EFFICACY AND ETHICAL IMPLICATIONS OF MASS SURVEILLANCE SYSTEMS USING CONVOLUTIONAL NEURAL NETWORKS

by

NICHOLAS AKIN

(Under the Direction of Jeremy Davis)

ABSTRACT

In recent years, law enforcement agencies have used artificial intelligence to predict crime both domestically and internationally. With the development of technology also comes changes to the legal landscape to protect an individual's privacy. The need to regulate privacy is indisputable, so a mass surveillance system built with standards to protect privacy can help law enforcement incarcerate or exonerate individuals. This thesis argues that a well-designed surveillance system can achieve this balance using state-of-the-art convolutional neural networks and a human-in-the-loop architecture. Once technical and ethical concerns have been addressed and appropriately resolved, mass surveillance systems can help policing agents exonerate the innocent, hold individuals responsible for their actions, and keep societies safe.

INDEX WORDS: Mass Surveillance, Convolutional Neural Networks, Privacy, Accountability, Regulation, Police, Order

EFFICACY AND ETHICAL IMPLICATIONS OF MASS SURVEILLANCE SYSTEMS USING CONVOLUTIONAL NEURAL NETWORKS

by

Nicholas Akin

B.S., University of Georgia, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

EFFICACY AND ETHICAL IMPLICATIONS OF MASS SURVEILLANCE SYSTEMS USING CONVOLUTIONAL NEURAL NETWORKS

by

Nicholas Akin

Major Professor:	Jeremy Davis
Committee:	Kimberly Van Orman
	Lefteris Anastasopoulos

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

DEDICATION

I dedicate this to my family for their love and support.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jeremy Davis for his advice through countless revisions of my thesis. I would also like to thank Dr. Kimberly Van Orman and Dr. Lefteris Anastasopoulos for their assistance while serving on my committee. Lastly, I would like to thank my friends for their support throughout my time at the University of Georgia.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Recently, artificial intelligence (AI) has gained significant attraction in the policing field. AI systems have been used in various applications such as predictive policing, as well as mass surveillance systems. Predictive systems are used to help law enforcement predict crime. While surveillance systems are able to analyze footage to identify criminals through facial recognition. Some policing units have even paired surveillance systems with predictive systems to monitor high density areas to prevent crime (Brayne, 2020).

Despite the benefit of AI in policing systems to reduce crime, the use of AI raises significant ethical concerns, most notably, surrounding privacy. As revealed by Edward Snowden in 2013, the NSA was monitoring the United States through intercepting telecommunications, video, and other surveillance techniques (Scheuerman, 2014). Predictive systems that utilize artificial intelligence operate through the analysis of large quantities of data to identify patterns and trends. By doing so, these systems provide policing agents with valuable insights that enable them to increase their effectiveness in deploying resources to serve the public. However, these systems use "big data" that have underlying biases to generate predictions. These biased predictions are then used by law enforcement.

Throughout this thesis, we will discuss the biases, success, and limitations of predictive and surveillance systems and provide reasons a surveillance system could

serve law enforcement and the community better than a predictive system. Also, we will establish the technical and ethical challenges that must be faced before implementing a mass surveillance system. Chapter 2 will introduce Convolutional Neural Networks and various state-of-the-art algorithms that are used in computer vision systems. Such networks have been rigor tested many times and have shown exceptional accuracy (>95%) in facial recognition. In Chapter 3, we will discuss the methods in which we evaluate network architectures and propose a surveillance system equipped with facial recognition that can be utilized by policing forces. Chapter 4 will review current uses of predictive and surveillance systems by policing agents. Chapter 5 will discuss various ethical views of privacy and the factors that must be evaluated to implement a mass surveillance system with facial recognition capabilities. We conclude with Chapter 6 and the future of surveillance systems in law enforcement.

CHAPTER 2

CONVOLUTIONAL NEURAL NETWORKS IN COMPUTER VISION SYSTEMS

2.1 History of Neural Networks

Throughout this chapter, we will provide a background on the architectures used to perform computer vision tasks (e.g. facial recognition, object detection). We will also discuss our selection of architectures to perform facial recognition in a mass surveillance system.

Neural networks are the backbone of computer vision tasks. The history of neural networks can be traced back to the 1940's. In 1943, Warren McCulloch and Walter Pitts authored a paper describing how neurons might work and modeled a simple neural network using electrical circuits (McCulloch & Pitts, 1990). Their model was able to classify a binary output which denotes whether an input belongs or does not belong to a given class, however, weights that determined the class label had to be adjusted manually (McCulloch & Pitts, 1990). In 1957, Frank Rosenblatt came up with the perceptron and illustrated that neurons can learn from data and adjust weights used to classify inputs without human contribution through algorithms such as Stochastic Gradient Descent (SGD) (Rosenblatt, 1957). The ability to learn to adjust weights automatically throughout the training stage of a model became the standard for about a decade. Unfortunately, research into neural networks stagnated after a 1969 publication by Marvin Minsky and Seymour Papert. Minsky and Papert demonstrated that a perceptron with a linear

activation function was only a linear classifier and unable to solve nonlinearly separable problems (Figure 1) (Minsky & Papert, 1969).

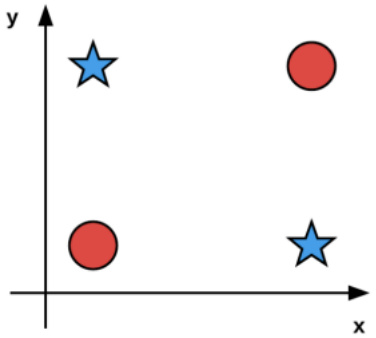


Figure 1: Nonlinear problem XOR

A nonlinearly separable problem (Figure 1) is a classification inseparable by a straight line. Attempting to divide the graph in Figure 1 with a straight line grouping the shapes of a 'star' and a 'circle' would prove futile. Fortunately, research was able to progress once again after the development of the backpropagation algorithm (Werbos, 1974; LeCun, 1998). The backpropagation algorithm enabled training for multi-layer feedforward neural networks (Werbos, 1974; LeCun, 1998). Combined with nonlinear activation functions, neural networks were now able to solve nonlinear problems (Werbos, 1974; LeCun, 1998). The backpropagation algorithm is still used to this day as a critical part of a modern neural network, but at its inception, slow hardware, and the lack of large, labeled training sets prevented researchers from training neural networks with more than two hidden layers (Le, 2019).

Today, research into neural networks is called 'deep learning'. Specialized hardware and large labeled training datasets enable networks to be trained with multiple hidden layers where lower layers learn simple concepts and the higher layers of the network learn abstract patterns. Convolutional neural networks are considered the most

robust architecture for state-of-the-art solutions to computer vision applications. Throughout the remainder of this chapter, we will discuss our selection of networks for facial recognition, their respective architectures, and the system architecture of a surveillance system.

2.2 Related Work

Surveillance systems can work simply by reviewing recorded video; but facial recognition systems are faster than manual review, accurate, and can work in real time. The task of facial recognition can be divided into two distinct problems. The first task at hand is the image localization/detection. Image localization involves finding the target inside an image in which the target is determined by the problem statement. The target in our case is the face of a person. The second task for a recognition system is classification. As our problem is identifying faces of individuals, once you have the face of the subject squared in a still frame, we can perform facial recognition through state-of-the-art Convolutional Neural Networks (CNNs).

Recently, CNNs have shown very promising success in large-scale image and video recognition (Krizhevsky et al., 2012; Sermanet et al. 2014; Simonyan & Zisserman, 2014; Zeiler & Fergus, 2013). Many of these recent CNN systems have been able to take advantage of high-performance computation systems with multiple GPU's or large-scale distributed clusters (Dean et al., 2012). The majority of CNNs are based on the original architecture by Krizhevsky et al (2012). Many of the advancements and improvements to the CNN architecture are largely due to the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) and similar image recognition conferences (Asian Conference on

Computer Vision, 2022; Computer Vision and Pattern Recognition, 2023; European Conference on Computer Vision, 2022; International Conference on Computer Vision, 2023; Russakovsky et al., 2015). The ILSVRC consists of a dataset containing images compiled to aid in research of image localization and classification. The dataset contains more than 20,000 categories and challenges competitors to create neural networks that correctly detect and classify images and videos. Developers compete to have the best performing accuracy with most able to classify within 95% accuracy (Russakovsky et al., 2015). The state-of-the-art success in recent years enables us to take advantage of robust architectures.

Facial recognition during most computer vision conferences is primarily performed on front facing, fully exposed faces. Facial recognition typically struggles with occluded faces, which is usually the case in real-world situations. Faces can be obscured for many reasons with a variety of clothing and headwear. With the rise of SARS-CoV-2 virus (COVID-19) in the Spring of 2020 people across the globe were advised by the World Health Organization to quarantine, social distance, and wear masks while in public spaces (Chang et al., 2021). KN-95 respirators, surgical masks, and home-made cloth masks became common place whether sick or not. These masks all have different characteristics for particulate filtering, but also shape and size on the face. Many of these masks when worn with a cap, glasses, or hood leave very little of the face, if any part, exposed.

The task of recognition through 3-d body models could be further explored and might yield results comparable to the methodologies in which we as humans perform identification, but for the purposes of this thesis, we will focus on facial identification and

classification. Researchers took note of deficiency in occlusions and have developed ways to combat occlusions in facial recognition. The two main approaches are to removal and suppression of undesirable features in an image (Long et al., 2017; Qiu et al., 2021). The technique of removing attempts to remove/suppress the parts of the face that are occluded by facial recognition. Recent approaches at identification with facial occlusions have shown that facial recognition systems have been able to achieve a 96.03% accuracy with a lower face occlusion and an 81.12% accuracy with upper face occlusions (Qiu et al., 2021). These results suggest that the upper face contains more highly identifiable information than the lower face. Success in computer vision research allows us to visit robust facial detection and facial recognition algorithms that are invariant to changes in pose, luminance, and occlusions.

2.3 Basics of a Convolutional Neural Network

Before we go into our selection of CNNs, we must first discuss how they are made. A CNN is an evolution of a multi-layer neural network. CNNs are built primarily by stacking multiple different layers each with different tasks. Most CNNs follow a standard architecture consisting of alternating convolutional (CONV) layers and pooling layers. The final layers of a CNN are usually fully-connected (FC) layers with a softmax classifier used to predict the final class output. Training of a CNN utilizes a backpropagation algorithm such as SGD to automatically learn the weights that minimize the loss functions that perform the prediction.

Convolutional layers are used to apply CONV filters. A CONV filter is a matrix of numbers (e.g., 3x3, 5x5) that extract features from the input matrix (an input matrix is

a matrix of pixel values that represent the input image). The output of the CONV filter is called a feature map. The feature map is generated by sliding the CONV filter over the input matrix and computing the dot product of the values. The result is summed into a single value representing the feature which, as a result, reduces the dimensionality of the image.

The next layer used in a CNN are pooling layers. Pooling layers are used to reduce the spatial size (width and height) of an input matrix. Thus, pooling preserves features in the image while reducing the number of parameters and computation needed in a CNN. The pooling operation is typically completed through a maximum (max) pooling or average (avg) pooling. Pooling is performed by using a matrix (e.g., 2x2) to take the average or maximum value within the filter. This results in a spatial size reduction and is performed throughout the feature map in a manner that mirrors the CONV layer.

Lastly, once the input matrix is sufficiently small, fully-connected (FC) layers are applied to have a final output prediction. FC layers are based on neurons which are also seen in regular neural network architectures. The neurons are used to discriminate features from the feature map. The weights for each neuron are updated through back propagation and the last set of neurons are used to make a class prediction. The last FC layer has *n* neurons where, for a classification problem, each neuron represents a distinct class that the FC layers attempt to predict.

