipment dedicated for enhanced neural network processing.



Figure 7.1. This image shows a face in the top right behind the main face that was not detectable by the crowd workers in most of the AdaptiveFocus methods.

become a guessing game, "Will we get a crowd worker that excels at this task?" and that is not an option. For example figure 7.1 is an example of a face that was not detected by most of the AdaptiveFocus methods (except for the two-step method, which reduces the filter significantly in the second step). This issue shows that the model can benefit from a larger dataset. The current dataset only contained 185 faces from 60 images. The problem faced during the initial perception study was acquiring as accurate data as possible. People

were not used to extracting faces from filtered images. So to collect more data, an improved experiment that reduces the number of participants needed for each step is required.

The IntoFocus method had the highest accuracy in face detection across all the methods tested in the results of the AdaptiveFocus filter (table 5.7). The method would improve with the inclusion of additional information in the filter extraction phase. It is currently based solely on face sizes. Other factors exist in determining an appropriate filter level to uses.

REFERENCES

- [1] J. Howe, “The Rise of Crowdsourcing,” *Wired*, Jun. 2006, ISSN: 1059-1028. [Online]. Available: <https://www.wired.com/2006/06/crowds/>.
- [2] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton, “The Future of Crowd Work,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW ’13, New York, NY, USA: ACM, 2013, pp. 1301–1318, ISBN: 978-1-4503-1331-5. DOI: [10.1145/2441776.2441923](https://doi.org/10.1145/2441776.2441923). [Online]. Available: <http://doi.acm.org/10.1145/2441776.2441923>.
- [3] P. J. Phillips, A. N. Yates, Y. Hu, C. A. Hahn, E. Noyes, K. Jackson, J. G. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J.-C. Chen, C. D. Castillo, R. Chellappa, D. White, and A. J. O’Toole, “Face recognition accuracy of forensic examiners, super-recognizers, and face recognition algorithms,” en, *Proceedings of the National Academy of Sciences*, vol. 115, no. 24, pp. 6171–6176, Jun. 2018, ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1721355115](https://doi.org/10.1073/pnas.1721355115). [Online]. Available: <https://www.pnas.org/content/115/24/6171>.
- [4] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, “Pushing the Limits of Unconstrained Face Detection: A Challenge Dataset and Baseline Results,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, ISSN: 2474-9680, Oct. 2018, pp. 1–10. DOI: [10.1109/BTAS.2018.8698561](https://doi.org/10.1109/BTAS.2018.8698561).
- [5] V. Jain and E. Learned-Miller, “Fddb: A benchmark for face detection in unconstrained settings. University of Massachusetts,” *Amherst, Tech. Rep. UM-CS-2010-009*, vol. 2, no. 7, p. 8, 2010.
- [6] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, “CMS-RCNN: Contextual Multi-Scale Region-Based CNN for Unconstrained Face Detection,” en, in *Deep Learning for Biometrics*, ser. Advances in Computer Vision and Pattern Recognition, B. Bhanu and A. Kumar, Eds., Cham: Springer International Publishing, 2017, pp. 57–79, ISBN: 978-3-319-61657-5. DOI: . [Online]. Available: .
- [7] A. Alshaibani, S. Carrell, L.-H. Tseng, J. Shin, and A. Quinn, “Privacy-Preserving Face Redaction Using Crowdsourcing,” en, *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 8, no. 1, pp. 13–22, Oct. 2020, Number: 1. [Online]. Available: <https://ojs.aaai.org/index.php/HCOMP/article/view/7459>.

- [8] A. J. Quinn and B. B. Bederson, “Human computation: A survey and taxonomy of a growing field,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’11, Vancouver, BC, Canada: Association for Computing Machinery, May 2011, pp. 1403–1412, ISBN: 978-1-4503-0228-9. DOI: [10.1145/1978942.1979148](https://doi.org/10.1145/1978942.1979148). [Online]. Available: <http://doi.org/10.1145/1978942.1979148>.
- [9] E. D. Karnin, E. Walach, and T. Drory, “Crowdsourcing in the Document Processing Practice,” en, in *Current Trends in Web Engineering*, F. Daniel and F. M. Facca, Eds., ser. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 2010, pp. 408–411, ISBN: 978-3-642-16985-4. DOI: .
- [10] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh, “VizWiz: Nearly Real-time Answers to Visual Questions,” in *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’10, New York, NY, USA: ACM, 2010, pp. 333–342, ISBN: 978-1-4503-0271-5. DOI: [10.1145/1866029.1866080](https://doi.org/10.1145/1866029.1866080). [Online]. Available: <http://doi.acm.org/10.1145/1866029.1866080>.
- [11] A. Sorokin, D. Berenson, S. S. Srinivasa, and M. Hebert, “People helping robots helping people: Crowdsourcing for grasping novel objects,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, Oct. 2010, pp. 2117–2122. DOI: [10.1109/IROS.2010.5650464](https://doi.org/10.1109/IROS.2010.5650464).
- [12] S. Swaminathan, R. Fok, F. Chen, T.-H. (Huang, I. Lin, R. Jadvani, W. S. Lasecki, and J. P. Bigham, “WearMail: On-the-Go Access to Information in Your Email with a Privacy-Preserving Human Computation Workflow,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’17, Québec City, QC, Canada: Association for Computing Machinery, Oct. 2017, pp. 807–815, ISBN: 978-1-4503-4981-9. DOI: [10.1145/3126594.3126603](https://doi.org/10.1145/3126594.3126603). [Online]. Available: <http://doi.org/10.1145/3126594.3126603>.
- [13] J. Deng, J. Krause, and L. Fei-Fei, “Fine-Grained Crowdsourcing for Fine-Grained Recognition,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013, pp. 580–587. DOI: [10.1109/CVPR.2013.81](https://doi.org/10.1109/CVPR.2013.81).
- [14] W. S. Lasecki, Y. C. Song, H. Kautz, and J. P. Bigham, “Real-time Crowd Labeling for Deployable Activity Recognition,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW ’13, New York, NY, USA: ACM, 2013, pp. 1203–1212, ISBN: 978-1-4503-1331-5. DOI: [10.1145/2441776.2441912](https://doi.org/10.1145/2441776.2441912). [Online]. Available: <http://doi.acm.org/10.1145/2441776.2441912>.
- [15] W. Lasecki, M. Gordon, J. Teevan, E. Kamar, and J. Bigham, “Preserving Privacy in Crowd-Powered Systems,” in *In AAMAS 2015 Workshop on Human-Agent Interaction Design and Models*, ser. HAIDM 2015, Istanbul, Turkey, 2015.

- [16] H. Kaur, M. Gordon, Y. Yang, J. P. Bigham, J. Teevan, E. Kamar, and W. S. Lasecki, “CrowdMask: Using Crowds to Preserve Privacy in Crowd-Powered Systems via Progressive Filtering,” en, in *Fifth AAAI Conference on Human Computation and Crowdsourcing*, Sep. 2017. [Online]. Available: <https://www.aaai.org/ocs/index.php/HCOMP/HCOMP17/paper/view/15938>.
- [17] W. S. Lasecki, M. Gordon, W. Leung, E. Lim, J. P. Bigham, and S. P. Dow, “Exploring Privacy and Accuracy Trade-Offs in Crowdsourced Behavioral Video Coding,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’15, New York, NY, USA: ACM, 2015, pp. 1945–1954, ISBN: 978-1-4503-3145-6. DOI: [10.1145/2702123.2702605](https://doi.org/10.1145/2702123.2702605). [Online]. Available: <http://doi.acm.org/10.1145/2702123.2702605>.
- [18] W. S. Lasecki, J. Teevan, and E. Kamar, “Information Extraction and Manipulation Threats in Crowd-powered Systems,” in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW ’14, New York, NY, USA: ACM, 2014, pp. 248–256, ISBN: 978-1-4503-2540-0. DOI: [10.1145/2531602.2531733](https://doi.org/10.1145/2531602.2531733). [Online]. Available: <http://doi.acm.org/10.1145/2531602.2531733>.
- [19] H. Kajino, Y. Baba, and H. Kashima, “Instance-Privacy Preserving Crowdsourcing,” en, in *Second AAAI Conference on Human Computation and Crowdsourcing*, Sep. 2014. [Online]. Available: <https://www.aaai.org/ocs/index.php/HCOMP/HCOMP14/paper/view/8946>.
- [20] L. von Ahn, R. Liu, and M. Blum, “Peekaboom: A Game for Locating Objects in Images,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’06, New York, NY, USA: ACM, 2006, pp. 55–64, ISBN: 978-1-59593-372-0. DOI: [10.1145/1124772.1124782](https://doi.org/10.1145/1124772.1124782). [Online]. Available: <http://doi.acm.org/10.1145/1124772.1124782>.
- [21] A. Das, H. Agrawal, L. Zitnick, D. Parikh, and D. Batra, “Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions?” *Computer Vision and Image Understanding*, Language in Vision, vol. 163, pp. 90–100, Oct. 2017, ISSN: 1077-3142. DOI: [10.1016/j.cviu.2017.10.001](https://doi.org/10.1016/j.cviu.2017.10.001). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314217301649>.
- [22] D. Gurari, S. D. Jain, M. Betke, and K. Grauman, “Pull the Plug? Predicting If Computers or Humans Should Segment Images,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR ’16, Las Vegas, NV, USA, Jun. 2016, pp. 382–391. DOI: [10.1109/CVPR.2016.48](https://doi.org/10.1109/CVPR.2016.48).
- [23] M. B. Lewis and A. J. Edmonds, “Face Detection: Mapping Human Performance,” en, *Perception*, vol. 32, no. 8, pp. 903–920, Aug. 2003, ISSN: 0301-0066. DOI: [10.1068/p5007](https://doi.org/10.1068/p5007). [Online]. Available: <https://doi.org/10.1068/p5007>.

- [24] D. S. Lindsay, P. C. Jack, and M. A. Christian, "Other-race face perception," eng, *The Journal of Applied Psychology*, vol. 76, no. 4, pp. 587–589, Aug. 1991, ISSN: 0021-9010. DOI: [10.1037/0021-9010.76.4.587](https://doi.org/10.1037/0021-9010.76.4.587).
- [25] G. Rhodes, V. Locke, L. Ewing, and E. Evangelista, "Race Coding and the Other-Race Effect in Face Recognition," en, *Perception*, vol. 38, no. 2, pp. 232–241, Feb. 2009, ISSN: 0301-0066. DOI: [10.1068/p6110](https://doi.org/10.1068/p6110). [Online]. Available: <https://doi.org/10.1068/p6110>.
- [26] K.-W. Wong, K.-M. Lam, and W.-C. Siu, "An efficient algorithm for human face detection and facial feature extraction under different conditions," en, *Pattern Recognition*, vol. 34, no. 10, pp. 1993–2004, Oct. 2001, ISSN: 0031-3203. DOI: [10.1016/S0031-3203\(00\)00134-5](https://doi.org/10.1016/S0031-3203(00)00134-5). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320300001345>.
- [27] C. Tsai, W. Cheng, J. Taur, and C. Tao, "Face Detection Using Eigenface and Neural Network," in *2006 IEEE International Conference on Systems, Man and Cybernetics*, ISSN: 1062-922X, vol. 5, Oct. 2006, pp. 4343–4347. DOI: [10.1109/ICSMC.2006.384817](https://doi.org/10.1109/ICSMC.2006.384817).
- [28] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," en, *Pattern Recognition*, vol. 39, no. 3, pp. 444–455, Mar. 2006, ISSN: 0031-3203. DOI: [10.1016/j.patcog.2005.09.009](https://doi.org/10.1016/j.patcog.2005.09.009). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320305003791>.
- [29] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: An application to face detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, Jun. 1997, pp. 130–136. DOI: [10.1109/CVPR.1997.609310](https://doi.org/10.1109/CVPR.1997.609310).
- [30] B. Heiselet, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, ISSN: 1063-6919, vol. 1, Dec. 2001, pp. I–I. DOI: [10.1109/CVPR.2001.990537](https://doi.org/10.1109/CVPR.2001.990537).
- [31] F. A. Pujol, M. Pujol, A. Jimeno-Morenilla, and M. J. Pujol, "Face Detection Based on Skin Color Segmentation Using Fuzzy Entropy," en, *Entropy*, vol. 19, no. 1, p. 26, Jan. 2017. DOI: [10.3390/e19010026](https://doi.org/10.3390/e19010026). [Online]. Available: <https://www.mdpi.com/1099-4300/19/1/26>.
- [32] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, Jan. 1998, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: [10.1109/34.655647](https://doi.org/10.1109/34.655647).

- [33] S. S. Farfade, M. J. Saberian, and L.-J. Li, “Multi-view Face Detection Using Deep Convolutional Neural Networks,” in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ser. ICMR '15, Shanghai, China: Association for Computing Machinery, Jun. 2015, pp. 643–650, ISBN: 978-1-4503-3274-3. DOI: [10.1145/2671188.2749408](https://doi.org/10.1145/2671188.2749408). [Online]. Available: <http://doi.org/10.1145/2671188.2749408>.
- [34] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, “Integrating Utility into Face De-identification,” in *Proceedings of the 5th International Conference on Privacy Enhancing Technologies*, ser. PET '05, Berlin, Heidelberg: Springer-Verlag, 2006, pp. 227–242, ISBN: 978-3-540-34745-3. DOI: . [Online]. Available: .
- [35] E. M. Newton, L. Sweeney, and B. Malin, “Preserving privacy by de-identifying face images,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, Feb. 2005, ISSN: 1041-4347. DOI: [10.1109/TKDE.2005.32](https://doi.org/10.1109/TKDE.2005.32).
- [36] L. Sweeney, “K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, Oct. 2002, ISSN: 0218-4885. DOI: [10.1142/S0218488502001648](https://doi.org/10.1142/S0218488502001648). [Online]. Available: <https://www.worldscientific.com/doi/abs/10.1142/S0218488502001648>.
- [37] A. Jourabloo, X. Yin, and X. Liu, “Attribute preserved face de-identification,” in *2015 International Conference on Biometrics (ICB)*, ISSN: 2376-4201, May 2015, pp. 278–285. DOI: [10.1109/ICB.2015.7139096](https://doi.org/10.1109/ICB.2015.7139096).
- [38] R. Collins, R. Gross, and J. Shi, “Silhouette-based human identification from body shape and gait,” in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, May 2002, pp. 366–371. DOI: [10.1109/AFGR.2002.1004181](https://doi.org/10.1109/AFGR.2002.1004181).
- [39] P. Agrawal and P. J. Narayanan, “Person De-Identification in Videos,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 299–310, Mar. 2011, ISSN: 1558-2205. DOI: [10.1109/TCSVT.2011.2105551](https://doi.org/10.1109/TCSVT.2011.2105551).
- [40] A. Ghanbari and M. Soryani, “Contour-Based Video Inpainting,” in *2011 7th Iranian Conference on Machine Vision and Image Processing*, ISSN: 2166-6784, Nov. 2011, pp. 1–5. DOI: [10.1109/IranianMVIP.2011.6121586](https://doi.org/10.1109/IranianMVIP.2011.6121586).
- [41] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain, “Face detection in color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: [10.1109/34.1000242](https://doi.org/10.1109/34.1000242).