2.4 Facial Localization Selection

Now that we have established the basics of a CNN, we can use them for the task of facial recognition. As described earlier, the first task of facial recognition is facial localization/detection. Through this process, we obtain a cropped image of just the face we are trying to predict. For the task of localization, there are a couple of privately owned state-of-the-art solutions, but the publicly available CNN based DNN (Deep Neural Network) from OpenCV provides us an open-source state-of-the-art solution. For our task, we take advantage of the robust architecture of OpenCV's DNN which we can call "YuNet" named after one of its creators Dr. Shiqi Yu (Feng et al., 2021). It has been proven to be adequately fast and the most robust out of the other popular detectors (Nelson, 2022; Gupta, 2022; Rosebrock, 2021). Alternatives to OpenCV's DNN include Dlib's HoG and MMOD detectors; however, they are not as robust as OpenCV's DNN for our application but in some cases have a speed advantage. For our purposes, we take advantage of the robust nature of YuNet in our surveillance system but acknowledge the great achievements of OpenCV's and Haar Cascade Face Detector, Dlib's HoG (Histogram of Oriented Gradients), and Dlib's MMOD (Max-Margin Object Detection).

Comparing the various face detection methods, it is shown that YuNet is the best of the other methods. YuNet's architecture is a lightweight CNN famous for its accuracy and robustness being invariant to changes in pose, luminance, occlusion, and size (Nelson, 2022). Despite it being slower than some of the other architectures discussed (Figure 2), it performs in what is considered real time and its robust architecture makes it one of the best for a surveillance application. YuNet is able to identify faces in many

conditions of luminance, scale, and non-frontal images. The ability for YuNet to detect faces under real-world imagery conditions makes it ideal for a surveillance application.

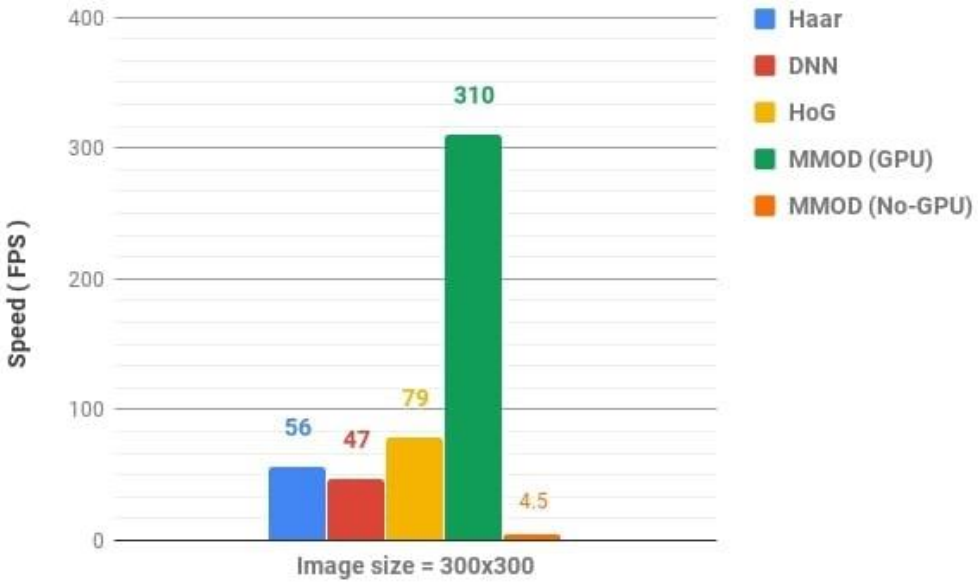


Figure 2: Speed in Frames per Second (FPS) at which each detector achieves

We start our comparison with the Haar Cascade Classifier. Proposed in 2001, Viola and Jones (V-J) created the first ever object detection framework known as Haar Cascades (Rahmad et al., 2020). The V-J face detection algorithm works through feature detection. Applied to facial detection, it scans an input image for features that resemble a human face (Rahmad et al., 2020). The Haar features classify the intensity of the pixels in the region and are represented by rectangles shown in Figure 3. The Haar cascade architecture excels at speed and scale of the input images but falls short when identifying images with different poses and occlusions which results in many false positives.

Figure 3: HAAR Cascade Classifier feature detection

Next, we look at the HoG (Histogram of Oriented Gradients) from Dlib. Proposed in 2005, the HoG architecture is equipped with a HoG feature descriptor with aspects from a SVM (Support Vector Machine) to perform facial detection (Rahmad et al., 2020). The HoG portion of the architecture works by dividing the image into cells then computing the histogram for each cell (Rahmad et al., 2020).

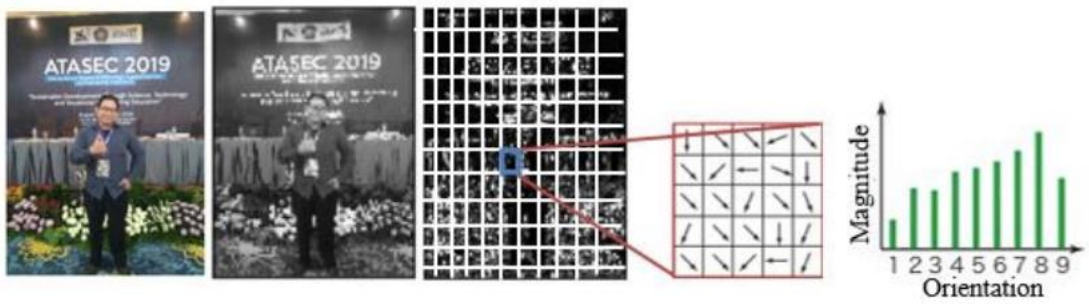


Figure 4: Histogram of Oriented Gradients (HOG) gradient mapping

After the HoG (Figure 4) is computed, it is passed to a trained SVM which determines a binary classification of if the vectors represent a person or not. HoG and SVM excels at CPU only performance and is able to identify with minor occlusions performing better than Haar Cascade (Rahmad et al., 2020). Though, HoG falls short when trying to identify faces with small representation (< 80x80 pixels), as well as identification with substantial occlusions (Gupta, 2022). Another downside to HoG and SVM as a facial detector is the bounding box produced by the detector. The detector fails to produce a bounding box in the extremes of off-center views and eliminates usable portions of the face creating a tighter crop (Figure 5).

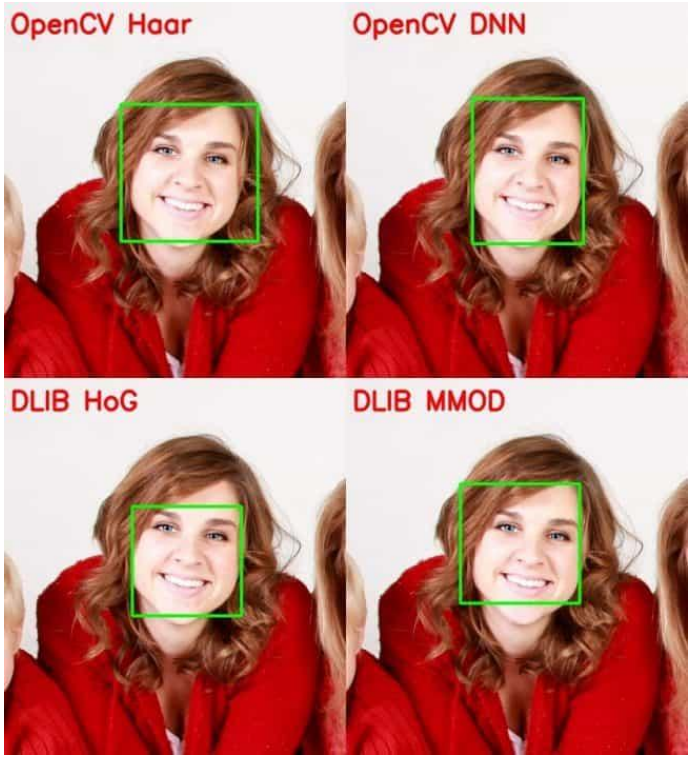


Figure 5: Bounding Box Comparison between facial detectors

The last detector we consider is Dlib's MMOD (Max-Margin Object Detection). Like YuNet, MMOD is a CNN based architecture. The CNN is applied to input images at different scales and positions to detect faces of varying sizes. MMOD uses a set of anchor boxes to define the possible locations and sizes of the faces and during training, the model adjusted the location and size of the boxes to match the face location. MMOD introduced the max-margin loss function which optimizes larger margins between positive and negative samples to distinguish face from background in the image. This helps the robust nature of the classifier through invariance to changes in pose, luminance, and expression. MMOD's architecture has the same benefits as YuNet through both of their robust natures and fast performance on GPU (Figure 2) (Gupta, 2022). Though,

MMOD's larger architecture is slower on CPU than YuNet and does not work well with small faces (< 80x80).

All of the various facial detectors have their benefits, but YuNet proves to be the most robust. It is fast on GPU enabled devices and is able to overcome problems in facial detection by the ability to identify faces in off-axis views, faces at the largest range of scale, and detection of occluded faces. The YuNet architecture is easily deployable on a large-scale system while not taking as many computational resources as other architectures. For our application, we utilize the YuNet architecture because of its robust characteristics.

2.5 Facial Classification Selection

With the help of YuNet, we can now perform facial classification. We can use the bounding boxes produced by YuNet to crop the image to just the face which reduces noise in the image. The facial crop of the image lets us take advantage of research into facial embedding generation for classification through the CNNs VGGFace and Resnet50. As discussed in Chapter 3, the ability for the CNNs to make meaningful embeddings from the facial crops will be compared through a k-NN (K-Nearest Neighbors) model. Next, we will consider and evaluate the performance of the VGGFace architecture, and ResNet-50 architecture to generate representative facial embeddings for facial recognition.

The VGGFace Model is based on the original architecture inspired by Ciresan et al. (2011) and Krizhevsky et al. (2012). Made in 2014 by the Visual Geometry Group from the University of Oxford, the group made a couple different versions of their VGG

architecture (Simonyan & Zisserman, 2014). The VGG group made VGG-16 and VGG-19 where the numbers in the CNN's names only differentiate the total numbers of the layers in the network. The original VGG architectures were trained to perform object classification learning weights from the ImageNet dataset. Later, the VGG group decided to see how their model performed with facial classification and developed what is coined VGGFace.

VGGFace consists of a VGGFace dataset and VGGFace model. The VGGFace dataset consists of 2.6 million facial images of 2,622 unique identities (Cao et al., 2018). Their custom dataset was used to train their model. The VGGFace model is based on the original VGG-16 architecture (Figure 6). The difference between the original VGG-16 architecture and the VGGFace model is the FC layers which contain 2,622 neurons to output to the 2,622 identities of the VGGFace dataset instead of the 1,000 identities of the ImageNet dataset in the original VGG-16 architecture. The weights learned by the models are, of course, different as well. The purpose of the VGGFace model was to take the effective performance of image classification that VGG-16 produced and apply the architecture to facial recognition tasks (Cao et al., 2018). Instead of training the VGGFace model on the ImageNet dataset the model is trained on the VGGFace dataset to learn the weights as described above to produce accurate results for facial recognition.

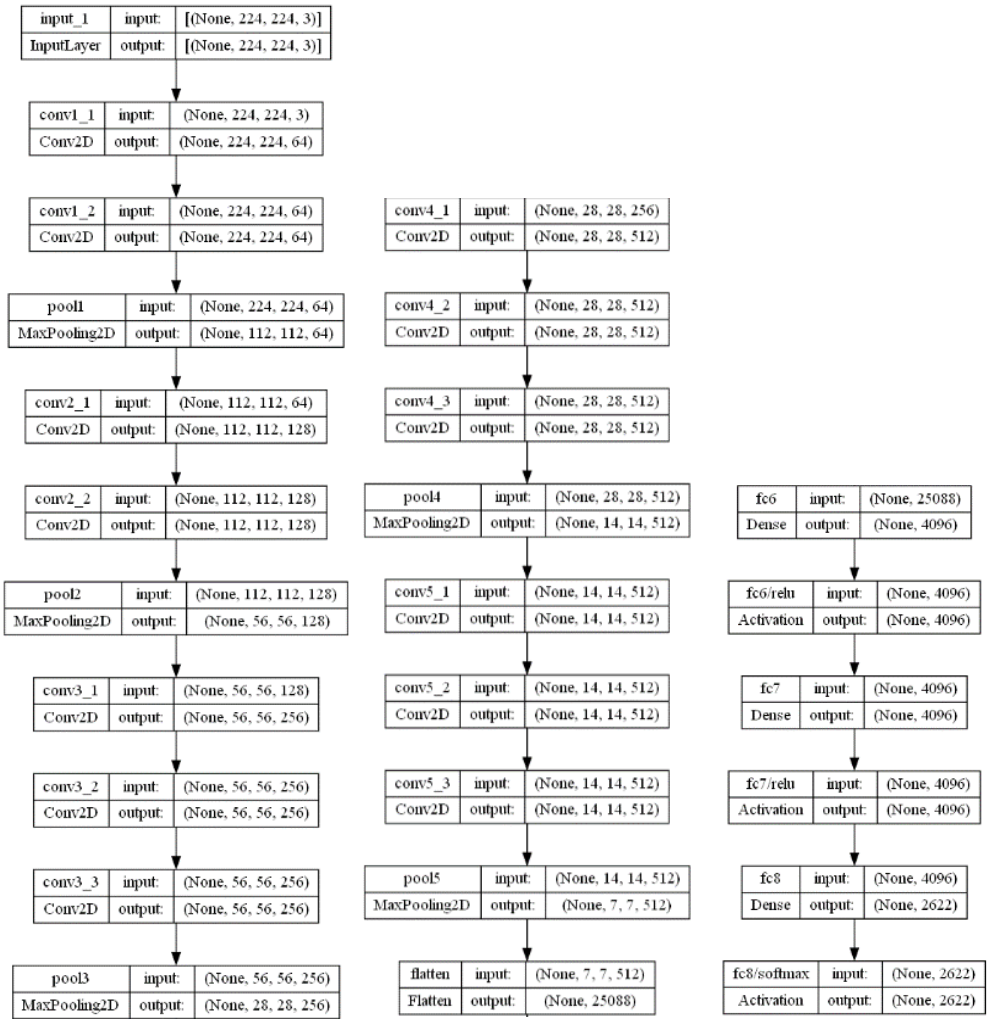


Figure 6: Left - Input to 3rd block. Center - 4th block to flatten layer. Right - FC layers to prediction layer

The VGGFace model consists of 6 distinct blocks. The 1st block contains the input layer which accepts an image of the dimensions of 224x224 and 3 channels (RGB). The block then has 2 convolutional layers followed by a max pooling layer. The 2nd block follows the same structure as the 1st but without the input layer. The 3rd, 4th, and 5th blocks each have 3 convolutional layers followed by a max pooling layer. The 5th block ends a flatten layer to change the dimensions for the fully connected (FC) layers. The 6th block is home to the fully connected layers and output layer.

One of the main contributions by the original VGG architecture was the use of small convolutional filters. In the VGG architecture, the convolutional filters have a size of 3x3 which is the smallest filter size that can capture the notion of up/down, left/right, and center (Simonyan & Zisserman, 2014). The convolutional filter is applied on the matrix of pixel values with a stride of 1 which shifts the filter matrix by 1 pixel after each CONV is applied. The VGGFace model consists of the same base architecture, so utilizes the same 3x3 filters.

As depicted above, each convolutional stack is followed by spatial pooling performed through 5 max-pooling layers. These max-pooling layers reduce the spatial size of the feature map. Each max pool filter is a 2x2 window with a stride of 2.

The VGGFace model ends with 3 fully connected layers followed by a softmax activation function. The fully connected layers are each equipped with a ReLU activation function for benefit of use in nonlinearly separable problems (Figure 7). The output layer of the network has 2,622 neurons to match predictions of the VGGFace dataset's 2622 classes.

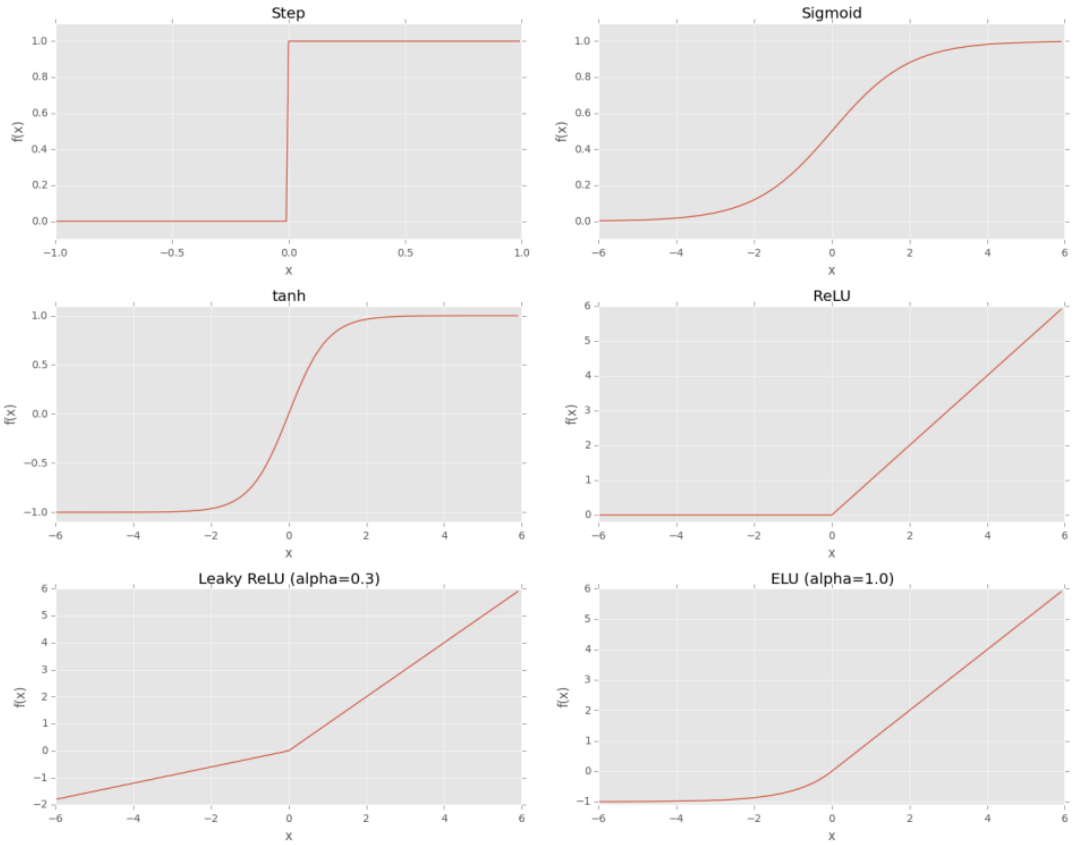


Figure 7: Various activation functions used in neural networks

Next, we evaluate the ResNet architecture which is based on the original VGG architecture. Proposed by a research group from Microsoft in 2015, the group developed multiple ResNet models with the most common being ResNet50 and ResNet101 (Figure 8) (He et al., 2015). The researchers were able to go up to 152 layers while still having a lower complexity than the VGG nets (He et al., 2015).

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 8: ResNet architectures

The lower complexity allows for the model to be trained quicker with fewer parameters all while being significantly deeper than the VGG-16/19 architecture. Previous deep models have struggled with an increased error as model depth increased. Prior to the ResNet architecture, deeper networks were theorized to increase performance, but suffered from the problem of vanishing gradients. The problem of vanishing gradients occurs as more layers are added using improper activation functions. The gradients that are found during back propagation become very small and approach 0 which makes training a deep neural network difficult. Activation functions such as the sigmoid function cause large inputs to be mapped into a small input space between 0 and 1. So, a large change to the input will result in a small change to the output. Luckily, this problem has been addressed by through the exploration of alternative activation functions for deep neural networks such as ReLU (Figure 7) (He et al., 2015). Another problem with deeper neural networks is when they begin converging a degradation problem starts occurring. As the network grows deeper, the accuracy starts degrading, then decreases rapidly (He et al., 2015). The authors address the depth related degradation problem

through a deep residual learning framework that utilizes shortcut connections to perform identity mappings (Figure 9).

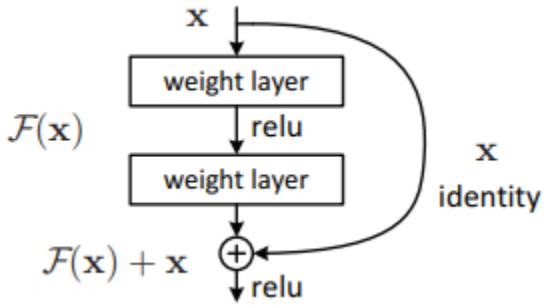


Figure 9: Building block for shortcut identity connections in ResNet

To perform identity mappings, the output is added to the outputs of the stacked layers. A network utilizing shortcut connections can be trained end-to-end by SGD with back propagation while not adding computational time or an extra parameter (He et al., 2015). The building blocks to the residual networks are formally defined as:

$$y = F(x,\{W_i\}) + x \quad (1)$$

In Eqn. (1), the x and y are the input and output vectors of the respective layers considered. The function $F(x,\{W_i\})$ is representative of the residual mapping that is learned. The *x* added to the end of Eqn. (1) is the identity connection. This allows initial information to the block to be added to the output of the function *F*. Effectively, this identity mapping allows the model to preserve information from the original input that could be lost during the forward pass through the function *F*.

The authors introduced their residual design through inspiration of the VGG architectures. The first network they made is called the plain network which has 3x3 convolutional filters with the following 2 design rules:

1) For the same output feature map size, the layers have the same number of filters (He et al., 2015).
2) If the feature map size is halved, the number of filters is doubled to preserve the time complexity per layer (He et al., 2015).

The plain networks have convolutional filters with a stride of 2 and performs down sampling after the convolutional layers. The architecture ends with a global average pooling layer and a 1000-way fully connected layer with a softmax function totaling 34 layers (Figure 10).