- [42] H. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, Jan. 1998, ISSN: 1939-3539. DOI: [10.1109/34.655647](https://doi.org/10.1109/34.655647).
- [43] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” en, *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004, ISSN: 1573-1405. DOI: [10.1023/B:VISI.0000013087.49260.fb](https://doi.org/10.1023/B:VISI.0000013087.49260.fb). [Online]. Available: <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>.
- [44] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, “A convolutional neural network cascade for face detection,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, Jun. 2015, pp. 5325–5334. DOI: [10.1109/CVPR.2015.7299170](https://doi.org/10.1109/CVPR.2015.7299170).
- [45] X. Sun, P. Wu, and S. C. H. Hoi, “Face detection using deep learning: An improved faster RCNN approach,” en, *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018, ISSN: 0925-2312. DOI: [10.1016/j.neucom.2018.03.030](https://doi.org/10.1016/j.neucom.2018.03.030). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231218303229>.
- [46] S. Yang, P. Luo, C.-C. Loy, and X. Tang, “From Facial Parts Responses to Face Detection: A Deep Learning Approach,” 2015, pp. 3676–3684. [Online]. Available: .
- [47] H. Jiang and E. Learned-Miller, “Face Detection with the Faster R-CNN,” in *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, May 2017, pp. 650–657. DOI: [10.1109/FG.2017.82](https://doi.org/10.1109/FG.2017.82).
- [48] B. Zhang, J. Li, Y. Wang, Y. Tai, C. Wang, J. Li, F. Huang, Y. Xia, W. Pei, and R. Ji, “ASFD: Automatic and Scalable Face Detector,” *arXiv:2003.11228 [cs]*, Mar. 2020, arXiv: 2003.11228. [Online]. Available: <http://arxiv.org/abs/2003.11228>.
- [49] G. Guo, H. Wang, Y. Yan, J. Zheng, and B. Li, “A fast face detection method via convolutional neural network,” en, *Neurocomputing*, vol. 395, pp. 128–137, Jun. 2020, ISSN: 0925-2312. DOI: [10.1016/j.neucom.2018.02.110](https://doi.org/10.1016/j.neucom.2018.02.110). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231219309075>.
- [50] S. Yang, P. Luo, C. C. Loy, and X. Tang, “WIDER FACE: A Face Detection Benchmark,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, Jun. 2016, pp. 5525–5533. DOI: [10.1109/CVPR.2016.596](https://doi.org/10.1109/CVPR.2016.596).
- [51] B. Yang, J. Yan, Z. Lei, and S. Z. Li, “Fine-grained evaluation on face detection in the wild,” in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, May 2015, pp. 1–7. DOI: [10.1109/FG.2015.7163158](https://doi.org/10.1109/FG.2015.7163158).

- [52] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep Learning Face Attributes in the Wild,” 2015, pp. 3730–3738. [Online]. Available:
- [53] L. R. Varshney, “Privacy and Reliability in Crowdsourcing Service Delivery,” in *2012 Annual SRII Global Conference*, ser. SRII ’12, San Jose, CA, USA, Jul. 2012, pp. 55–60. DOI: [10.1109/SRII.2012.17](https://doi.org/10.1109/SRII.2012.17).
- [54] R. Hardie, K. Barnard, and E. Armstrong, “Joint MAP registration and high-resolution image estimation using a sequence of undersampled images,” *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997, Conference Name: IEEE Transactions on Image Processing, ISSN: 1941-0042. DOI: [10.1109/83.650116](https://doi.org/10.1109/83.650116).
- [55] R. A. Hummel, B. Kimia, and S. W. Zucker, “Deblurring Gaussian blur,” *Computer Vision, Graphics, and Image Processing*, vol. 38, no. 1, pp. 66–80, Apr. 1987, ISSN: 0734-189X. DOI: [10.1016/S0734-189X\(87\)80153-6](https://doi.org/10.1016/S0734-189X(87)80153-6). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0734189X87801536>.
- [56] A. Clark, *Pillow: Python Imaging Library*, 2020. [Online]. Available: <https://python-pillow.org>.
- [57] D. L. Hansen, P. J. Schone, D. Corey, M. Reid, and J. Gehring, “Quality control mechanisms for crowdsourcing: Peer review, arbitration, & expertise at familysearch indexing,” in *Proceedings of the 2013 conference on Computer supported cooperative work*, ser. CSCW ’13, San Antonio, Texas, USA: Association for Computing Machinery, Feb. 2013, pp. 649–660, ISBN: 9781450313315. DOI: [10.1145/2441776.2441848](https://doi.org/10.1145/2441776.2441848). [Online]. Available: <https://doi.org/10.1145/2441776.2441848>.
- [58] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, “Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions,” *ACM Computing Surveys*, vol. 51, no. 1, 7:1–7:40, Jan. 2018, ISSN: 0360-0300. DOI: [10.1145/3148148](https://doi.org/10.1145/3148148). [Online]. Available: <https://doi.org/10.1145/3148148>.
- [59] A. Marcus, D. Karger, S. Madden, R. Miller, and S. Oh, “Counting with the crowd,” *Proceedings of the VLDB Endowment*, vol. 6, no. 2, pp. 109–120, Dec. 2012, ISSN: 2150-8097. DOI: [10.14778/2535568.2448944](https://doi.org/10.14778/2535568.2448944). [Online]. Available: <https://doi.org/10.14778/2535568.2448944>.
- [60] S.-W. Huang and W.-T. Fu, “Enhancing reliability using peer consistency evaluation in human computation,” in *Proceedings of the 2013 conference on Computer supported cooperative work*, ser. CSCW ’13, San Antonio, Texas, USA: Association for Computing Machinery, Feb. 2013, pp. 639–648, ISBN: 9781450313315. DOI: [10.1145/2441776.2441847](https://doi.org/10.1145/2441776.2441847). [Online]. Available: <https://doi.org/10.1145/2441776.2441847>.

- [61] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2005, ISBN: 9780385721707.
- [62] R. Rothe, R. Timofte, and L. Van Gool, “DEX: Deep EXpectation of Apparent Age from a Single Image,” in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec. 2015, pp. 252–257. DOI: [10.1109/ICCVW.2015.41](https://doi.org/10.1109/ICCVW.2015.41).
- [63] R. Rothe, R. Timofte, and L. Van Gool, “Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks,” en, *International Journal of Computer Vision*, vol. 126, no. 2, pp. 144–157, Apr. 2018, ISSN: 1573-1405. DOI: [10.1007/s11263-016-0940-3](https://doi.org/10.1007/s11263-016-0940-3). [Online]. Available: <https://doi.org/10.1007/s11263-016-0940-3>.
- [64] N. B. Shah and D. Zhou, “Double or Nothing: Multiplicative Incentive Mechanisms for Crowdsourcing,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, ser. NIPS ’15, Cambridge, MA, USA: MIT Press, 2015, pp. 1–9. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2969239.2969240>.
- [65] S. Zafeiriou, C. Zhang, and Z. Zhang, “A survey on face detection in the wild: Past, present and future,” *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, Sep. 2015, ISSN: 1077-3142. DOI: [10.1016/j.cviu.2015.03.015](https://doi.org/10.1016/j.cviu.2015.03.015). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314215000727>.
- [66] D. Triantafyllidou, P. Nousi, and A. Tefas, “Fast Deep Convolutional Face Detection in the Wild Exploiting Hard Sample Mining,” *Big Data Research*, Selected papers from the 2nd INNS Conference on Big Data: Big Data & Neural Networks, vol. 11, pp. 65–76, Mar. 2018, ISSN: 2214-5796. DOI: [10.1016/j.bdr.2017.06.002](https://doi.org/10.1016/j.bdr.2017.06.002). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2214579617300096>.
- [67] J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, “Platemate: Crowdsourcing Nutritional Analysis from Food Photographs,” in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’11, New York, NY, USA: ACM, 2011, pp. 1–12, ISBN: 978-1-4503-0716-1. DOI: [10.1145/2047196.2047198](https://doi.org/10.1145/2047196.2047198). [Online]. Available: <http://doi.acm.org/10.1145/2047196.2047198>.
- [68] mturk, *Amazon Mechanical Turk - Welcome*, 2017. [Online]. Available: <https://www.mturk.com/mturk/welcome>.
- [69] A. N. Kolmogorov and A. T. Bharucha-Reid, *Foundations of the Theory of Probability: Second English Edition*, en. Courier Dover Publications, Apr. 2018, Google-Books-ID: ZZ5ODwAAQBAJ, ISBN: 9780486821597.
- [70] Microsoft, *Facial Recognition — Microsoft Azure*, en. [Online]. Available: <https://azure.microsoft.com/en-us/services/cognitive-services/face/>.

- [71] Amazon, *Data protection in Amazon Rekognition - Amazon Rekognition*, 2021. [Online]. Available: <https://docs.aws.amazon.com/rekognition/latest/dg/data-protection.html>.
- [72] R. Yadav and Priyanka, “An Overview of Recent Developments in Convolutional Neural Network (CNN) Based Face Detector,” en, in *Computational Methods and Data Engineering*, V. Singh, V. K. Asari, S. Kumar, and R. B. Patel, Eds., ser. Advances in Intelligent Systems and Computing, Singapore: Springer, 2021, pp. 243–258, ISBN: 9789811579073. DOI: .
- [73] H. Qin, J. Yan, X. Li, and X. Hu, “Joint Training of Cascaded CNN for Face Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, Jun. 2016, pp. 3456–3465. DOI: [10.1109/CVPR.2016.376](https://doi.org/10.1109/CVPR.2016.376).
- [74] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, ISSN: 1939-3539. DOI: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [75] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, Jun. 2014, pp. 580–587. DOI: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [76] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, Jul. 2017, pp. 6517–6525. DOI: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).
- [77] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single Shot MultiBox Detector,” en, in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2016, pp. 21–37, ISBN: 9783319464480. DOI: .
- [78] Y. Liu, F. Wang, B. Sun, and H. Li, “MogFace: Rethinking Scale Augmentation on the Face Detector,” *arXiv:2103.11139 [cs]*, Mar. 2021, arXiv: 2103.11139. [Online]. Available: <http://arxiv.org/abs/2103.11139>.
- [79] S. Agarwal, A. Awan, and D. Roth, “Learning to detect objects in images via a sparse, part-based representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, Nov. 2004, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: [10.1109/TPAMI.2004.108](https://doi.org/10.1109/TPAMI.2004.108).

- [80] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, Jun. 2012, pp. 2879–2886. DOI: [10.1109/CVPR.2012.6248014](https://doi.org/10.1109/CVPR.2012.6248014).
- [81] N. I. Glumov, E. I. Kolomiyetz, and V. V. Sergeyev, "Detection of objects on the image using a sliding window mode," *Optics & Laser Technology*, vol. 27, no. 4, pp. 241–249, 1995, ISSN: 0030-3992. DOI: [https://doi.org/10.1016/0030-3992\(95\)93752-D](https://doi.org/10.1016/0030-3992(95)93752-D). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/003039929593752D>.
- [82] H. G. R. Gouk and A. M. Blake, "Fast Sliding Window Classification with Convolutional Neural Networks," in *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*, ser. IVCNZ '14, New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 114–118, ISBN: 978-1-4503-3184-5. DOI: [10.1145/2683405.2683429](https://doi.org/10.1145/2683405.2683429). [Online]. Available: <http://doi.org/10.1145/2683405.2683429>.
- [83] J. Seo and H. Ko, "Face detection using support vector domain description in color images," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, ISSN: 1520-6149, vol. 5, May 2004, pp. V–729. DOI: [10.1109/ICASSP.2004.1327214](https://doi.org/10.1109/ICASSP.2004.1327214).
- [84] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, arXiv: 1409.1556. [Online]. Available: <http://arxiv.org/abs/1409.1556>.
- [85] N. Qian, "On the momentum term in gradient descent learning algorithms," eng, *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 12, no. 1, pp. 145–151, Jan. 1999, ISSN: 1879-2782. DOI: [10.1016/s0893-6080\(98\)00116-6](https://doi.org/10.1016/s0893-6080(98)00116-6).
- [86] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," en, *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, ISSN: 2196-1115. DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0). [Online]. Available: <https://doi.org/10.1186/s40537-019-0197-0>.
- [87] F. J. Moreno-Barea, F. Strazzera, J. M. Jerez, D. Urda, and L. Franco, "Forward Noise Adjustment Scheme for Data Augmentation," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Nov. 2018, pp. 728–734. DOI: [10.1109/SSCI.2018.8628917](https://doi.org/10.1109/SSCI.2018.8628917).
- [88] L. Taylor and G. Nitschke, "Improving Deep Learning using Generic Data Augmentation," *arXiv:1708.06020 [cs, stat]*, Aug. 2017, arXiv: 1708.06020. [Online]. Available: <http://arxiv.org/abs/1708.06020>.
- [89] face++, *Face Detection - Face++ Cognitive Services*, 2021. [Online]. Available: <https://www.faceplusplus.com/face-detection/>.