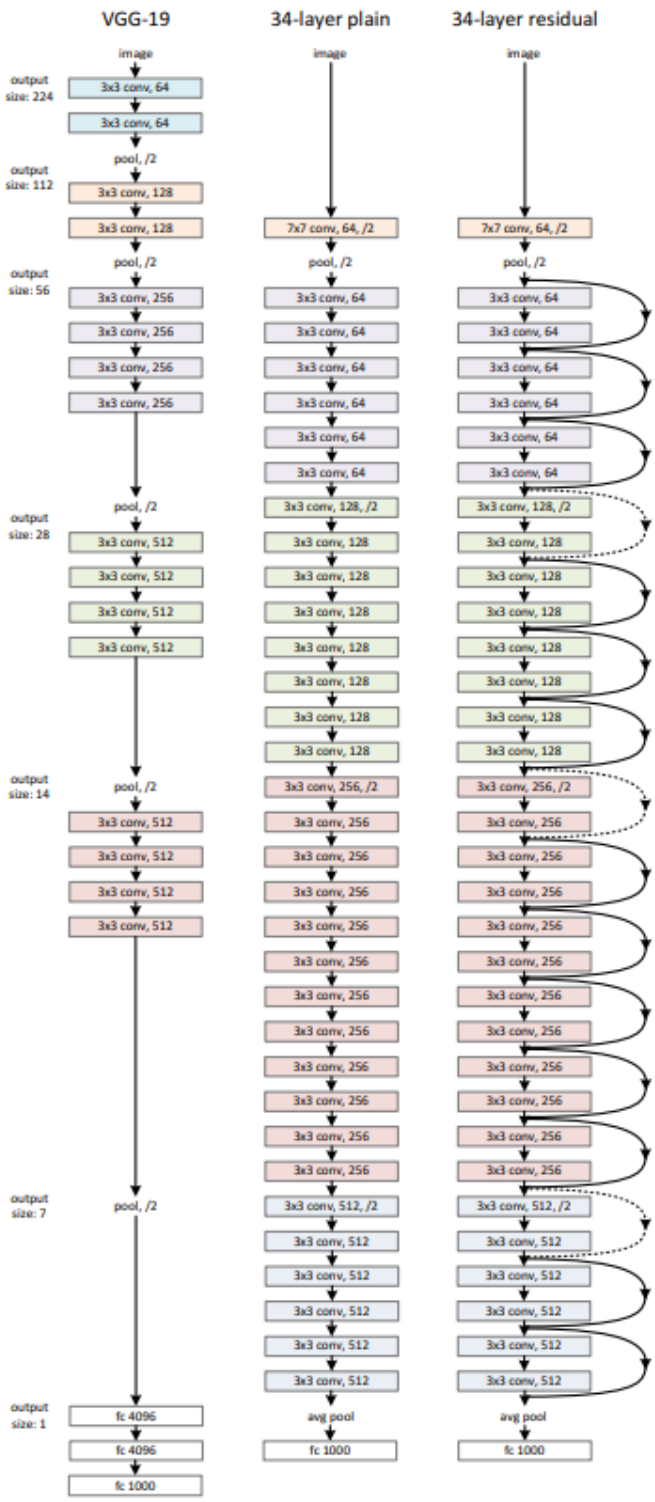


Figure 10: Left – VGG-19 Model. Center – Plain network. Right – Residual Network

The residual network (Figure 10, right) is based on the initial plain network. It is equipped with the shortcut connections which turn it into a residual network. The identity connection can be directly utilized by the network when the input and output are the same dimensions (solid line shortcut connections). Alternatively, matching dimensions can be performed by 1x1 convolutions where both cases use a convolutional stride of 2. When the dimensions are increased (dotted line shortcut connections), the shortcut can perform identity mapping with zero padding to account for the increased dimensions. The described architectures apply for ResNet-18 and ResNet-34, but due to a concern for training time, the residual building block described in (Figure 11, right) is modified for the larger ResNet architectures.

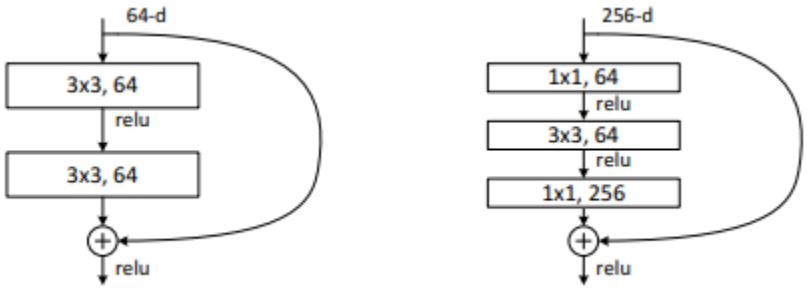


Figure 11: Left – Building block for ResNet-18/34. Right – Building block for deeper ResNet architectures

In deeper ResNet architecture, the residual blocks utilize a stack of 3 layers instead of 2. This replacement results in a 50-layer ResNet architecture. The 101 and 152-layer ResNet architectures are built using more 3-layer blocks all while still maintaining a lower complexity than the VGG network architectures. The authors found that the deeper ResNet architectures are more accurate than the ones using the 2-layer building block and are not prone to the degradation problem, so the authors were able to experiment with very deep ResNet architectures with increased accuracy. When using the deeper

architectures, they also trained an excessively deep 1202-layer architecture and only found a small decrease in performance (~1.5%) (He et al., 2015). They attributed this to overfitting, so as increasingly large datasets become readily available, a very deep ResNet architecture might become the best solution.

CHAPTER 3

EXPERIMENTATION

3.1 Dataset

For our experiments with the two CNN models, we want to evaluate their accuracy when generating facial embedding. To experiment, we utilize the Labeled Faces in the Wild (LFW) dataset. The LFW dataset was created by researchers at the University of Massachusetts for the purpose of studying unconstrained face recognition. The dataset is a commonly used classification benchmark for machine learning models used in many experiments. The dataset contains over 13,000 images of faces collected from the internet with name labels of the individual pictured. The images were collected using the Viola-Jones Object Detection Framework (Huang et al., 2012). V-J combines the concepts of Haar-like Features, Integral Images, the AdaBoost Algorithm, and the Cascade Classifier to make a system for accurate object detection. Of the 13,000+ pictures, there are 1680 classes (name labels) that contain more than 2 distinct photos in the dataset. LFW was updated throughout its lifetime to include 3 subsets of the original. The subsets contain different types of "aligned" images. The aligned images include "funneled" images. These images include an unpublished method of image alignment. There is also a "deep funneled" images set. According to the authors, LFW-a (funneled images) and "deep funneled images" produce the most accurate results for facial classification and recognition algorithms as demonstrated in ICCV 2007. For our tests, we used the "deep funneled" images in pursuit of the most accurate results from our models. We divide the

LFW dataset into a directory structure (Figure 12) in which each class holds the images for the respective identity.

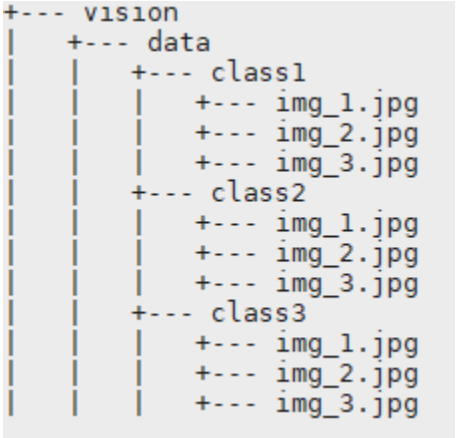


Figure 12: Directory structure of images divided into classes

For the purposes of facial classification, the set is divided into 4 parts (Table 1). Within each LFW split, the number represents the number of images per class.

Table 1: Partitions of LFW Dataset

	LFW_50	LFW_20	LFW_10	LFW_5
# Images	1553	1240	1580	2115
# Classes	12	62	158	423

The exception to the rule is the 'LFW_50' set where each class has at least 50 images (many have more). The 'LFW_50' split has an average of 129 images per class and is uncapped to evaluate the performance of ample comparison data. The other 3 classes (LFW_20/LFW_10/LFW_5) are capped at 20/10/5 images per class respectively to test the performance of set limits on the number of images per class.

LFW is relatively small compared to datasets such as VGGFace (2,622 identities, 2M+ images) and the Microsoft Celeb dataset (100,000 identities, 8M+ images) and suffers from known issues, most notably bias in gender and in skin shade. For our

purposes, we use the dataset to test the ability of the various models to make representative embeddings. The significantly larger datasets mentioned are better equipped and must be used when training a model for facial recognition. Datasets with representational gender and skin shades, with the help of specialized CNNs, have been tested and shown to be generalizable to diverse populations (Gong et al., 2020). Our use of the model does not learn new weights from the VGGFace dataset and shows that a CNN trained for facial recognition can be generalizable to any data.

3.2 Implementation

Our goal through our test is to see if the VGG-16 based model and the ResNet-50 based model can generate facial embeddings of an individual that are significantly differentiable from other facial embeddings to perform facial identification. To do this, we make use of Keras. Keras is a high-level API that runs on top of Tensorflow (an open-source deep learning framework by Google) to facilitate ease when building deep learning models (Chollet et al., 2015). Keras allows us to utilize the pretrained weights from both of the models we test.

For the purpose of generating facial embeddings, we do not want the model to predict an output class from the training set. Traditional facial recognition performed by CNNs utilizes the last FC layer to perform a class prediction. However, this can only be performed after the model is trained and weights are learned for a specific dataset. This process needs sufficiently large datasets in order to learn the weights for the model. For a surveillance system that continuously adds individuals to identify, having to retrain such a model would be inefficient. So, to perform facial recognition we take advantage of the

learned weights of the VGGFace and ResNet-50 models to generate facial embeddings. We can obtain the embeddings the model generates by removing the FC layers from the models so that they end in an average pooling layer instead. We can then compare the embeddings to perform facial recognition rather than using the output prediction of the FC layers. This lets us utilize weights learned when training on sufficiently large datasets to generate facial embeddings without having to constantly retrain the model.

To generate facial embeddings, we evaluate on the LFW dataset splits as described in Section 3.1. Each image is ensured to be of the correct size (224x224) and converted from an image into an array of pixels. The array instances are of shape (224x224x3) which represent the width, height, and 3 color channels of an image (depth). These pixels are stored in a list with the labels (identity name converted into numerical label) for that set of pixels in a separate list. Each image instance is then passed into the modified VGGFace and ResNet50 models to generate facial embeddings. To store the embeddings and respective labels for comparison, we compress them into a '.npz' (NumPy zip) file.

3.3 Evaluation Metrics

We assess the VGGFace model and Resnet model using the partitions represented in Table 1. The partitions in the dataset serve to evaluate the accuracy, and impact of outliers/misclassifications produced by the model. To evaluate the VGGFace and ResNet architectures, we use a k-NN (k-Nearest Neighbors) algorithm. A k-NN allows us to determine the class corresponding to an embedding using their Euclidean distance (straight line distance). We access the k-NN from the reported accuracy. We also

visualize a k-NN's results in a scatter plot where the clusters resemble identities of the same class.

When evaluating the performance of a model, it is important to evaluate the model with other metrics alongside accuracy. Accuracy often does not indicate if the model is predicting with an even distribution or is just predicting a dominate class within a dataset. To further evaluate the performance of the models, we also utilize precision, recall, and F-measure.

Precision is a metric that evaluates the number of True Positives by the number of True Positives and False Positives. The precision metric indicates whether the model performs accurately. A high precision indicates a low number of False Positives, and a low precision indicates a high number of False Positives. A low precision indicates that the model is over-predicting a datapoint belonging to a class when it does not (false positives on an identity).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall evaluates the number of True Positives over the number of True Positives and False Negatives. A high recall indicates a low number of false negatives whereas a low recall indicates a high number of false negatives. A low recall reveals that the model is under-predicting a datapoint belonging to a class when it does.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

The F-measure is used to increase or decrease the importance of precision and recall for evaluation. The F-Score is evaluated through the formula below. The *beta* is

typically of value (.5, 1, or 2). Where .5 increases importance on precision. The value of 1 equally balances precision and recall, and the value of 2 provides more weight to recall.

$$F\text{-}Score = \frac{(1 + beta^2) \quad x \quad Precision \; x \; Recall}{(beta^2 \; x \; Precision) + Recall}$$

An F0.5-score (*beta* = 0.5) value is useful for cases where minimizing false positive is important rather than minimizing false negatives. An F1-score (*beta* = 1) balances the importance of precision and recall by using a *beta* value of 1. An F2-score (*beta* = 2) increases the importance of minimizing false negatives rather than minimizing false positives. In a surveillance system equipped with facial recognition, we value the minimization of false positives, so we evaluate it using a *beta* value of 0.5.

3.4 Results and Analysis

We evaluate both models through use of a k-NN. First, the data corresponding to the LFW split we are trying to compare is retrieved from the respective '.npz' file. To use a k-NN for comparison, we first have to train it. We divide the embedding data and labels into a training and testing by use of an 80-20 train-test split. The k-NN is trained on the majority of the data where it learns the relationship between the embeddings and the labels. When the test set is evaluated, the k-NN determines the distance of the embedding through the Euclidean Distance algorithm. With the number of neighbors set to '3', the k-NN looks at the 3 closest embeddings according to their Euclidean Distance and applies the appropriate class label. The results of the k-NN testing are listed in Table 2.

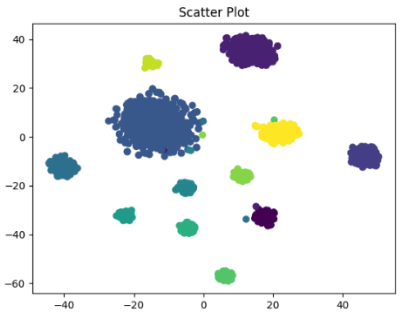
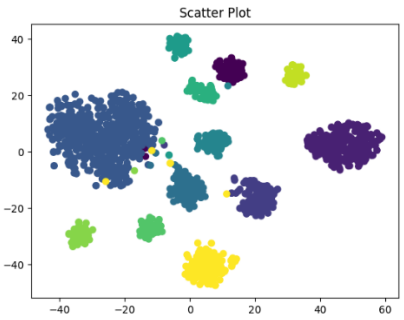
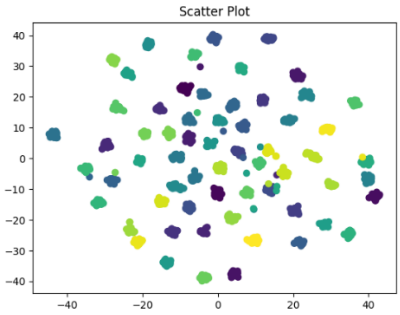
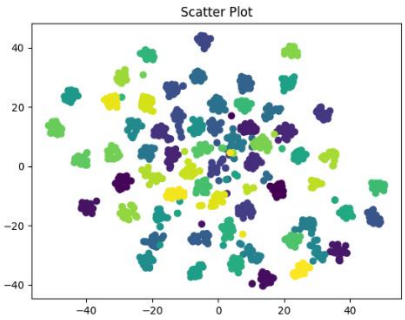
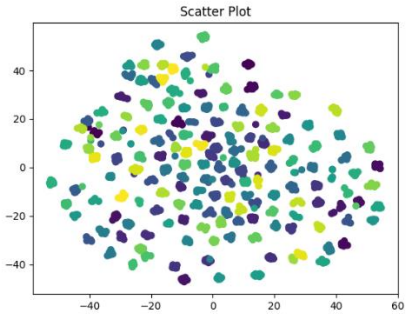
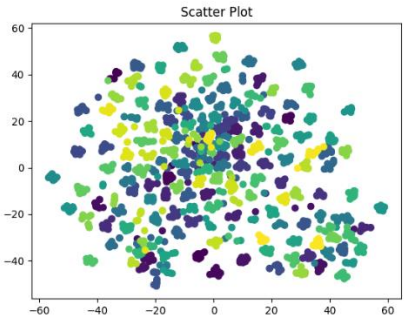
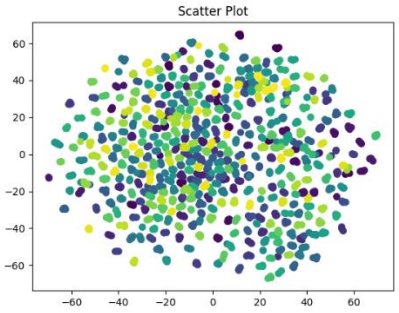
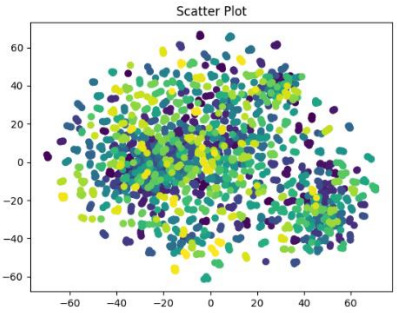
Table 2: Metrics for each model (% accuracy)

		ResNet-50	VGGFace
LFW_50	Accuracy:	99.66	99.04
	Precision:	99.92	99.70
	Recall:	99.82	98.80
	F0.5-Score:	99.90	99.51
LFW_20	Accuracy:	99.57	97.18
	Precision:	99.45	97.85
	Recall:	99.59	97.73
	F0.5-Score:	99.48	97.83
LFW_10	Accuracy:	98.97	80.51
	Precision:	98.24	86.74
	Recall:	97.77	86.69
	F0.5-Score:	98.14	86.73
LFW_5	Accuracy:	91.52	72.10
	Precision:	89.98	64.73
	Recall:	90.18	66.93
	F0.5-Score:	90.02	65.16

The results of the k-NN tests show that both the ResNet-50 and VGGFace models both perform well. Each architecture was able to achieve a 99.66% and 99.04% accuracy respectively with ample training data in the LFW_50 split. However, it is clear that the ResNet based architecture performs the best. It still achieves a 91.52% in the LFW_5 split compared to the VGG architecture which achieved a 72.10% accuracy. This achievement in accuracy of ResNet-50 is attributed to the depth of the network. The network contains many more filters than the VGG architecture which allows the network to learn more significant features about the input image.

Scatter plots allow us to see how the models are able to differentiate and group each embedding. The scatter plots for each dataset split are listed below (Table 3) for each model configuration.

Table 3: Scatter plots for each model

	ResNet-50	VGG-16
LFW_50:		
LFW_20:		
LFW_10:		
LFW_5		

The differences between the two models can be attributed to the quality of the facial embeddings. For the k-NN to differentiate the embeddings, they have to be similar enough to each other to not be mistaken as belonging to another embedding cluster.

ResNet is shown to generate tighter clusters; thus, the ResNet-50 architecture generates more representative embeddings than the VGG-16 based architecture. As seen in Table 3, there are some outliers in various clusters. These outliers are seen as a different colored dot in the middle or in the vicinity of a different colored cluster. These outliers are differences in the embeddings that the models create. The model produced an embedding that more closely resembles a different class. Thus, we see that misclassifications are a possibility with a k-NN comparison. However, algorithmic predictions should not be the only factor used to positively identify an individual as we will further discuss later.

With high accuracies, the representative embeddings ResNet-50 is able to produce allows us to utilize facial embeddings versus traditional CNN prediction in a surveillance system for policing agents. As the number of people the model needs to identify would be continuously changing, a traditional CNN prediction system would need to be retrained. By utilizing facial embedding comparison for prediction, the system can leverage precise predictions while maintaining the flexibility to modify the dataset under consideration. With this method, it would also be simple to supplement the ResNet-50 architecture with another state-of-the-art CNN architecture as models are constantly being developed. A substitution to the network could be the 1202-layer deep architecture that the authors of ResNet-50 experiment with or even a deeper architecture (He et al., 2015). The authors attributed the slight decrease in performance of the 1202-layer deep architecture to overfitting (He et al., 2015). But, for an architecture that could potentially be trained on everyone in the United States, or even world, deeper architectures like the 1202-layer ResNet architecture might prove useful. The final model used in a surveillance system can even be finetuned on the data it uses as a comparison over time to improve

performance to the dataset. The promising results of facial recognition with facial embeddings as shown allows us to create an architecture with and without the use of machine learning models that have to be retrained.

Despite the high accuracies achieved by the facial embedding analysis, the data from the LFW dataset has a few known problems. LFW was not created with the intention of vetting commercial facial recognition applications. Thus, LFW contains disproportionate representations of minority groups, and genders favoring white males. The benefit of LFW is the unconstrained facial recognition. It contains images of faces in the wild which are prone to occlusions and off axis views. Exhaustive, balanced datasets that include gender and race (Fitzpatrick skin type classification system) need to be utilized for training deployable models at a large scale. Ultimately, our experiment shows that a CNN can learn representative facial embeddings without having to be fine-tuned to a dataset when trained properly.

3.5 Surveillance System Architecture

Throughout Chapters 2 and 3, we discussed various methods in which facial recognition can be accomplished and have settled on the use of Convolutional Neural Networks for their state-of-the-art accuracy. The CNN based facial detector YuNet provides a very robust architecture in which we can extract faces from an image. Paired with the CNN based ResNet-50 model, a face can be extracted, and made into a facial embedding. Accurate results in facial detection and recognition allow us to create an effective surveillance system with facial recognition abilities. In this section, we will briefly describe an architecture (see Figure 13) in which facial recognition can be easily

performed from an image, or video feed.

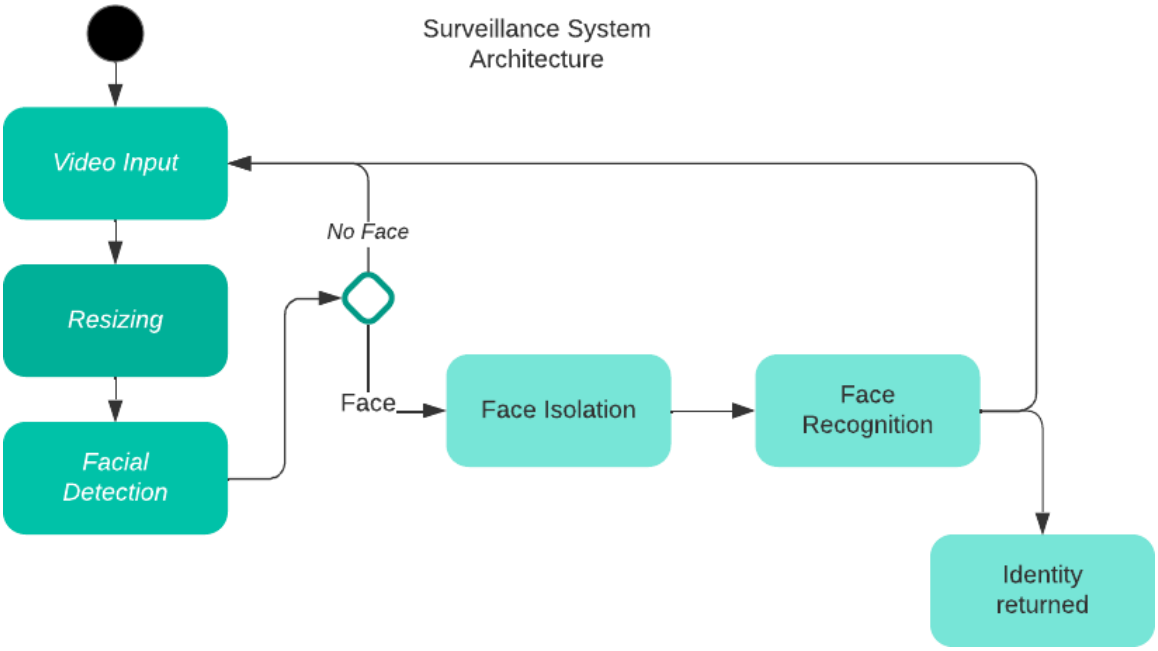


Figure 13: Simple surveillance system architecture diagram

The 1st step for any facial recognition application is the input. A video feed can be broken down into frames in which each frame is simply an image, so we will discuss the input as an image. First, an input image needs to be resized appropriate to that of the detector. For our purposes, we resized the image to 224 x 224 pixels as the size is a common choice for computer vision frameworks. This is largely due to efficiency and performance. A 224x224 image is large enough to capture the important details in an image, but not large enough to require the use of expensive resources. Though, more capable hardware might be able to take advantage of larger images with a higher resolution to preserve detail.

The 2nd step in our recognition problem is facial detection. We utilize the CNN detector YuNet to perform facial detection. This can be substituted for other architectures based on system requirements, but we utilize YuNet for its robust nature. The resized

image is passed into YuNet, and it outputs the bounding box for the face. With this bounding box, we crop the face out of the image to isolate it for recognition purposes.

For the 3rd step, we create the embedding of the isolated face. We utilize ResNet-50 for this purpose. Passing the image into our ResNet-50 model (as described in Section 4.4), the model generates, then outputs a facial embedding.

Our 4th step will be described with the help of a decision tree. Having the embedding of the individual under consideration and presumably a database of known embeddings, we have a couple of options as depicted in Figure 14.