- [90] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, issn: 1558-2361. DOI: [10.1109/LSP.2016.2603342](https://doi.org/10.1109/LSP.2016.2603342).
- [91] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” *arXiv: 2104.00298 [cs]*, Jun. 2021, arXiv: 2104.00298. [Online]. Available: <http://arxiv.org/abs/2104.00298>.

A. EXPERIMENT IMAGES

This appendix contains all the images that were used in the experiments. The images are organized based on the the experiments that used them. The images include the test images and the attention check images.

A.1 IntoFocus Experiment 1 images



Figure A.1. Grid 1 of the images used in the experiments in chapter 3



Figure A.2. Grid 2 of the images used in the experiments in chapter 3



Figure A.3. Grid 3 of the images used in the experiments in chapter 3

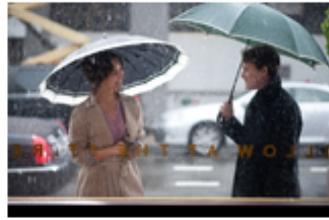


Figure A.4. Grid 4 of the images used in the experiments in chapter 3



Figure A.5. Grid 5 of the images used in the experiments in chapter 3



Figure A.6. Grid 6 of the images used in the experiments in chapter 3



Figure A.7. Grid 7 of the images used in the experiments in chapter 3

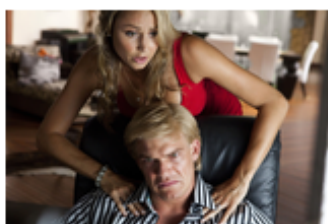


Figure A.8. Grid 8 of the images used in the experiments in chapter 3



Figure A.9. Grid 9 of the images used in the experiments in chapter 3



Figure A.10. Grid 10 of the images used in the experiments in chapter 3



Figure A.11. Grid 11 of the images used in the experiments in chapter 3



Figure A.12. Grid 12 of the images used in the experiments in chapter 3



Figure A.13. Grid 13 of the images used in the experiments in chapter 3



Figure A.14. Grid 14 of the images used in the experiments in chapter 3



Figure A.15. Grid 15 of the images used in the experiments in chapter 3

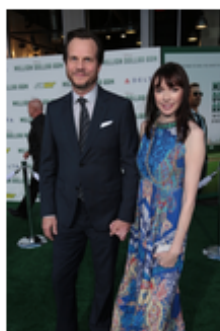
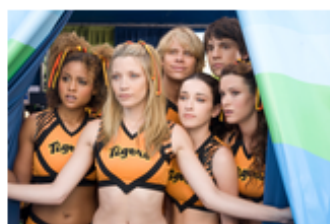