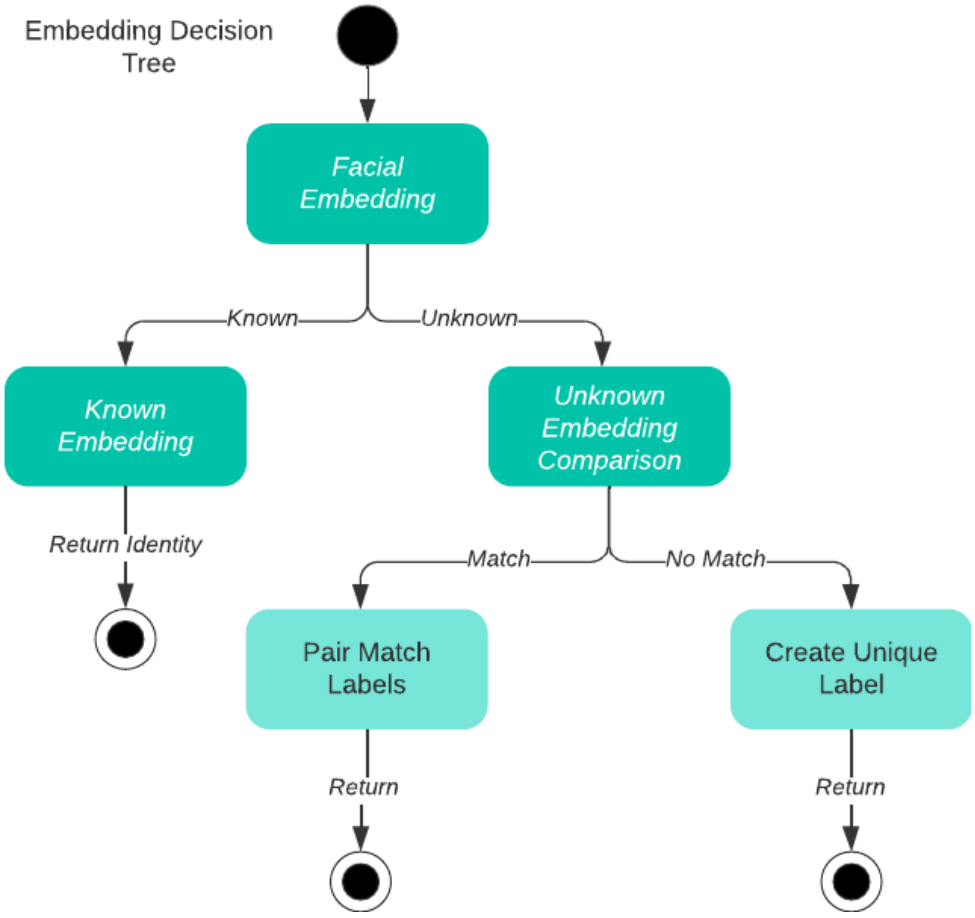


Figure 14: Decision Tree showing potential paths for a facial embedding

We first compare the embedding to known embeddings using their Euclidean distance. If there is a match (under a threshold value), then the identity is returned. This

threshold value can be reduced to create a tighter threshold to emphasize positive classifications over potential misclassifications. If there is no match, then compare it to the set of unknown identities. If the embedding matches an unknown identity, match their labels. If it does not match an identity, then store the embedding as a unique label for future reference.

To ensure proper classification of embeddings by the system, it is fairly simple to insert a human into the decision tree process of the prediction using a human-in-the-loop design (HITL). A HITL design is an architectural design principle coming from the study of human-computer interaction (HCI). A HITL is as intuitive as it sounds: a human-in-the-loop of a program. A human works to determine what should go where, and if the information that passes through is accurate (Wang, 2019). As we turn to surveillance systems that can be used to identify individuals through facial recognition, a human can work to assist the AI discriminate identities from within a video frame. A human can work to verify correct labels that the system predicts. A human can also work to ensure that training data is under the correct label before the model is trained.

For a human to work as a discriminator, they would need to see the cropped face before it is passed into ResNet-50 or another classification model. Then the human would compare the predicted label with images from that label's set. From there, a human can identify whether that image is a True Positive or a False Positive. In this instance, a human is working as a discriminator similar to the k-NN, but instead of using Euclidean Distance, the human would be using sight. For unknown labels that have found a match, again a human can serve as a discriminator. The individual can determine if the match is

in fact a match, or if it is simply a False Positive as they would be more common with less unknown identities.

The human-in-the-loop is used as a safeguard and audit of the classifications by the system. The human can work in place of a CNN to classify images, but manual classification with increasingly larger datasets would be too large of a task. The hope for CNNs is to attain a level of progressive improvement that results in 100% in their predictions. Even when that day comes, a human should still remain a part of the audit of a surveillance system to routinely ensure correct function. While a human can be used for routine audit by such a system, a human discriminator might be better utilized to initiate an investigation. With current uses of surveillance systems, an investigation or arrest might be launched when a facial recognition system identifies a criminal without the identity being manually confirmed or situation contextualized. A manual confirmation should be performed before action in any investigatory situation.

CHAPTER 4

AI IN LAW ENFORCEMENT

4.1 Algorithmic Biases

Algorithms are the basis of many computer systems. An algorithm is generally better at processing large amounts of data than a human. Mundane data movement or entry tasks do not take much cognitive effort, thus are prone to error because of a lack of focus. Luckily, we can leave algorithms to do mundane computational tasks to leave humans to do other work. Ideally, these algorithms could do any task asked perfectly given the right steps and input, but that is not always the case. Algorithms are prone to algorithmic bias, and data bias which produce harmful effects when not appropriately addressed.

Algorithmic bias is rooted in the intricacies of an algorithm. Algorithmic bias occurs when an algorithm favors one output over another. For example, if an algorithm is trying to predict gender based on an individual's age, income, and salary, the algorithm might have a bias and weighs higher incomes as 'male' more often than 'female'. Depending on how the algorithm was created, it could have a bias due to societal stereotypes, the weight it applies to each factor, or more often than not, a bias from the data itself.

For algorithms that learn the input data, data biases perpetuate harmful effects and cause incorrect outputs. Data bias introduces under-representation, over-representation, and reveals an inaccurate picture about the data being presented. Disproportionate

representation is the primary form of data bias. For example, imagine a basket of apples. Out of a basket of 100 apples, if 75 of the 100 are red and 25 of the 100 are green, what are the odds that a red apple is chosen? Probability states there is a 75% chance that a red apple is chosen and a 25% chance that a green apple is chosen. The prediction that a red apple is chosen is much higher than that of a green apple. The difference in probability is due to the underrepresentation of the green apple in the basket. Now consider a program that is trying to predict the color of an apple out of the basket. The algorithm could perform well by simply guessing red every time. This is called over-predicting and it is a problem when algorithms do not learn the data well. Instead they over predict the majority class (over-represented data). When trying to analyze the performance of a model or an algorithm, over-predicting can lead to a false sense of good performance. An algorithm can also be subject to bias through feedback loops. A feedback loop occurs when an algorithm uses biased data to make a prediction. In such a situation, the input bias is perpetuated through the algorithm and creates a biased prediction. Then, these new predictions are used as future additional data as input to the algorithm which creates a cycle of bias data being used as an input and produced as output. Bias in data and in algorithms can lead to harmful effects, especially when such information is used in the policing field.

4.2 History of Bias in Policing

Now, instead of the set within an apple basket, consider the set of total arrests. The United States has a history of social and economic oppression of African Americans which has led to an increased representation of African Americans in jails, and as persons

of interest by law enforcement (Mayson, 2019, La Vigine, Lowry, Markman, & Dwyer, 2011). The oppression of African Americans is rooted to the slave trade which peaked in the mid-17th century. The abolition of slavery occurred in the mid-18th century, and it was only in the 21st century that the Civil Rights Act of 1964 was passed (Thernstrom & Thernstrom, 2022). The Civil Rights Act of 1964 prohibited discrimination (prior discrimination was primarily by skin color) in public places, by employers, and by schools (Congress, 2015). Throughout these periods and still to this day, African Americans and non-Caucasians have been subject to heightened scrutiny by policing agents. Higher attention has led to a history of increased interaction with policing bodies, and increased arrests. Data generated from these arrests create disproportionate representations in predictive policing datasets of African Americans, in comparison to Caucasians which results in over-policing.

In recent years, the war on drugs has resulted in an increased racial bias and disproportionate scrutiny of minority groups. The drug war is a global effort aimed at reducing illegal drug trade and use. Even though African Americans are only 12.5% of all substance users in the United States, 30% of all drug-related arrests are from African American users (Pearl & Perez, 2018). For marijuana possession, African Americans are 32.7% more likely to be arrested than Caucasians (Rahmatulla, 2017). African Americans are also more likely to receive longer sentences than other ethnicities for drug violations (ACLU, 2013). These harsher punishments are not representative of equal punishment among drug users and reveal a discriminatory bias towards African Americans. It is possible that law enforcement concentrates on urban areas and lower-income communities and have personal biases that result in harsher punishment (Drug Policy

Alliance, n.d.). This approach to drug enforcement perpetuates the cycle of data being created that is bias towards minority groups.

Another example of biased data being created comes from the 1950's when the FBI developed the COINTELPRO (Counterintelligence Program) to disrupt the activities of the Communist Party (Federal Bureau of Investigation, n.d.). In the 1960's, COINTELPRO was expanded to include domestic groups and movements such as the Ku Klux Klan, the Black Panther Party, and the New Left (Federal Bureau of Investigation, n.d.). While this program ended in 1971, it was known for government bias towards targeted groups of individuals with progressive ideologies. Throughout the lifetime of the program, increased arrests and scrutiny of these groups again led to the perpetuation of biased data and biased sentiments towards minority groups.

Both the war on drugs and COINTELPRO ultimately targeted groups of individuals despite the intention of the program. Both initiatives were designed to be beneficial at the time, but their implementation eventually led to increased scrutiny and targeting of minority groups, particularly African Americans. The higher arrest rates of select individuals create more datapoints for certain ethnicities which are used as input data to the predictive systems we will address next.

4.3 Person-Based Predictive Policing

Predictive policing aims to predict crime in order to prevent it. Experimentation into prediction in criminal justice can be traced back to efforts of the Chicago School of Sociology back in the 1920s for criminal recidivism (Ferguson, 2017). Predictive policing efforts utilize data to predict who has a high potential of committing a crime, and

the locations of future crime. Before the advent of predictive policing, the theory behind policing was a reaction to crime. Policing forces would respond to reports of crime and manage the situation accordingly. The shift to prediction takes a forward-thinking approach in policing efforts through deterrence. The rise of predictive policing efforts extends past just the United States' COMPAS and PredPol Systems. Predictive systems have been in utilized Europe such as Crime Anticipation System, PreCobs, and Hunchlab (Hardyns, 2018). Japan has the Hitachi AI system which takes sets of biometrics and data analytics for crime prediction (Hung, 2021). Analysis by predictive algorithms gives policing agents data-backed predictions used for crime prevention. From the FBI (Federal Bureau of Investigation) and NSA (National Security Agency) to the DHS (Department of Homeland Security) as well as others, policing agents have used big data to drive predictive algorithms in order to reduce and prevent crime. Critics and scholars debate the suitability of existing laws and regulations that govern police activity in a digital world and consider decisions based on machine learning, and technological surveillance as civil rights violations that discriminate against minority groups (Brayne, 2020; Caplan, 2018). The two main methodologies for predictive policing in the United States involve:

i. Person-based: Who is more likely to commit a crime.

ii. Place-based: Where is a crime more likely to occur.

Person-based predictions aim at identifying individuals who are at a higher risk of committing a future crime. A product such as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a software program used for person-based policing. COMPAS is a criminal recidivism software program by Northpoint (now Equivant) that uses 137 factors to determine an individual's likelihood of committing a

future crime and assigns them a risk score (Angwin et al., 2016; Brayne 2020). The risk score is used by judges when determining parole, sentencing, and jail time (Brackey, 2015). One study in New Orleans, Louisiana on gun related violence illuminated the role of individuals in crime rates. It was found that out of 378,000 individuals, 1% (3,000 individuals) had the highest risk of being involved in a gun violence related incident (Ferguson, 2017). Police units were able to act on the tips of the person-based prediction and investigate. Because of the actions taken from the predictive assessment, the City of New Orleans murder rate fell 21.9% (Ferguson, 2017). Using predictive assessment such as COMPAS, higher risk individuals can be identified and sentenced appropriately.