Figure A.16. Grid 16 of the images used in the experiments in chapter 3



Figure A.17. Grid 17 of the images used in the experiments in chapter 3



Figure A.18. Grid 18 of the images used in the experiments in chapter 3

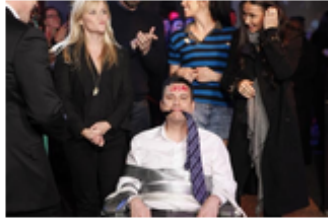


Figure A.19. Grid 19 of the images used in the experiments in chapter 3



Figure A.20. Grid 20 of the images used in the experiments in chapter 3



Figure A.21. Grid 21 of the images used in the experiments in chapter 3

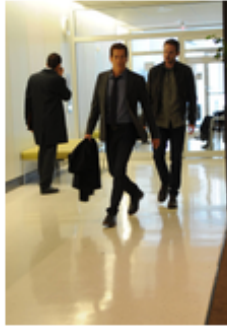


Figure A.22. Grid 22 of the images used in the experiments in chapter 3

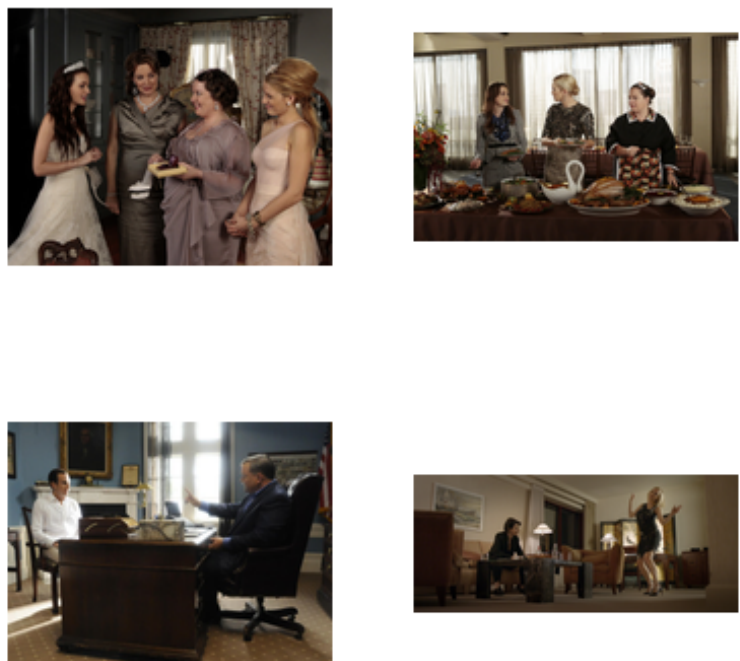


Figure A.23. Grid 23 of the images used in the experiments in chapter 3

A.2 IntoFocus Experiment 2 images



Figure A.24. Grid 1 of the images used in the experiments in chapter 3



Figure A.25. Grid 2 of the images used in the experiments in chapter 3

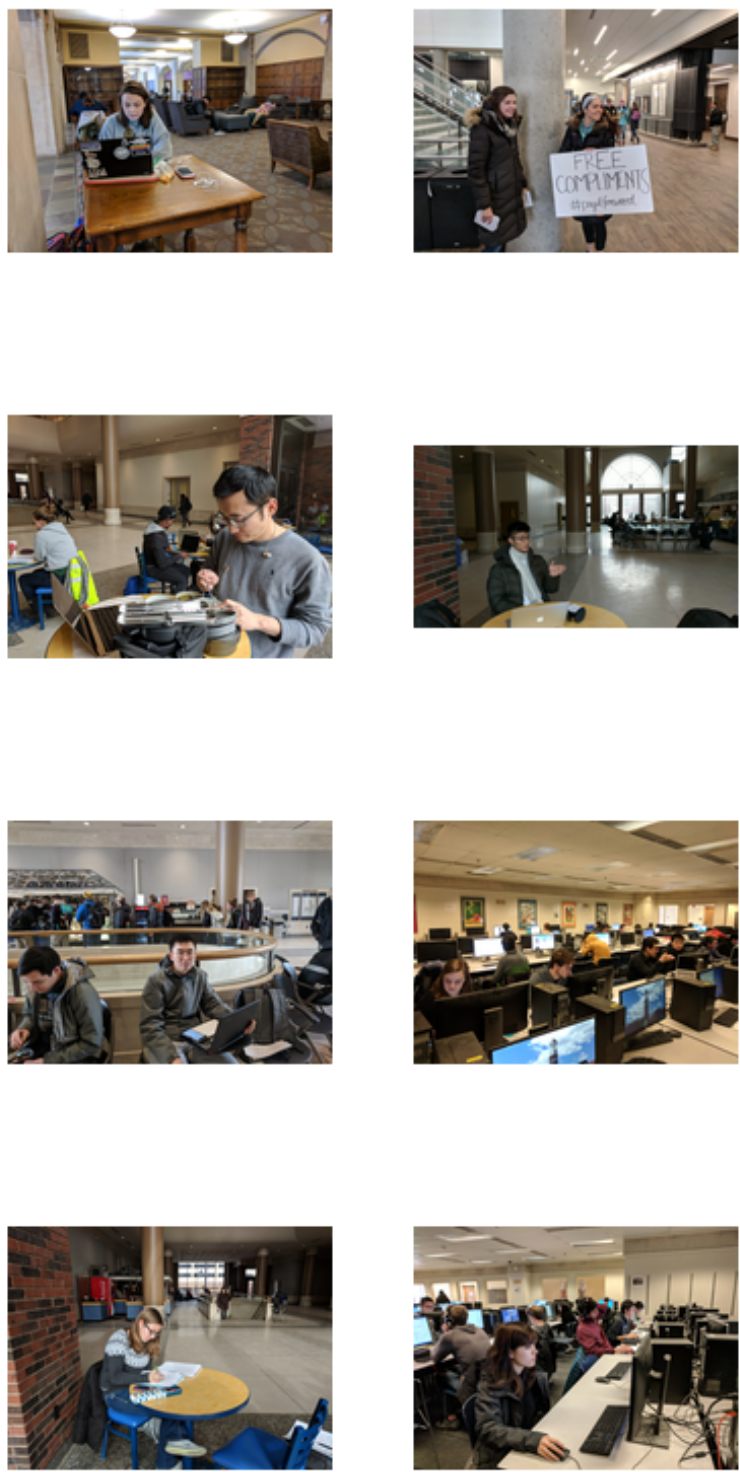


Figure A.26. Grid 3 of the images used in the experiments in chapter 3



Figure A.27. Grid 4 of the images used in the experiments in chapter 3



Figure A.28. Grid 5 of the images used in the experiments in chapter 3

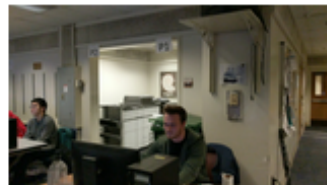


Figure A.29. Grid 6 of the images used in the experiments in chapter 3

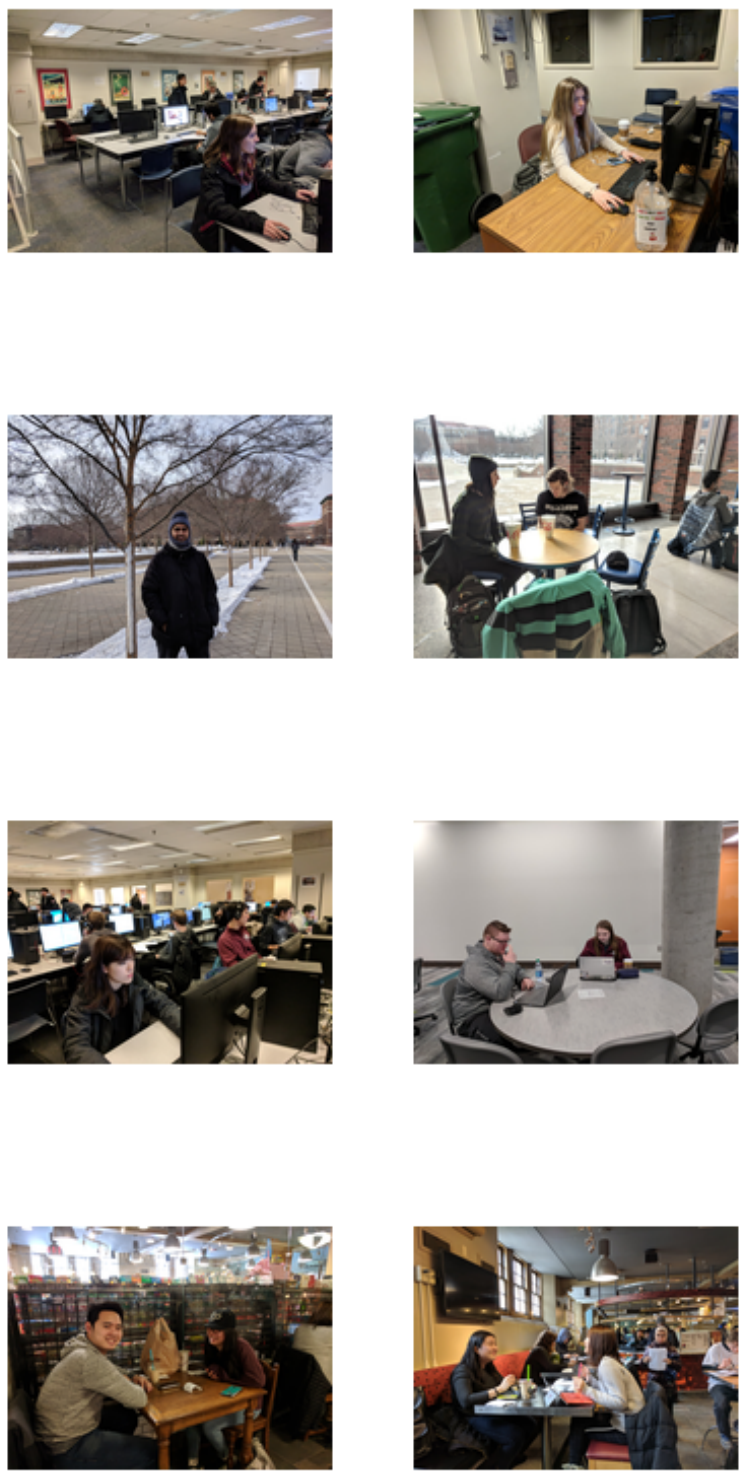


Figure A.30. Grid 7 of the images used in the experiments in chapter 3

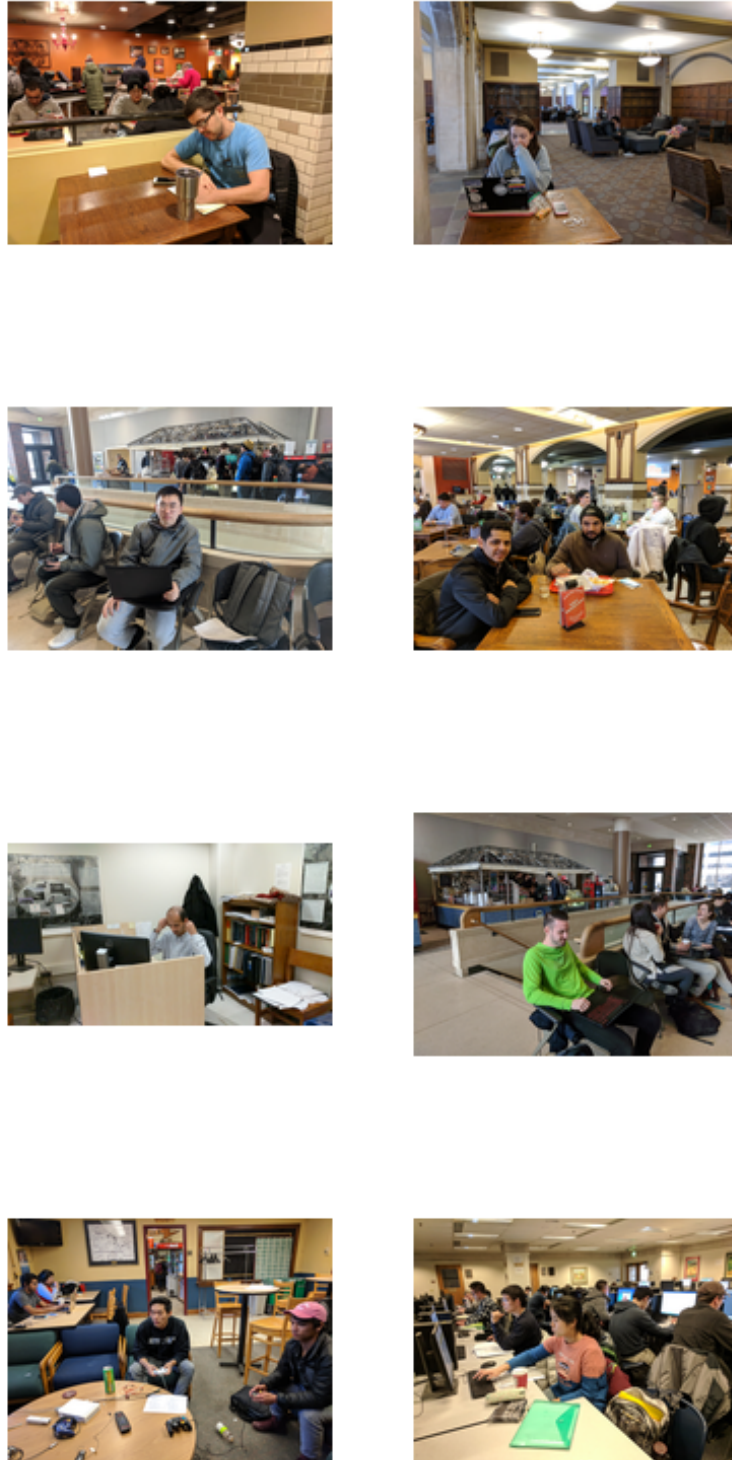


Figure A.31. Grid 8 of the images used in the experiments in chapter 3



Figure A.32. Grid 9 of the images used in the experiments in chapter 3

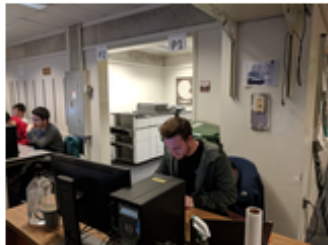


Figure A.33. Grid 10 of the images used in the experiments in chapter 3

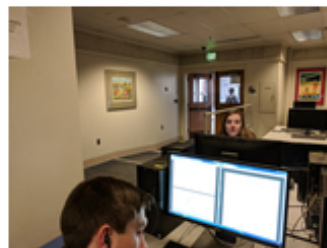
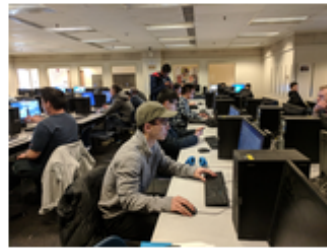
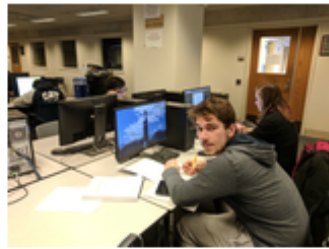
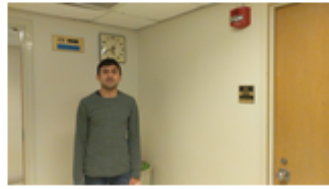


Figure A.34. Grid 11 of the images used in the experiments in chapter 3

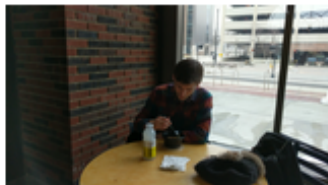


Figure A.35. Grid 12 of the images used in the experiments in chapter 3

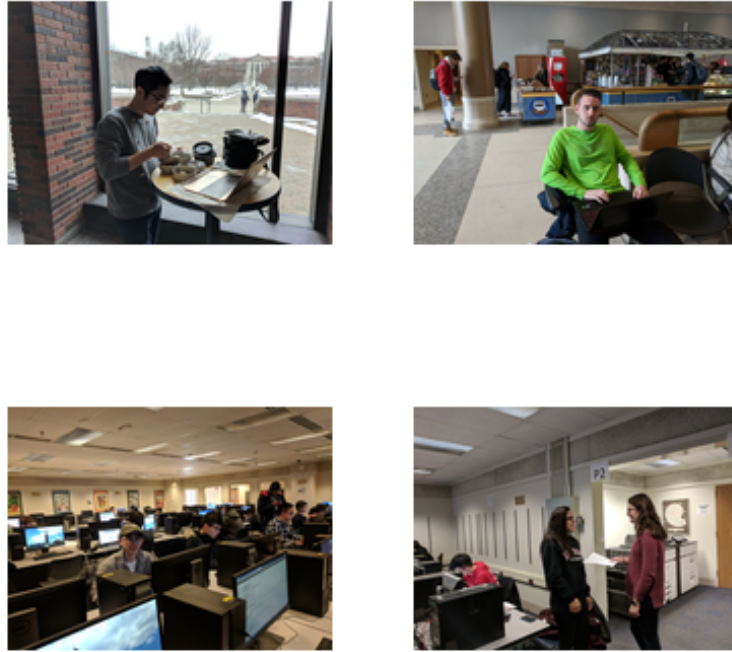


Figure A.36. Grid 13 of the images used in the experiments in chapter 3

A.3 The Human Perception Experiment Images



Figure A.37. Grid 1 of the images used in the experiments in chapter 4

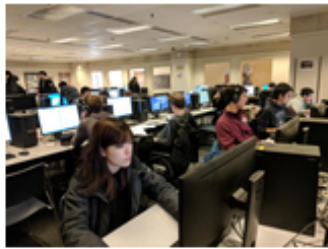


Figure A.38. Grid 2 of the images used in the experiments in chapter 4



Figure A.39. Grid 3 of the images used in the experiments in chapter 4



Figure A.40. Grid 4 of the images used in the experiments in chapter 4

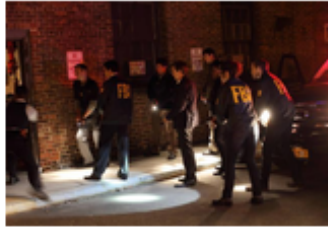
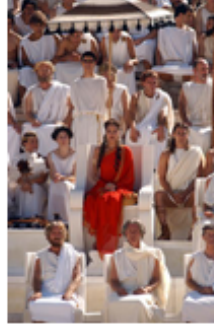
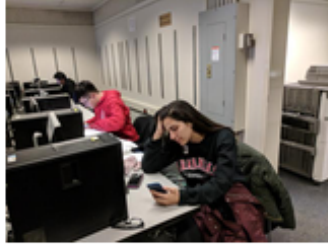


Figure A.41. Grid 5 of the images used in the experiments in chapter 4

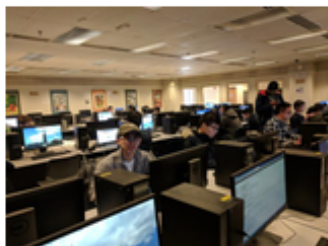


Figure A.42. Grid 6 of the images used in the experiments in chapter 4



Figure A.43. Grid 7 of the images used in the experiments in chapter 4



Figure A.44. Grid 8 of the images used in the experiments in chapter 4

A.4 The Improved IntoFocus Experiment Images

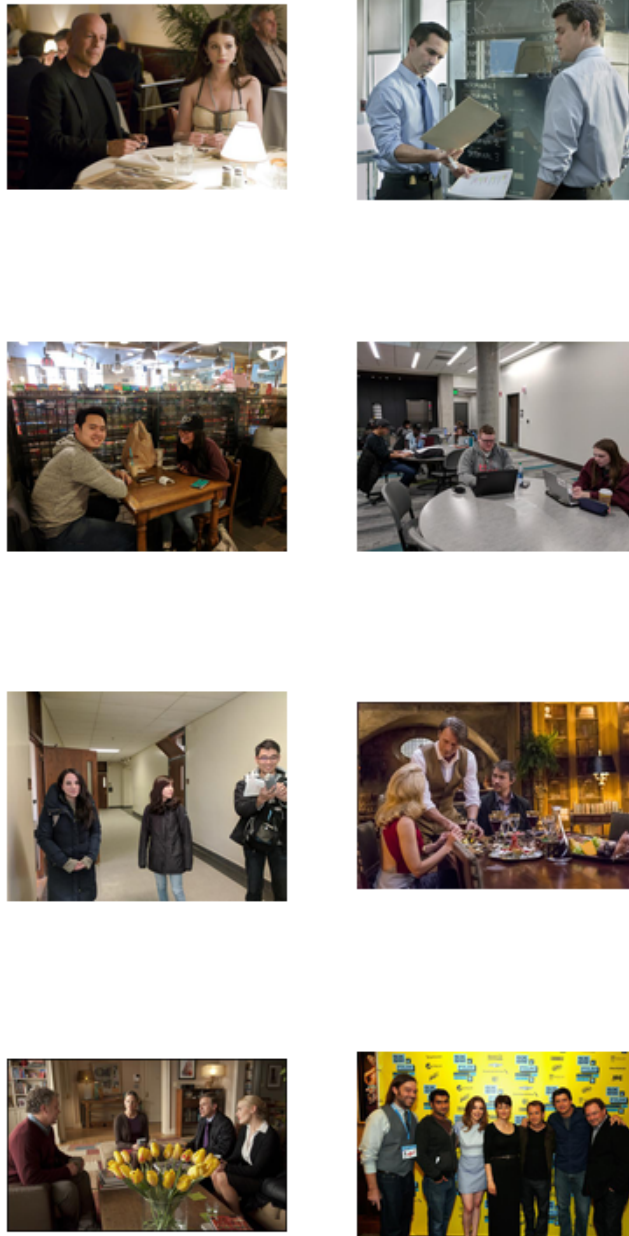


Figure A.45. Grid 1 of the images used in the experiments in chapter 4



Figure A.46. Grid 2 of the images used in the experiments in chapter 4



Figure A.47. Grid 3 of the images used in the experiments in chapter 4



Figure A.48. Grid 4 of the images used in the experiments in chapter 4

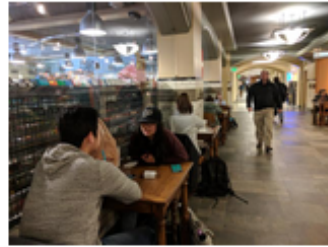


Figure A.49. Grid 5 of the images used in the experiments in chapter 4

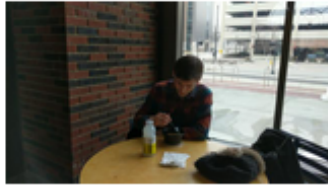


Figure A.50. Grid 6 of the images used in the experiments in chapter 4

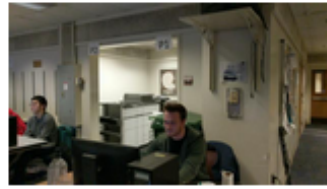


Figure A.51. Grid 7 of the images used in the experiments in chapter 4

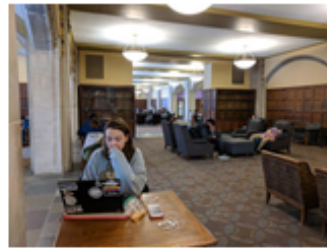
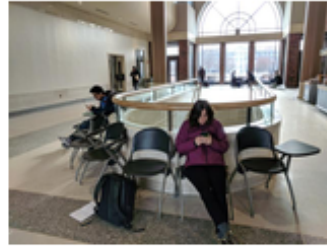


Figure A.52. Grid 8 of the images used in the experiments in chapter 4



Figure A.53. Grid 9 of the images used in the experiments in chapter 4



Figure A.54. Grid 10 of the images used in the experiments in chapter 4

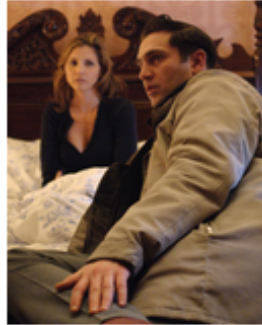


Figure A.55. Grid 11 of the images used in the experiments in chapter 4



Figure A.56. Grid 12 of the images used in the experiments in chapter 4



Figure A.57. Grid 13 of the images used in the experiments in chapter 4



Figure A.58. Grid 14 of the images used in the experiments in chapter 4

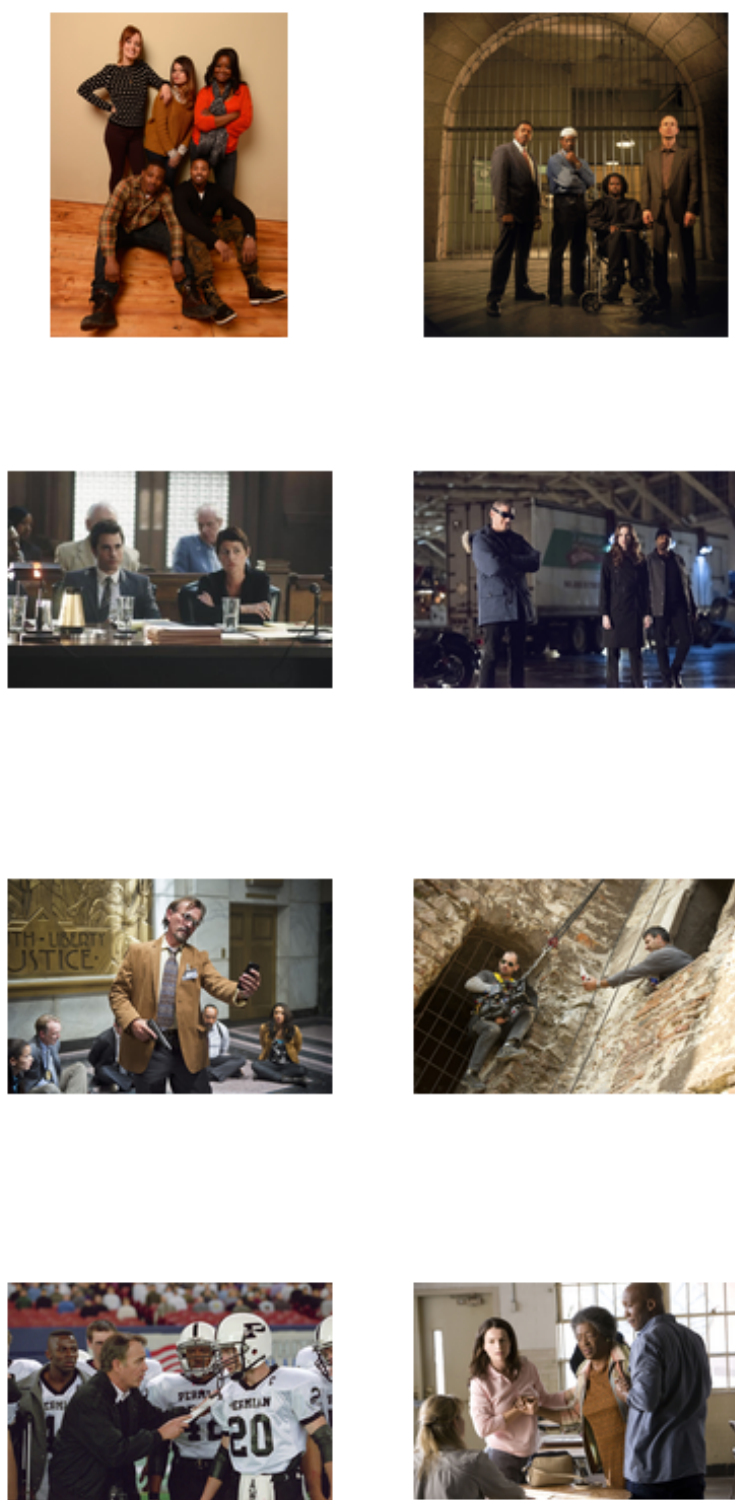


Figure A.59. Grid 15 of the images used in the experiments in chapter 4



Figure A.60. Grid 16 of the images used in the experiments in chapter 4

A.5 The Pterodactyl Experiment Images



Figure A.61. Grid 1 of the images used in the experiments in chapter 5



Figure A.62. Grid 2 of the images used in the experiments in chapter 5



Figure A.63. Grid 3 of the images used in the experiments in chapter 5



Figure A.64. Grid 4 of the images used in the experiments in chapter 5



Figure A.65. Grid 5 of the images used in the experiments in chapter 5



Figure A.66. Grid 6 of the images used in the experiments in chapter 5



Figure A.67. Grid 7 of the images used in the experiments in chapter 5



Figure A.68. Grid 8 of the images used in the experiments in chapter 5

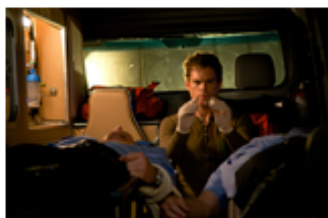


Figure A.69. Grid 9 of the images used in the experiments in chapter 5



Figure A.70. Grid 10 of the images used in the experiments in chapter 5



Figure A.71. Grid 11 of the images used in the experiments in chapter 5



Figure A.72. Grid 12 of the images used in the experiments in chapter 5



Figure A.73. Grid 13 of the images used in the experiments in chapter 5



Figure A.74. Grid 14 of the images used in the experiments in chapter 5



Figure A.75. Grid 15 of the images used in the experiments in chapter 5



Figure A.76. Grid 16 of the images used in the experiments in chapter 5



Figure A.77. Grid 17 of the images used in the experiments in chapter 5



Figure A.78. Grid 18 of the images used in the experiments in chapter 5



Figure A.79. Grid 19 of the images used in the experiments in chapter 5



Figure A.80. Grid 20 of the images used in the experiments in chapter 5



Figure A.81. Grid 21 of the images used in the experiments in chapter 5



Figure A.82. Grid 22 of the images used in the experiments in chapter 5



Figure A.83. Grid 23 of the images used in the experiments in chapter 5



Figure A.84. Grid 24 of the images used in the experiments in chapter 5



Figure A.85. Grid 25 of the images used in the experiments in chapter 5

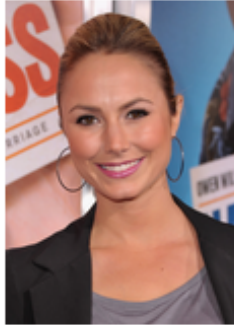


Figure A.86. Grid 26 of the images used in the experiments in chapter 5



Figure A.87. Grid 27 of the images used in the experiments in chapter 5

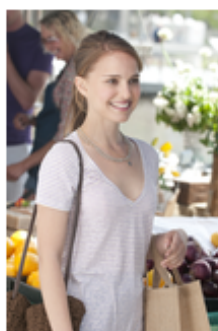


Figure A.88. Grid 28 of the images used in the experiments in chapter 5



Figure A.89. Grid 29 of the images used in the experiments in chapter 5



Figure A.90. Grid 30 of the images used in the experiments in chapter 5



Figure A.91. Grid 31 of the images used in the experiments in chapter 5



Figure A.92. Grid 32 of the images used in the experiments in chapter 5



Figure A.93. Grid 33 of the images used in the experiments in chapter 5



Figure A.94. Grid 34 of the images used in the experiments in chapter 5



Figure A.95. Grid 35 of the images used in the experiments in chapter 5

B. FACES

This appendix contains all the face images that were used to test for identification on all the experiments combined.

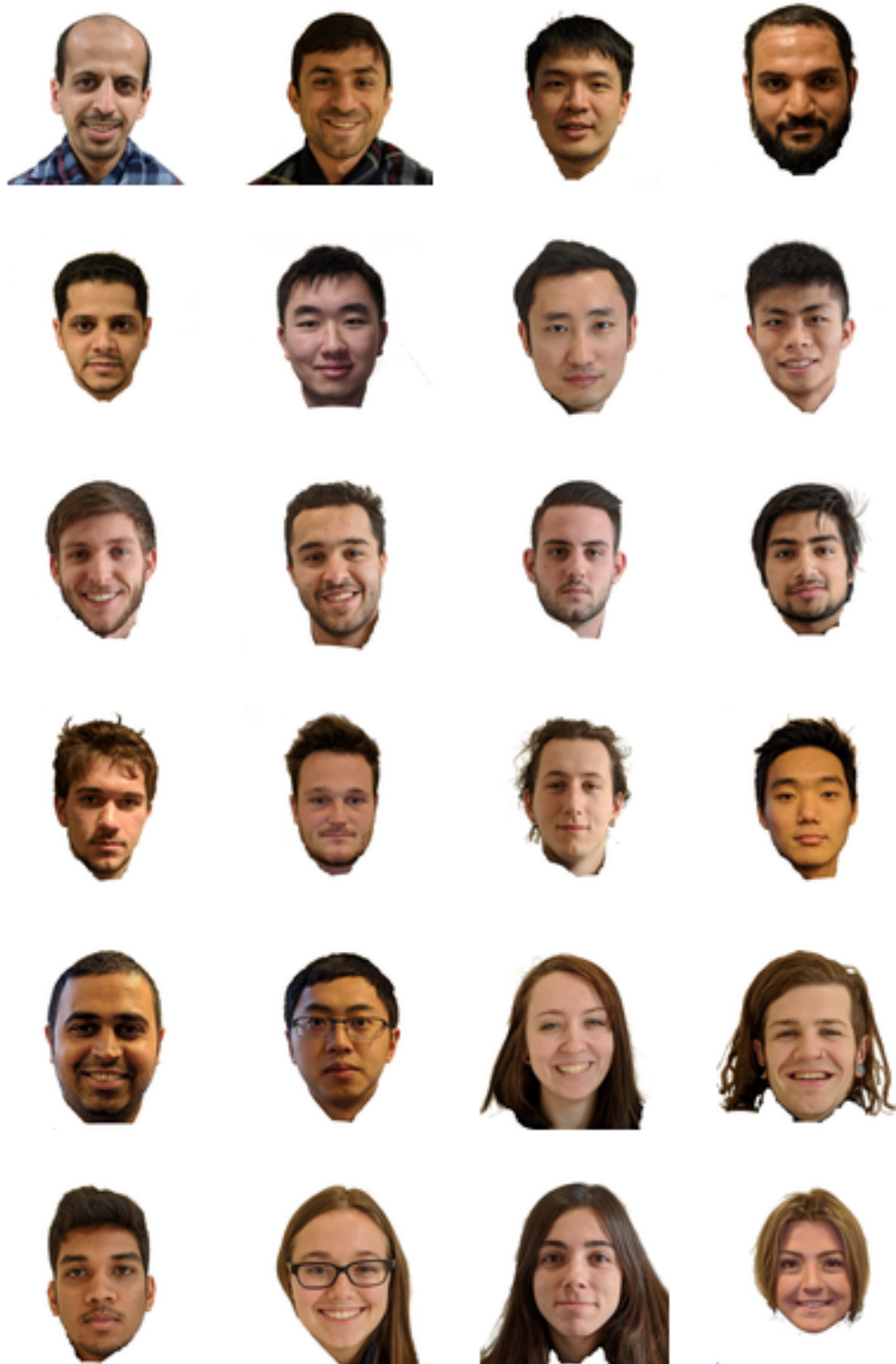


Figure B.1. Grid 1 of the faces used in the experiments.

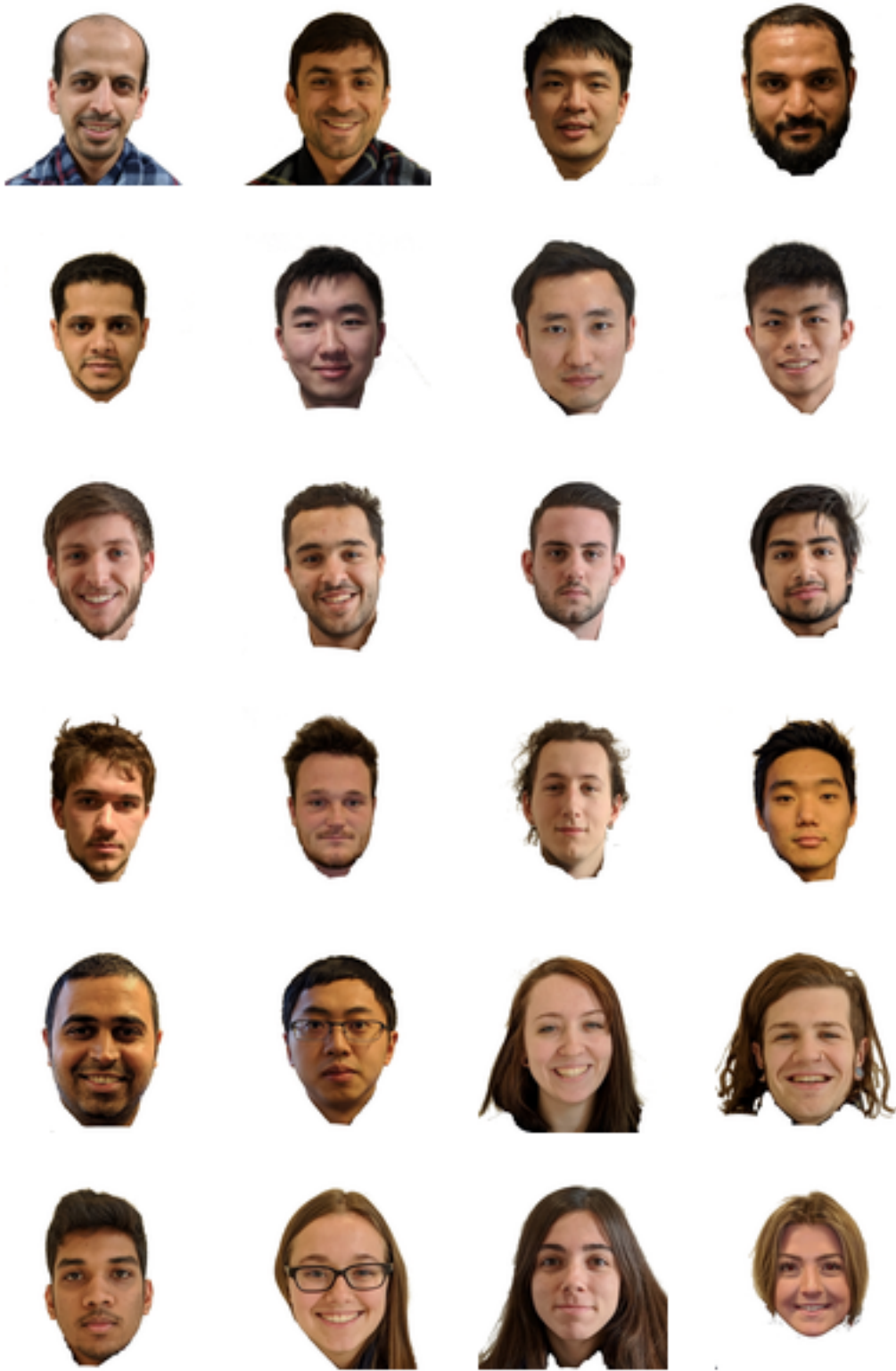


Figure B.2. Grid 1 of the faces used in the experiments.