Unfortunately, person-based policing strategies are subject to bias. As discussed previously, the United States has a history of racial bias. African Americans are incarcerated at six times the rate of Caucasians (Wachter-Boettcher, 2018; Brayne, 2020). Some studies show that recently African Americans and Caucasians are arrested at equal rates; however, COMPAS predicts that African Americans and Caucasians have very different recidivism rates (Mayson, 2019). The software predicted African American defendants to have a higher false positive rate (the false positive rate is the rate at which a prediction is true, and the condition is false) of 44.9% vs 23.5% for Caucasians (Angwin et al., 2016; Mayson, 2019). The false positive rate shows that African Americans were predicted to have a higher rate of rearrest but not rearrested. In other words, COMPAS labeled African Americans twice as likely to commit a future crime. On the other hand, COMPAS predicted a higher false negative rate (the false negative rate refers to the rate at which a prediction is false, and the condition is true) for Caucasians of 47.7% vs 28.0% (Angwin et al., 2016; Mayson, 2019). The false negative percentage shows that

Caucasians were rearrested at a higher rate than African Americans while predicted to not be rearrested. Equivant's response to the study was that the researchers grouped 'high' and 'medium' likelihood individuals both as 'high' which inflated false positive rates (Chawla, 2022). In the end, the software remains proprietary so it cannot be easily audited. Overall, COMPAS predicts with an accuracy of 65%; however, researchers at Dartmouth College attempted to predict recidivism with a group of random people. The participants predicted recidivism rate with an accuracy of 67% (Young, 2018). In their study, they also created an algorithm that just looked at an individual's age and number of prior convictions with a 67% accuracy (Young, 2018). In addition, they used various permutations of factors which did not achieve any better accuracy (Young, 2018). They concluded that there might be a ceiling on predicting criminal recidivism that technology cannot yet pass.

Person-based policing strategies also cause policing units to watch select individuals more closely. A list of 'chronic offenders' is used by many officers to conduct visits (Brayne, 2020). Those on the list are visited face-to-face by police detectives, social workers, or community leaders to give notice that they are being watched for potential criminal activity. Officers also use social network analysis where the relations between individuals are analyzed to reveal who has interactions with another individual (Brayne, 2020; Ferguson 2017). Interaction with an individual on the 'chronic offenders' list allows officers to put together connections and do visits with related individuals for their investigation. Visits with those on or related to high-risk individuals create a condition where people are subject to disproportionate attention by officers when compared to those not on the list. Someone on the list receives more police interaction

than an average citizen of a community and are prone to receive harsher punishment/longer jail time than those not on the list (Ferguson, 2017). If someone is on the list and is labeled as 'high risk' by COMPAS, a judge is likely to increase the severity of punishment.

Ultimately, person-based policing strategies create disproportionate attention for certain individuals. This attention creates a feedback loop where historically biased data is used to create a prediction, and minority individuals are subject to increased scrutiny; thus, an increased likelihood of getting arrested. These new arrests create more data points for minority individuals that are again used to create predictions. In the next section, we will talk about place-based predictive technology.

4.4 Place-Based Predictive Policing

The Los Angeles Police Department, with the help of federal agencies, leads law enforcement in the area of place-based predictive policing (Brayne, 2020). In 2011, the LAPD started using PredPol (Brayne, 2020). Developed in partnership with Jeff Brantingham from UCLA and George Mohler from Santa Clara University, PredPol became a staple in predictive policing software used by many police departments around the nation (Brayne, 2020). PredPol's algorithm for place-based predictions uses official crime data as parameters for analysis. The software analyzes location, type, and time of crime to predict hot spots with data within the past 10 years being weighted more heavily than older crime data (Brayne, 2020). The program places 500 square foot boxes that overlay maps of the areas that the officer is patrolling during their shift. The predictive boxes are used to encourage officers in the area to patrol the locations of higher crime

rate. Certain areas might have a higher potential for crime for a variety of reasons. Clubs introduce the aspect of alcohol, drugs, and can be tied to errors in judgement (Ferguson, 2017). Some streets and alleys might have gang affiliations or be a hot spot. This can be due to the decreased visibility and traffic. These prior knowledge factors and crime precursors have allowed data to be collected and analyzed for sophisticated place-based predictions.

Place-based predictive policing or hot spot policing has proven effective in various situations. Hot spot policing has been utilized to reduce crime in many instances related to robberies, shootings, and gang-related activities (Weisburd & Telep, 2014; Kleck & Barnes, 2014, Sherman, 1995). One of the first hot spot studies, Minneapolis in 1995, suggested the effectiveness of hot spot policing, and a subsequent study by Koper in 1995 revealed that the survival time (amount of time after an officer departed a hotspot before disorderly conduct occurred) of a hot spot was increased by 23% per minute in a hot spot with diminishing returns after 15 minutes (Weisburd & Telep, 2014). The study concluded that the ideal time for an officer to be in a location was 14-15 minutes. Criminologists have since developed pilot programs for place-based prediction practices in Boston, Baltimore, Kansas City, Las Vegas, and Los Angeles to test the effectiveness of a place-based strategy (Ferguson, 2017). In Boston, it was found that less than 5% of Boston's streets amounted for 74% of fatal and non-fatal shootings between 1980 and 2008 with 65 of the locations being hot spots (Ferguson, 2017). Departments in Boston were able to utilize this analysis and create a plan for deploying department resources. Effective deployment of officers resulted in a 17.3% reduction in violent crime (Ferguson, 2017). In Chicago, PredPol targeted gun violence through examination into

precursors of gun related crimes. The analysis found that the algorithm could predict 50% of gun homicides specifically citing an elevated risk of homicide 30 through 100 days after a handgun crime within half a mile of the predicted location (Ferguson, 2017). Another study in Sacramento, California assessed the effectiveness of predictive policing. The officers randomly rotated between hot spots and stayed for 15 minutes each which resulted in a significant reduction in crime (Weisburd & Telep, 2014). An alternative approach to patrols was revealed in 2008 where it was determined that hot spot policing can also be effective through the reduction of crime attractiveness (Braga & Bond, 2008). Reducing crime attractiveness can be achieved through demolishing abandoned buildings, cleaning graffiti, and increasing the risk for offenders (Braga & Bond, 2008). Traditional patrol, foot patrol, and reducing crime attractiveness have all been promising methods of hot spot policing. The respective effectiveness of each comes from increased police presence and risk of arrest. However, increased police presence as a deterrent is not flawless.

Hot spot policing can work well to deter crime but is disproportionately used in minority and poor communities. The data used to formulate the hot spots in policing efforts has an underlying bias for disadvantaged communities as more crime data is reported in those areas (Weisburd, 2014). While the data may suggest a higher crime rate in certain areas, the data might not always tell the entire truth. With more police allocated to hot spot locations, there is an increased chance for police interactions, scrutiny, and arrests. This data is then fed into future predictions which suggest departments allocate additional resources to the same locations. Predictive policing systems ultimately create a feedback loop which creates a higher perceived crime in certain areas. The

disproportionate scrutiny these areas face is similar to the effects of person-based predictive systems. A study in Shreveport, Louisiana in 2012 was conducted to determine the effectiveness of predictive policing efforts. Throughout the study, the program did not reveal any statistically significant reduction in property crime (Hunt et al., 2014). Another study was conducted to examine whether increasing the number of law enforcement personnel in hot spots would discourage criminal activity through increased risk of arrest. The study selected 54 large urban counties deemed representative of the 75 largest counties responsible for the most crime in the United States (Kleck & Barnes, 2014). In conclusion of the study, they determined that there is no relation to increasing manpower and the risk of arrest.

In conclusion, person-based and place-based predictive policing solutions have shown their effectiveness in select areas. However, predictive policing creates feedback loops that disproportionately punish minority groups due to feedback loops and the history bias in policing throughout the United States. Over time, there is a risk that those being policed will perceive heightened police presence or police interactions as being targeted by policing efforts rather than protected by policing agents (Weisburd, 2014). When not addressed, both have the potential to erode the relation between the community and policing agents which can cause distrust in authority. In the next chapter, we will discuss the use of surveillance technology as an alternative to predictive policing methods.

4.5 Surveillance in Law Enforcement

Predictive based policing systems, in theory, would be a great tool for law enforcement. However, they suffer from significant feedback loops and historical data biases that affect their real-world performance. As an alternative, this thesis proposes that law enforcement should look towards mass surveillance systems (as described in Chapters 2 and 3) paired with other surveillance technologies to aid in crime deterrence and criminal investigations. Surveillance technologies have been used in the past, but their implementation was eventually abused. The benefit of surveillance systems is they are not prone to as severe historical data bias, nor the same feedback loops as predictive systems. Throughout the remainder of this chapter, we will discuss the current uses of surveillance technology by law enforcement, as well as address their faults.

Surveillance systems work as a crime prevention measure through deterrence. Similar to how if an officer were standing in front of you, you would probably be deterred from stealing. The deterrence of a surveillance systems works in the same way - through the perceived certainty of punishment for wrongful actions (Alexandrie, 2017). Modern surveillance techniques rely on portable imagery with the evolution of cameras, phones, and other imagery devices.

One of the most notable uses of mass surveillance in the United States in recent history developed after 9/11. The Patriot Act of 2001 and Foreign Intelligence Surveillance Act (FISA) in 2008 was passed to authorize the surveillance of private communications as a method to catch terroristic threats (Toomey & Gorski, 2016; USA Patriot Act, 2001). The use of mass surveillance via imagery monitoring, and telecommunications was justified to prevent another terrorist incident. The vast

surveillance by the government and private entities was unknown to the public until the whistle blower case of Edward Snowden. Snowden was known for being a contractor for the National Security Agency (NSA) who leaked classified documents about U.S. surveillance programs of its citizens (Davies, 2020). Some of the documents outlined the U.S. government keeping records of private phone calls and harvesting big data from internet companies including Google, Facebook, and Microsoft (Davies, 2020). The previously undisclosed program called Prism was used to collect data ranging from emails and file transfers, to live chats (Davies, 2020). Prism, alongside other private data collection and surveillance techniques by the U.S. government has since been deemed unconstitutional by the USA Freedom Act in 2015 (USA FREEDOM Act, 2015). Private companies (i.e. Google, Amazon, Ring) have been scrutinized for selling or providing the government with metadata from their users creating many privacy concerns (Harwell, 2019; Manfredi, 2022). The use of mass private data collection for anti-terrorism and policing measures is something that was previously (FISA) seen as "okay" in society, but reform of those programs was rooted in the issue of privacy for citizens which we will discuss in detail throughout Chapter 5.

In the United States, the two most common forms of surveillance technologies seen come in the form of traffic cameras and CCTV (closed-circuit television) systems. The goal of traffic cameras and ALPRS (automatic license plate readers) is to deter traffic violations. These systems both scan plates and either send tickets or alert officers of an incident involving the vehicle. Alternatively, these same cameras can also be utilized for real time surveillance. ALPR coverage in Los Angeles along with automated traffic surveillance and control cameras (ATSAC), and tolls allow law enforcement to collect

this license plate information, but more importantly, data about where people are and when. This data is important in creating timelines of people's movements and activities. An LAPD officer explained that this data can be used as a digital footprint (Brayne, 2020). This digital footprint can tell a different story than one could assume. Using these data points, an individual's intentions and behaviors can be investigated based on video evidence of their actions (Brayne, 2020). This video evidence can be used to create a story and provide context for someone's actions, but, as discussed later, it is still up to the legal system, not the surveillance technology to prove whether surveillance footage is incriminating or exculpatory. In other words, there must be articulate evidence and a case built before the surveillance information can be used. The system cannot be the basis of a case.