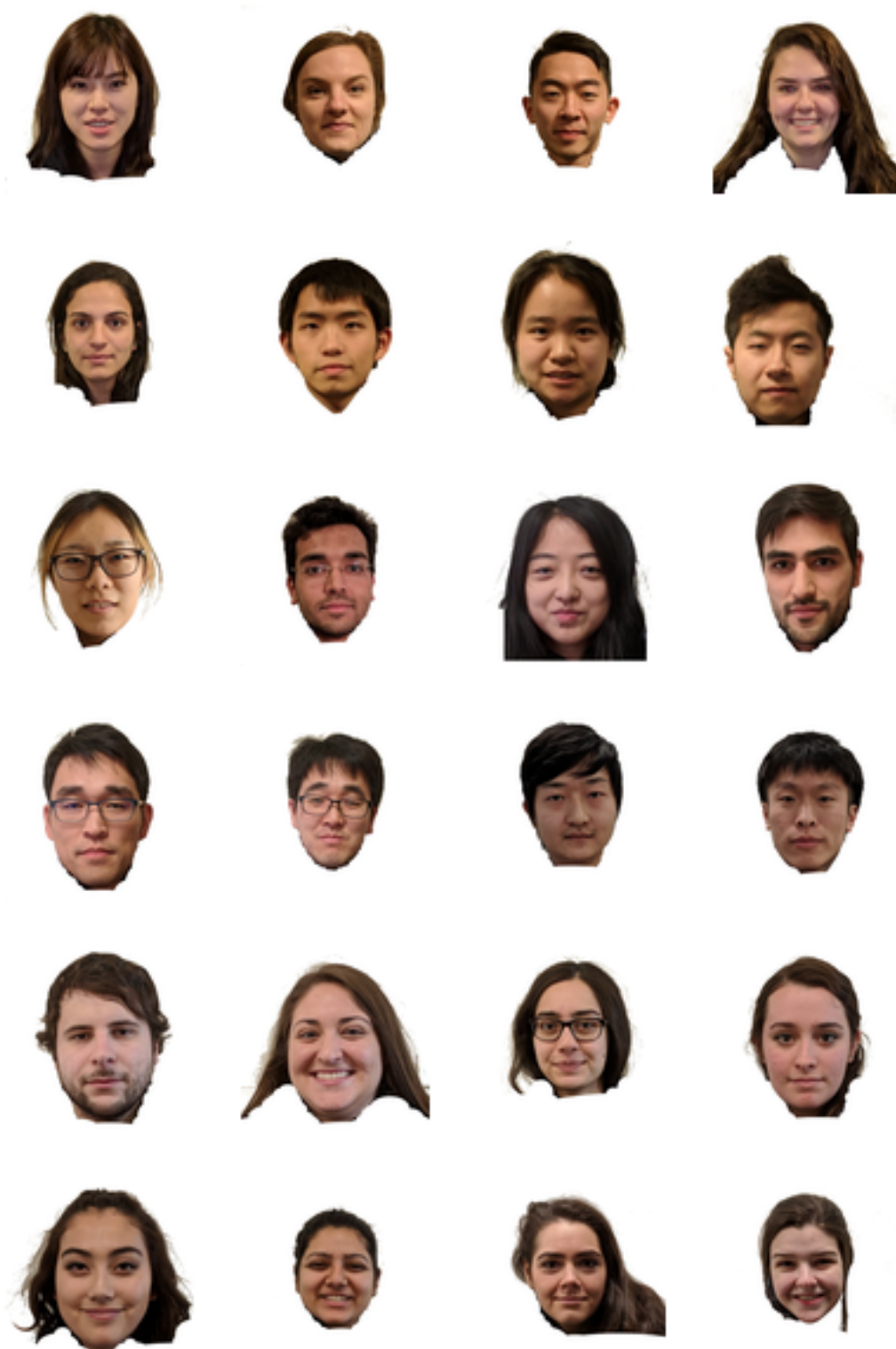


Figure B.3. Grid 2 of the faces used in the experiments.

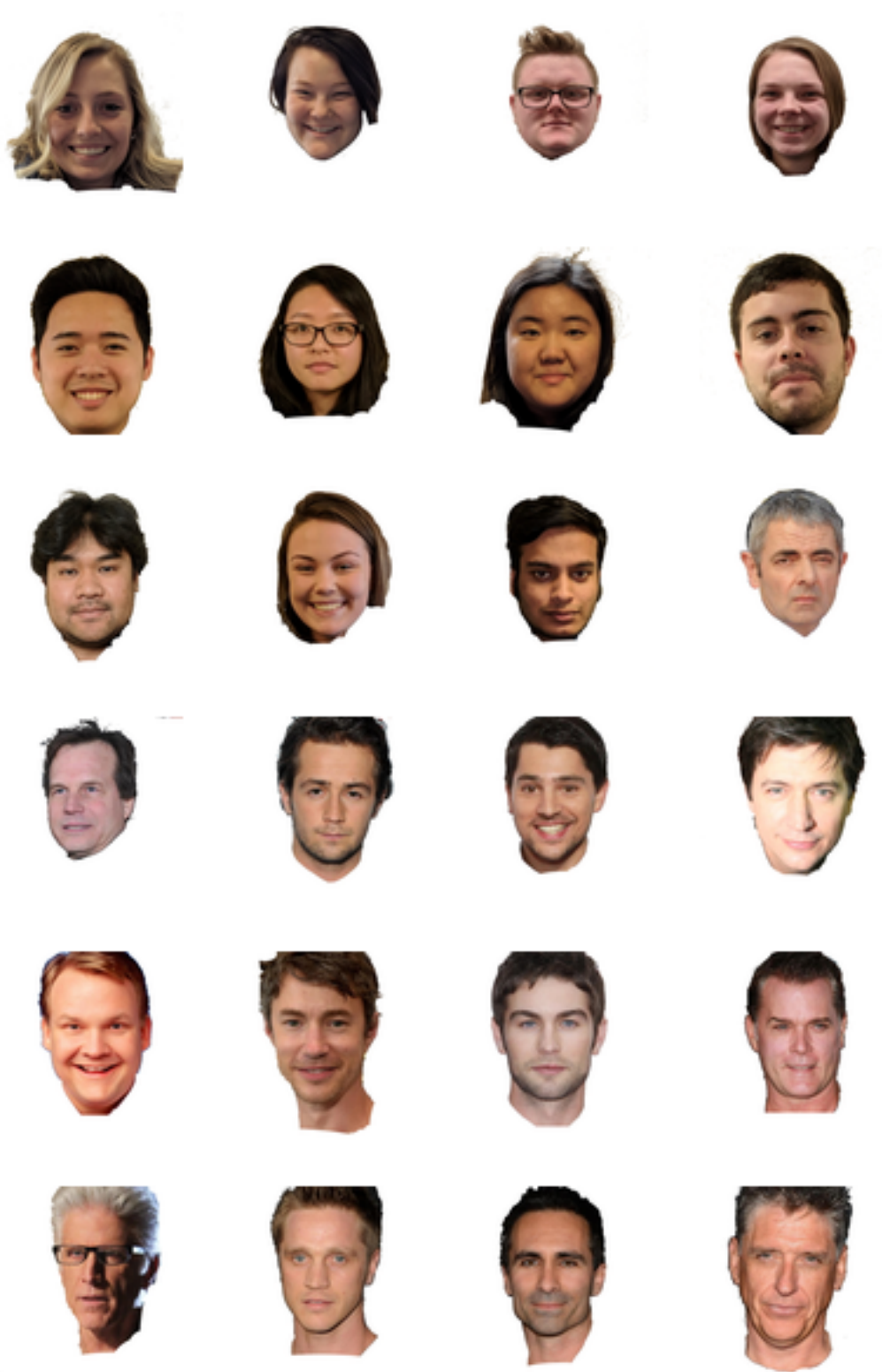


Figure B.4. Grid 3 of the faces used in the experiments.

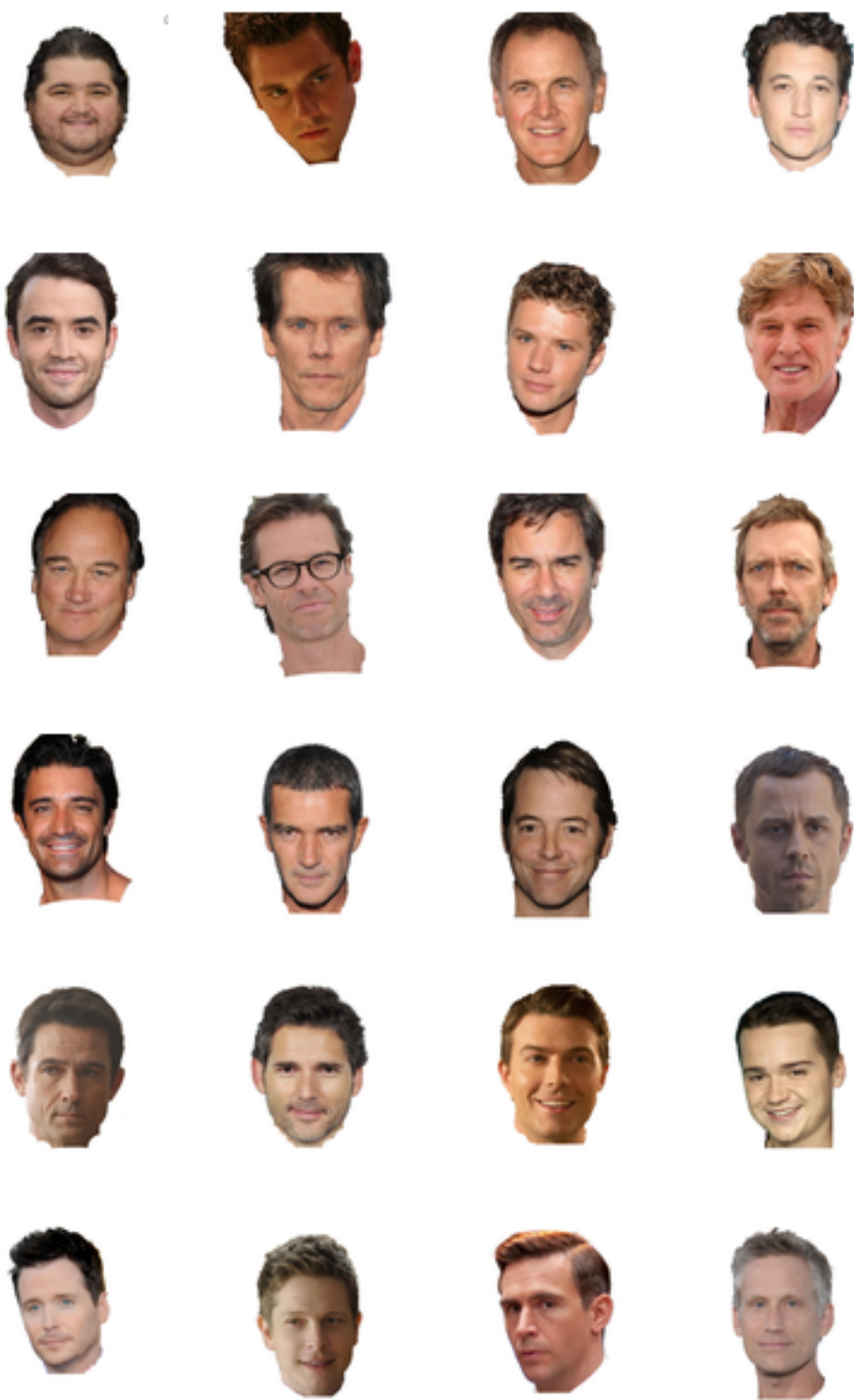


Figure B.5. Grid 4 of the faces used in the experiments.



Figure B.6. Grid 5 of the faces used in the experiments.

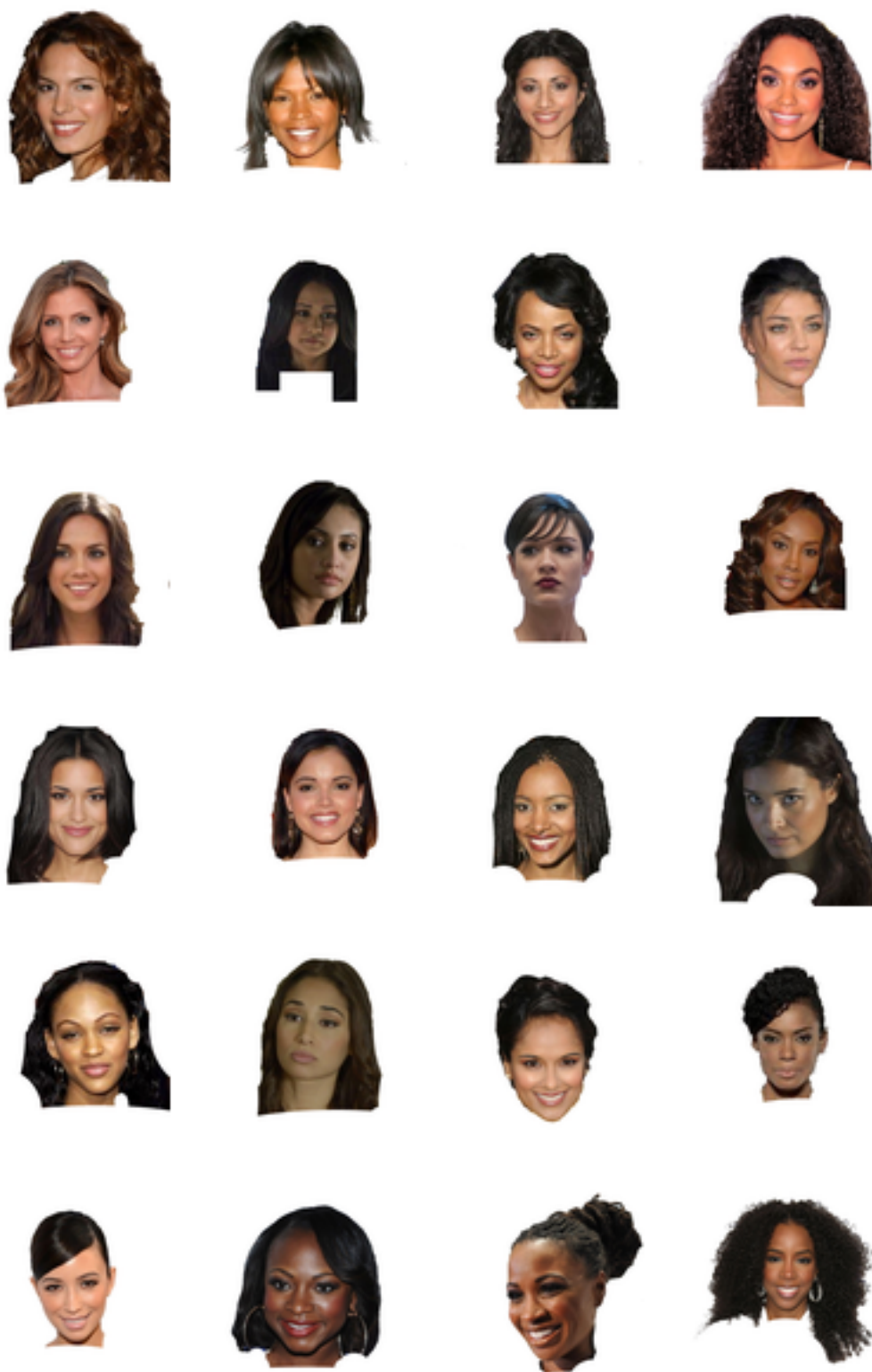


Figure B.7. Grid 6 of the faces used in the experiments.

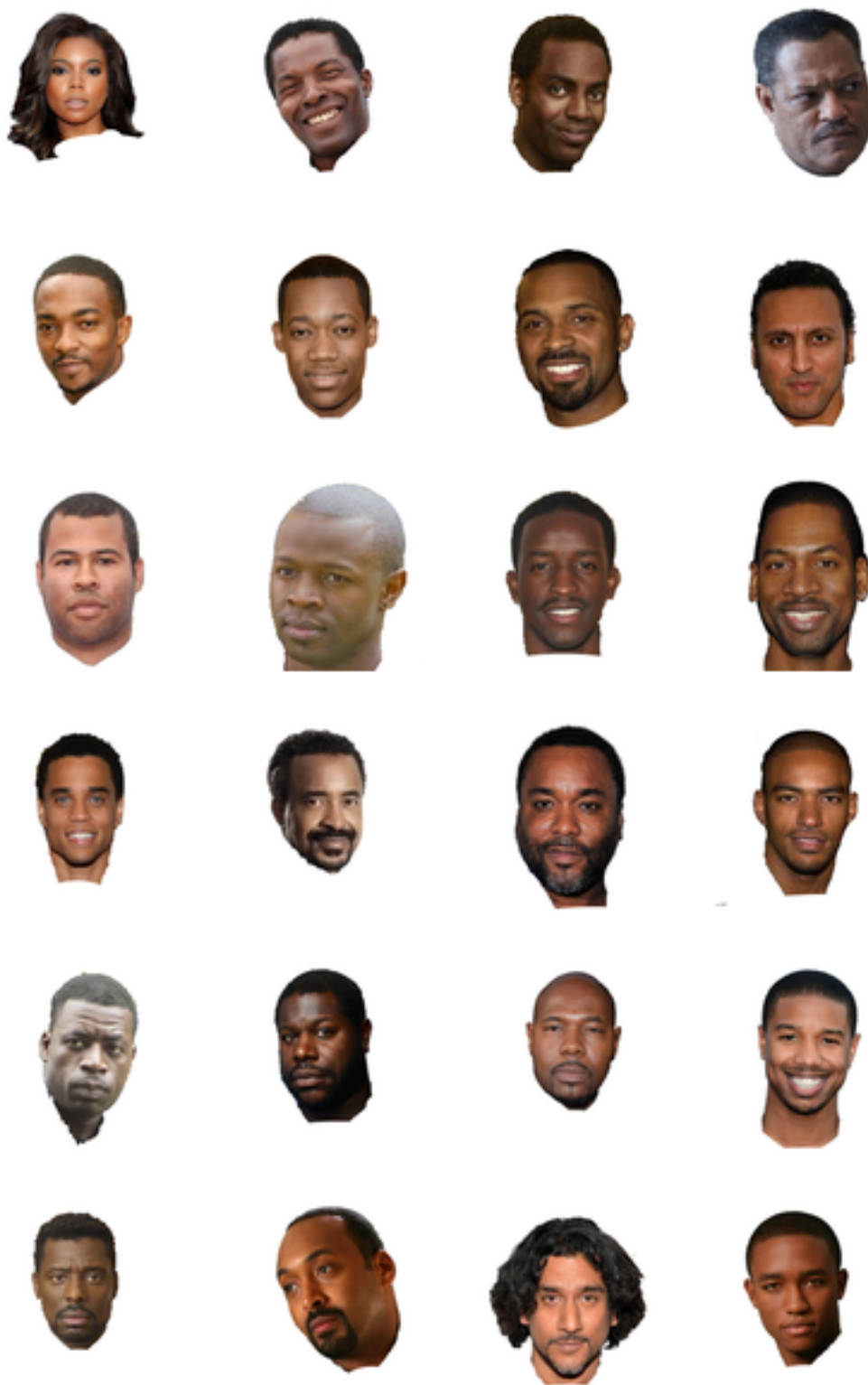


Figure B.8. Grid 7 of the faces used in the experiments.



Figure B.9. Grid 8 of the faces used in the experiments.



Figure B.10. Grid 9 of the faces used in the experiments.



Figure B.11. Grid 10 of the faces used in the experiments.


C. INTERFACES

This appendix contains all the interfaces used in the process of testing and evaluating the system.

C.1 IntoFocus Experiment Interfaces

The following interfaces were used in the IntoFocus experiments in chapter 3.

1. Highlight any of the following that you find in the image below: faces, tattoos, phone numbers. Use the smallest highlight possible to cover all the personal information. From the faces on the right select an actor/actress that can be seen in the image.












Change Highlight Width ▾ Undo Highlight

2. Can you see where any faces are (even if you can't recognize them)?

☐ Yes

☐ No

3. Try to find one of these actors in the image (at left). It might be difficult to find. **BONUS:** \$0.16 if you select the correct actor/actress; or \$0.02 if you select *Don't know*.

4. What allowed you to identify the actor?

☐ The scene ☐ The face is visible

☐ The clothes ☐ Tattoo

☐ Remember the movie

☐ Was not able to identify ☐ Other

5. Other comments on this image

Figure C.1. Interface of the IntoFocus experiment one in chapter 3

Image 1 Image 2 Image 3 Image 4 Image 5

1. Highlight the entire face, hair, hat (if any). Important: Read instructions at top before you begin.

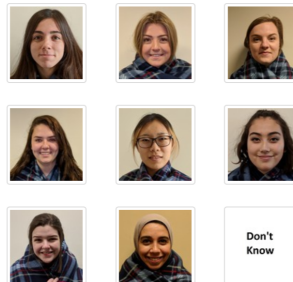


Change Highlight Width Undo Highlight

2. Can you see where any faces are (even if you can't recognize them)?

- ☐ Yes
- ☐ No

3. Try to find one of these people in the image. It might be difficult to find.



4. Comments on this image

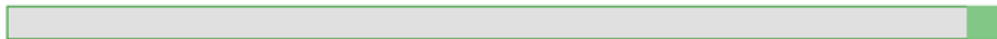
Figure C.2. Interface of the IntoFocus experiment two in chapter 3

C.2 Human Perception Study Experiment Interfaces

The following are the interfaces used in the chapter [4](#).



Detectability. Move the slider until the image is just clear enough to discern the location of any face (regardless of whether it is clear enough to be identified).



Identifiability. Move the slider until the image is just clear enough to discern whether or not you recognize the largest face (regardless of whether you actually know the person).

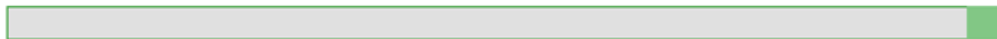


Figure C.3. Interface of the first perception study in chapter [4](#)

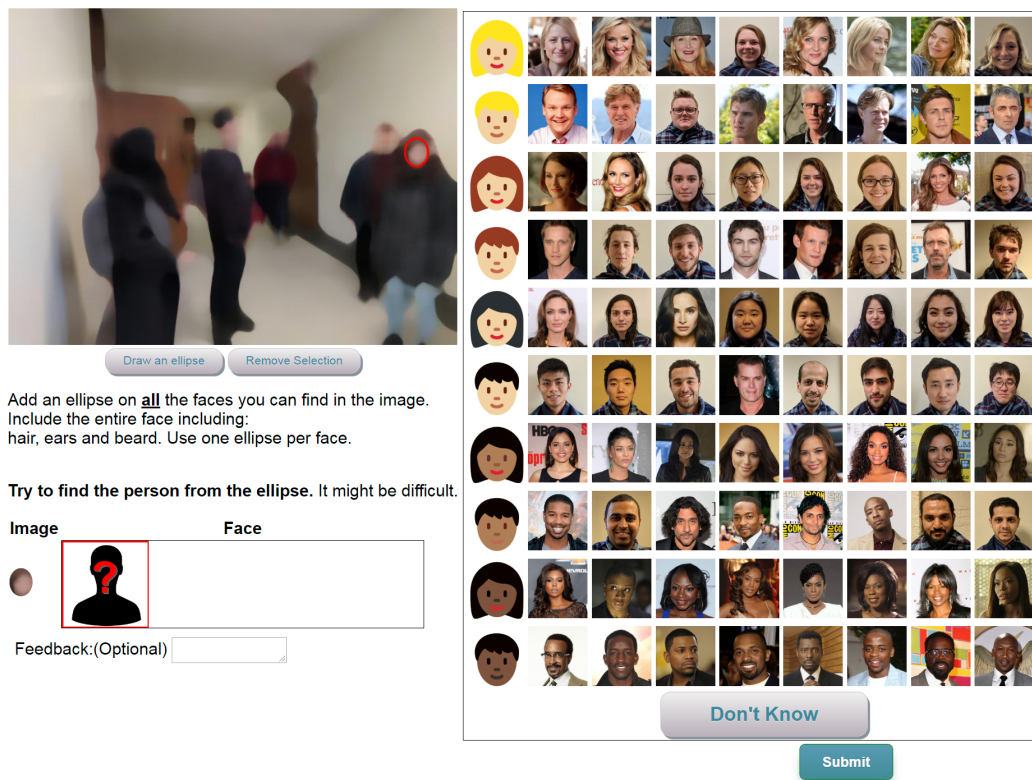
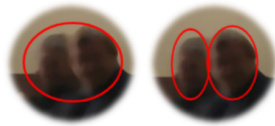


Figure C.4. Interface of the second perception study in chapter 4

Guidelines

Each ellipse covers only one face



✗
Wrong

✓
Right

Cover the entire face including hair and facial hair



✗
Wrong

✓
Right

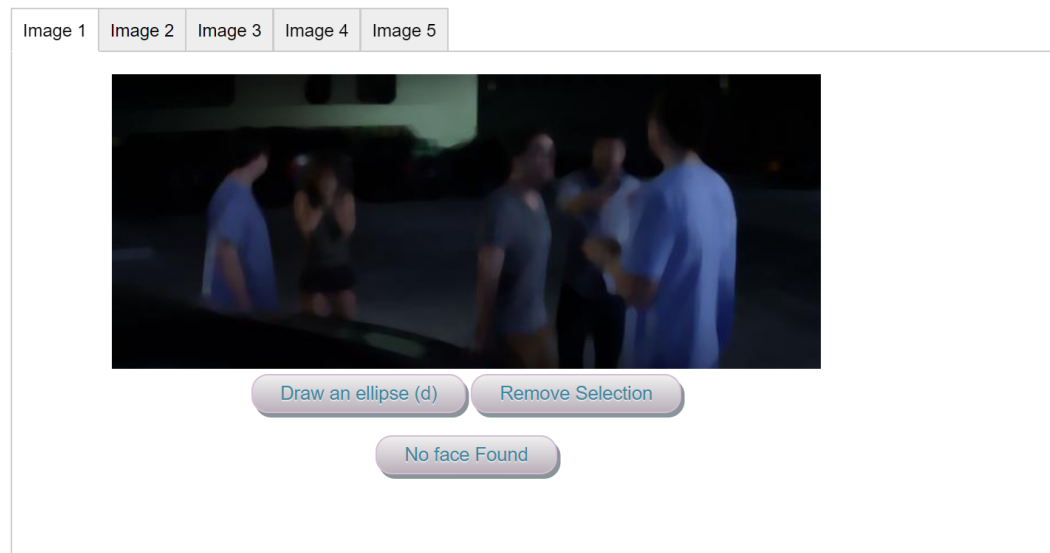
If unsure if something is a face do not mark it



✗
Wrong

✓
Right

Shortcuts: Press "d" to start drawing an ellipse.



Submit

Figure C.5. Interface of the detection aspect of the perception study in chapter 4

The red ellipse in the image below indicates the location of a face in the image. Try to identify the person shown in the red ellipse by selecting the person(s) you think it might be from the smaller faces to the right. Select all of the smaller faces that might be the same person as the one in the red ellipse. If you are confident that it matches only one of the reference faces (at right), then select only that one reference face. This HIT will only work on Google Chrome.

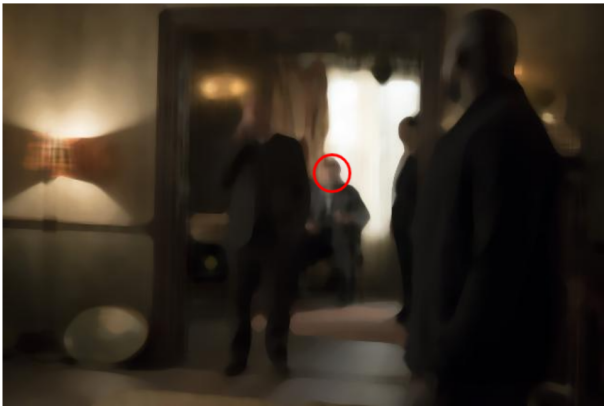
Image 1









Image 2

Image 3

Image 4

Image 5





Any of these

None of these

Submit

Figure C.6. Interface of the identification aspect of the perception study in chapter 4

C.3 Pterodactyl System Evaluation Interface

The following are the interfaces used in the chapter 4.

Add ellipses on ALL faces, including hair. Be sure you have covered all parts of the eyes, nose, lips, eyebrows, forehead, cheeks, ears, chin. From the faces on the right, select the faces that can be seen in the image.

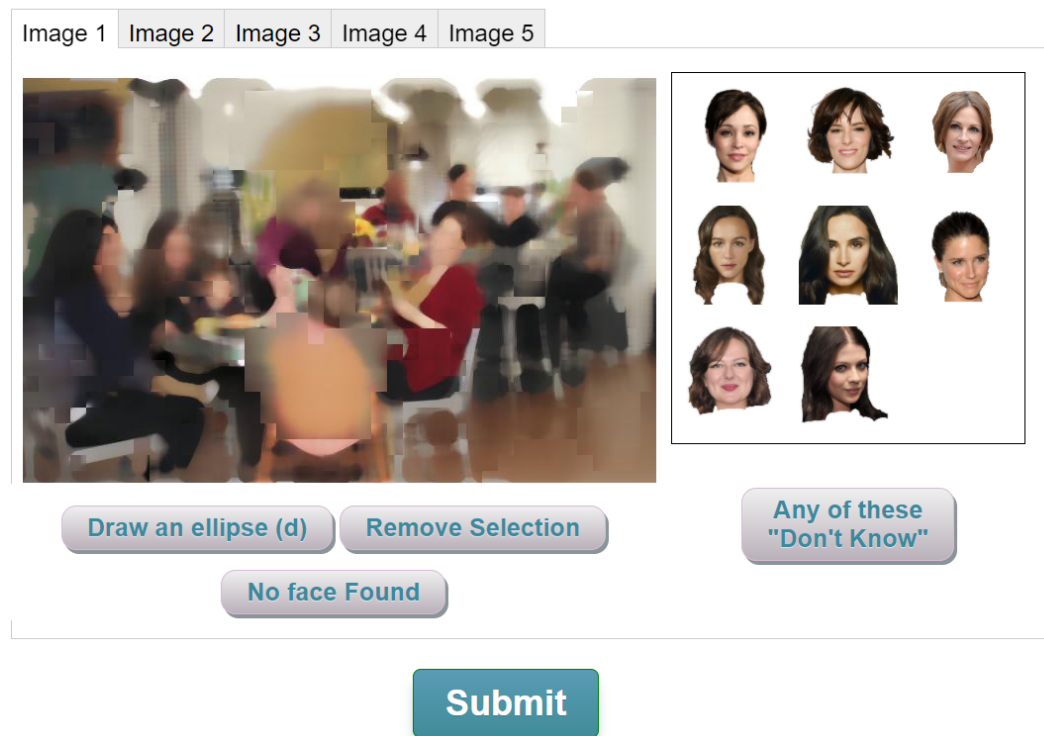
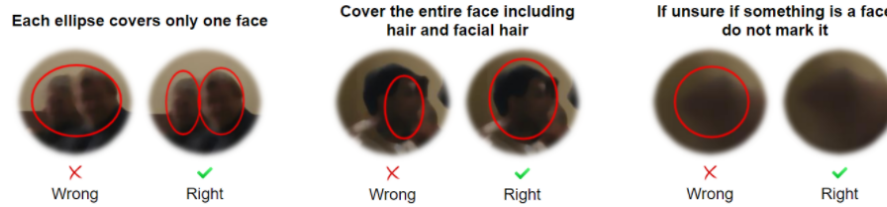


Figure C.7. Interface of the experiment to evaluate the Pterodactyl system in chapter 5

D. CONSENT FORMS

This chapter contains all the consent forms used in the experiments in this dissertation.

RESEARCH PARTICIPANT CONSENT FORM

Alexander J. Quinn

School of Electrical & Computer Engineering
Purdue University

Option A

What is the purpose of this study? We are trying to learn how to enable people to view online videos and get some kinds of information from them without revealing the identities of the people in the videos. This research is being conducted by Abdullah Alshaibani and Alexander J. Quinn at Purdue University.

What will I do if I choose to be in this study? You will be given a near vision test. If you have normal near vision (20/20), you will then be asked to view a few videos and tag people, objects and actions that you see.

How long will I be in the study? Participation is expected to last about 5 minutes per task.

What are the possible risks or discomforts? Exposure to sensitive video content is a potential risk. The videos may contain fictional depictions of violence, such as hitting, punching, kicking, or aiming guns at people. If you decide to participate, you may choose to stop at any time. Loss of confidentiality (e.g., your personal information) is a potential risk. The risks from participating are minimal, and no greater than in daily life.

Are there any potential benefits? This research is not designed to benefit you personally, but the results may help us build better systems for partial disclosure of videos (e.g., surveillance or police body camera videos) with the public.

Will I receive payment or other incentive? You will receive \$0.75 for each task that you complete. In addition, you will receive a bonus of \$0.10 for each correct tag minus \$0.10 for each incorrect tag. The bonus cannot be less than \$0.00 or more than \$2.00. You will be responsible for any taxes assessed on the compensation.

Will information about me and my participation be kept confidential? We will do our best to let you participate anonymously. Your identity will never be shared. All the collected data will be stored on password-protected server at Purdue, and may be retained indefinitely for possible use in future research. Project records will be accessible to the researchers (Alexander J. Quinn and Abdullah Alshaibani), and may also be reviewed by the departments at Purdue University that are responsible for regulatory and research oversight.

What are my rights if I take part in this study? Your participation in this study is voluntary. You may choose not to participate. If you agree to participate, you may withdraw your participation at any time without penalty or loss of benefits to which you are otherwise entitled.

Who can I contact if I have questions about the study? If you have questions, comments or concerns about this research project, please contact one of the researchers: Alexander J. Quinn or Abdullah Alshaibani.

If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (XXX) XXX-XXXX, email (XXX) or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.,
West Lafayette, IN 47907-2114

Documentation of Informed Consent. Clicking the “I agree” button below constitutes an electronic signature affirming that you are at least 18 years of age, you have read this consent form, all of your questions (if any) have been answered to your satisfaction, and you voluntarily agree to participate in this research study. You may print a copy of this consent form.

I agree

I do NOT agree

Figure D.1. The consent form used in the IntoFocus experiments in chapter 3.

RESEARCH PARTICIPANT CONSENT FORM

IntoFocus Face Perception Study
Alexander J. Quinn
School of Electrical & Computer Engineering
Purdue University

Key Information. Please take time to review this information carefully. This is a research study. Your participation in this study is voluntary which means that you may choose not to participate at any time without penalty or loss of benefits to which you are otherwise entitled. You may ask questions to the researchers about the study whenever you would like. If you decide to take part in the study, you will be asked to sign this form, be sure you understand what you will do and any possible risks or benefits.

We are studying people's ability to detect and recognize faces in images that have been blurred. The study will last for up to one hour. You may stop at any time.

What is the purpose of this study? We are trying to learn the difference between people's ability to detect and identify faces in images. People are being asked to participate to define the levels of blur affect the perception of human faces and their locations. We would like to enroll a maximum of 200 people in this study. This research is being conducted by Abdullah Alshaibani and Alexander J. Quinn at Purdue University.

What will I do if I choose to be in this study? You will be asked to look at blurred images on a computer and indicate (1) when faces can be detected, (2) locations of the faces, and (3) when the faces are clear enough to be recognizable to someone who knew the person in the image.

How long will I be in the study? Participation is expected to last at most two hours. You may stop at any time.

What are the possible risks or discomforts? There are no greater risks than you would encounter in daily life. Breach of confidentiality is always a risk with data, but we will take precautions to minimize this risk as described in the confidentiality section.

Are there any potential benefits? This research is not intended to benefit you personally. However, we hope the research may someday lead to new online work opportunities for crowd workers.

Will I receive payment or other incentive? You will receive \$20.00 for participating.

You must sign a form with your name, address, and social security number (or other identifier).

According to the rules of the Internal Revenue Service (IRS), payments that are made to you as a result of your participation in a study may be considered taxable income.

Will information about me and my participation be kept confidential? We will do our best to let you participate anonymously. Your identity will never be shared. All the collected data will be stored on password-protected server at Purdue, and may be retained indefinitely for possible use in future research. Project records will be accessible to the researchers (Alexander J. Quinn

Figure D.2. The first part of the consent form used in the in lab portion of the perception study experiments in chapter 4.

and Abdullah Alshaibani), and may also be reviewed by the departments at Purdue University that are responsible for regulatory and research oversight.

What are my rights if I take part in this study? You do not have to participate in this research project. If you agree to participate, you may withdraw your participation at any time without penalty.

Who can I contact if I have questions about the study? If you have questions, comments or concerns about this research project, please contact one of the researchers: Alexander J. Quinn or Abdullah Alshaibani.

To report anonymously via Purdue's Hotline see www.purdue.edu/hotline

If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (XXX) XXX-XXX, email ([XXX](#)) or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.
West Lafayette, IN 47907-2114

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am prepared to participate in the research study described above. I will be offered a copy of this consent form after I sign it.

_____ Participant's Signature	_____ Date
_____ Participant's Name	
_____ Researcher's Signature	_____ Date

Figure D.3. The second part of the consent form used in the in lab portion of the perception study experiments in chapter 4.

RESEARCH PARTICIPANT CONSENT FORM

IntoFocus Face Perception Study
Alexander J. Quinn
School of Electrical & Computer Engineering
Purdue University

Key Information. Please take time to review this information carefully. This is a research study. Your participation in this study is voluntary which means that you may choose not to participate at any time without penalty or loss of benefits to which you are otherwise entitled. You may ask questions to the researchers about the study whenever you would like. If you decide to take part in the study, you will be asked to sign this form, be sure you understand what you will do and any possible risks or benefits.

We are studying people's ability to detect and recognize faces in images that have been blurred. Participation is expected to last about 5 minutes. You may stop at any time.

What is the purpose of this study? We are trying to learn the difference between people's ability to detect and identify faces in images. People are being asked to participate to define the levels of blur affect the perception of human faces and their locations. We would like to enroll a maximum of 500 people in this study. This research is being conducted by Abdullah Alshaibani and Alexander J. Quinn at Purdue University.

What will I do if I choose to be in this study? You will be asked to look at blurred images on a computer and indicate (1) when faces can be detected, (2) locations of the faces, and (3) when the faces are clear enough to be recognizable to someone who knew the person in the image.

How long will I be in the study? Participation is expected to last about 5 minutes. You may stop at any time.

What are the possible risks or discomforts? There are no greater risks than you would encounter in daily life. Breach of confidentiality is always a risk with data, but we will take precautions to minimize this risk as described in the confidentiality section.

Are there any potential benefits? This research is not intended to benefit you personally. However, we hope the research may someday lead to new online work opportunities for crowd workers.

Will I receive payment or other incentive? You will receive \$0.75 for participating.

Will information about me and my participation be kept confidential? We will do our best to let you participate anonymously. Your identity will never be shared. All the collected data will be stored on password-protected server at Purdue, and may be retained indefinitely for possible use in future research. Project records will be accessible to the researchers (Alexander J. Quinn and Abdullah Alshaibani), and may also be reviewed by the departments at Purdue University that are responsible for regulatory and research oversight.

Figure D.4. The first part of the consent form used in the online portion of the perception study experiments in chapter 4 and the online experiments in the evaluation in chapter 5.

What are my rights if I take part in this study? You do not have to participate in this research project. If you agree to participate, you may withdraw your participation at any time without penalty.

Who can I contact if I have questions about the study? If you have questions, comments or concerns about this research project, please contact one of the researchers: Alexander J. Quinn or Abdullah Alshaibani.

To report anonymously via Purdue's Hotline see www.purdue.edu/hotline

If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (XXX) XXX-XXX, email ([XXX](#)) or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.
West Lafayette, IN 47907-2114

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am prepared to participate in the research study described above. I will be offered a copy of this consent form after I sign it.



Figure D.5. The second part of the consent form used in the online portion of the perception study experiments experiments in chapter 4 and the online experiments in the evaluation in chapter 5.

VITA

Abdullah Bader Alshaibani graduated *summa cum laude* from Florida Atlantic University in Boca Raton, Florida, in 2012 with his bachelor of science in Computer Engineering. He then attended Purdue University to obtain his Master's in 2014, followed by his Ph.D. under the direction of Professor Alexander J. Quinn in Electrical and Computer Engineering. His research interests include crowd-powered systems, image processing, privacy, face redaction, machine learning, haptics, and robotics. Upon completing his Ph.D., he will Join the faculty at Kuwait University as a tenure track Assistant Professor.