Surveillance has been used by public and private entities in localized areas making use of CCTV (closed-circuit television) and has been found beneficial in the reduction of crime. It is common to have to go through security screening in airports by the TSA (Transportation Security Administration). Surveillance technology implemented by the TSA include airport cameras, x-ray body screening and x-ray luggage screening. These security measures are taken in effort to keep mediums of mass transportation safe and secure, while attempting to prevent another event like 9/11 in the United States. Scholars argue that the effectiveness of TSA security measures because of their failure during controlled tests (Cayford & Pieters, 2018). Despite failing some controlled security tests, the real-world effectiveness of the security measures come in the form of the public's knowledge of the various security methods (Cayford & Pieters, 2018). The public and potential criminals are aware of the security measures through the screenings

and cameras that surround airport terminals and checkpoints. With the presence of TSA personnel, there is shown to be a real potential for punishment of criminal activities. Effectively, the surveillance and security screenings serve as a deterrence for crime rather than a real preventative measure because of their lack of effectiveness (Cayford & Pieters, 2018). Despite the measures being surpassed by determined criminals, the effectiveness of the various security measures is rooted in the physical presence, and plausible threat of being caught from surveillance and security screenings.

There have been many other instances where the effectiveness of CCTV systems as a surveillance technique have been proven to be effective in localized deployments. Deployment of CCTV cameras have been shown to produce a 24-28% reduction in violent crime in public streets and urban subways (Alexandrie, 2017). There has also been shown to be a significant reduction of property crime with the use of CCTV cameras (Alexandrie, 2017). The effectiveness of the deployment of CCTV systems in these isolated places are likely due to the coverage of the cameras. A CCTV system created for a smaller area such as a house, or a subway has a greater chance of having no blind spots compared to a larger area. However, city streets are an ideal area for implementation of mass surveillance techniques. Camera placement is predictable, which provides full coverage as well as the potential for those cameras to be spotted. With cameras having full coverage and being in plain sight, it is easy for someone to see if they are potentially being watched. The perceived potential of being caught performing an illegal act (La Vigne et al., 2011). Despite public criticism, the benefit of CCTVs as a deterrence is enhanced through public awareness of such systems being deployed for public safety.

Surveillance efforts have been a critical piece in the effectiveness of operations performed by the NSA but are heavily criticized. After receiving criticism by the public after leaks by Snowden, the NSA was required to reduce their surveillance operation. Even with the reduction, surveillance techniques have been and are still criticized by scholars of being an invasion of privacy (La Vigne et al., 2011; Cayford & Pieters, 2018). The goal of the NSA is to gain an advantage for the U.S. through computer network operations (National Security Agency, n.d.). Because of Snowden, it is publicly known that the NSA utilizes telecommunication and public surveillance to investigate and deter national security issues. Criticism falls on the NSA for the perception of the agency 'watching everyone'. Through their operation, they ingest about 1.6% of world data, while reviewing .025% of that 1.6% which works out to about .00004% of the surveillance data the NSA actually reviews (Cayford & Pieters, 2018). The argument against public surveillance techniques by the NSA is that the NSA has 'eyes' on everyone; however, the claim is false considering the very small amount of data that the NSA has and even less that the agency reviews. With surveillance proving to deter crime, the notion that the NSA ingests information about everyone may be a good thing. If the thought of that exists, then it is likely that people will refrain from criminal activities.

The crime deterrent effect of surveillance systems work because of increased public scrutiny. Real-time scrutiny and post-act scrutiny promote, with the threat of punishment, good behavior. We have seen surveillance systems used throughout history in a harmful manner as described. During the period after World War II, tensions caused the Soviet Union to erect a divide between East and West Germany to prevent people from leaving Soviet Controlled East Germany known as the Berlin Wall. The Stasi

(Ministry for State Security) wiretapped, opened letters, and tracked citizens of East Germany (Bailey, 2019). Effectively, East Germany was a police state where the Stasi were engaged in monitoring all communication efforts of the inhabitants. The massive amounts of surveillance efforts even caused some people to self-censor for fear of information being used against them (Bailey, 2019). The mass surveillance efforts by the Stasi resulted in fear for the residents of East Germany and caused a massive "chilling effect" of its residents.

As surveillance technologies develop, imagine if a surveillance system is equipped with better action detection. There are many small acts that society primarily accepts as a 'norm' but are in fact illegal. For example, it is common that people jaywalk, verbal harassment, and underage drink. All of these actions are common actions that are illegal, but often not punished as they are accepted. Consider a surveillance system that is programed to enforce the law with exceptional action detection. Small towns and cities would be different. College towns would be different. Alcohol influenced public social gatherings would be different. In these three cases, these acts are usually only punished when they become "out of control" and disruptive. How would a surveillance system be able to draw that line? According to U.S. Law, it might have to be programmed to punish even small acts that are technically illegal. Hence, there is a need for trust in the system that exists to punish criminal actions, but also considers social "norms" that it oversees all without being invasive. To create such a system and considering the current state of action recognition technology, it would be wise for the surveillance system to not directly punish individuals and have a human overseer as described in Chapter 3 (HITL).

Throughout Chapter 5, we will visit other privacy concerns of surveillance systems. We will also examine the necessary measures that need to be in place in order to operate a mass surveillance system that does not fall into the traps of abuse previous implementations have suffered.

CHAPTER 5

SURVEILLANCE AND PRIVACY

5.1 Privacy Issues of Mass Surveillance

As discussed in Section 4.4, surveillance systems can have a positive impact on policing efforts and provide a method for social accountability. Throughout Chapter 5, we will discuss leading arguments about privacy and provide examples where a mass surveillance system might infringe upon perceived privacy rights. We will end with a discussion of standards for a mass surveillance system implementation that account for privacy concerns.

In Section 4.4, we concluded surveillance systems have numerous beneficial applications. It can be used for the benefit of catching criminals or piecing together an investigation, so why would we be against its use? The answer is a bit complex and is rooted in privacy. Privacy is an elusive term to define, so we will work to paint a picture of privacy through various views. Privacy concerns the right to keep certain aspects about property, interests, and oneself private from others. We will first focus on Judith Thomson's account of privacy in which Thomson describes privacy through a simplifying hypothesis - the idea that rights exist over oneself, and property that an individual owns. Next, we will look at Thomas Nagel's definition of privacy which focuses primarily on the idea of voluntarily concealment of information from or revealing information to others. Lastly, Mark Tunick elaborates on a definition of privacy that includes informational privacy. As we will further discuss, informational privacy

overlaps the idea from Thomson definition through information about oneself, or information about something that an individual owns. It also overlaps Nagel's belief of voluntary exposure which also applies to information. All these accounts are crucial to why we should not ignore issues of privacy when implementing a surveillance system. They also show that privacy delineations intersect various different areas to protect the ways we make decisions on informational privacy and autonomy. A system as such that is implemented to enforce the law should be made in due regard to the privacy concerns that arise from it as there are many instances where we may be doing something that, while not illegal, may be considered a violation of privacy. These violations of privacy have the potential to expand into a case of wrongful punishment. This ushers a need to have a balance between privacy and protection.

5.2 Right to Privacy

One view on privacy states mass surveillance violates our right to privacy because of violations to an individual's cluster of rights. Judith Thomson illustrates that our right to privacy is not a distinct right, but part of a cluster of rights: protected property rights, right over oneself, and the right to not be subject to harm (Thomson, 1975). These rights are described as a "simplifying hypothesis" that describe rights over oneself, and property (Thomson, 1975). Throughout the discussion, Thomson illustrates an X-ray device that allows us to see an image someone has in a safe within their house. The individual that owns the image has the right to who views it and what is done with the image (Thomson, 1975). So, someone even with a legally acquired X-ray device that can view it would be violating the owner's privacy interest over the image. Mass surveillance can be paralleled

with the effects of X-ray devices. Someone with a camera or audio system pointed at an individual's house could pick up something via audio or video that the individual is trying to keep private and according to Thomson, has the right to keep private. A surveillance system on a mass scale would have a more impactful effect. A mass surveillance system has the potential to accumulate information through many different angles that might pick up an image from the inside of a private environment such as a home window. While we can debate whether someone acquiring information or imagery through an open window is within their legal right to do so, it is, according to Thomson's view, a violation of their moral right to privacy.

While privacy protects property, it also protects the body. Thomson claims privacy also dictates we have the right over oneself (Thomson, 1975). Our body is something that we have control over; thus, comes our ability to choose what happens to it (Thomson, 1975). Whether it be through sight or physical touch, it is almost undoubtedly our decision what happens to us (Thomson, 1975). When in public, most of us waive our right against being looked upon. Most would not consider it an issue of their privacy if they were seen in public as it is expected that they would be seen by others. On the street, we are in the public view and can be subjected to photography or gazing. While a long gaze might bring a feeling of unease, it is not wrong to simply look. However, a Muslim woman, in a culture where wearing a hijab is predominant, might want to stay covered. With an interest in keeping her skin private, a camera that picked up the woman's face while she was adjusting her hijab could be considered a violation of her privacy. This case can be escalated if the image is distributed on the internet or revealed to individuals without her consent. In this instance, the woman's privacy interests are influenced by her

culture. However, this issue is not just cultural and extends to individuals with an intent to keep themselves covered. Clothing that is worn to conceal parts of the body may be bypassed using surveillance systems. Cameras may unintentionally capture angles of private areas that people might not want to be seen. This might not be a violation of privacy, similar to an unintentional view. However, an effective surveillance system will always record and potentially capture imagery of private places and be disseminated to others. This unintentional consequence is a real privacy concern, especially with a mass-scale surveillance system. We will discuss a resolution to this issue through data privacy and anonymity in Section 5.5.

Thomson also illustrates the right to not be subjected to harm. Thomson highlights that violence or even a threat of violence should not be done onto another individual (Thomson, 1975). This belief is articulated by U.S. laws forbidding assault, battery, and murder (Crimes and Criminal Procedure, 2021). While these types of physical acts are not directly caused by surveillance, leaked imagery can lead to moral harm and wrongdoing. For example, a leaked surveillance video of someone cheating on their wife might provoke violence by either party. While being upset about the situation might be reasonable in the case of an affair, an act of violence in retaliation to almost any circumstance is not the appropriate response. However, similar to the case of the Muslim woman, the viewing of surveillance videos by unwanted individuals crosses the boundaries of privacy. The example of the affair shows that sometimes, violations of privacy can lead to further harm and wrong doings. Such an effect is precarious when implementing a mass surveillance system.

The rights that protect property, oneself, and to not be subject to harm, as previously mentioned, can all be violated by a surveillance system. Recently, Amazon-owned security company Ring has admitted to giving footage to law enforcement officers (Harwell, 2019; Manfredi, 2022). Ring is a well-known security company for its creation of doorbell security cameras that capture video and audio. Rather than using local storage, Ring's devices store footage in Amazon's extensive cloud network. Now, think about what kind of imagery that device picks up. Equipped with motion activation, the doorbell records when someone comes to your house and can help identify the individual. It also picks up audio from any conversation and as discussed before, this data can be used to timeline entry, exit and motion. While such features are great for a home security system, consider the many units Ring has sold and the number of homes on which they have been installed. Amazon is known for doing targeted data advertising and data selling to other companies (Harwell, 2019; Manfredi, 2022). With data from an individual's home, there is almost no doubt some sort of analysis is performed. Amazon has admitted to providing this data to law enforcement agents which causes an effect similar to if an officer is posted outside of your house (Harwell, 2019; Manfredi, 2022). For many, it could be seen as beneficial knowing that law enforcement is watching them and keeping them safe. But, without knowing the extent of the data that is being delivered to law enforcement nor how this data is used by private data sellers and companies, there must be safeguards implemented when it comes to surveillance capable devices. These varying examples involve the capture and distribution of imagery or audio of an individual or their property. We will continue to discuss the misapplication of captured imagery throughout this chapter.

5.3 Concealment and Exposure

As mentioned previously, mass surveillance poses the need for a discussion about concealment and exposure as illustrated by Thomas Nagel. Nagel asserts a fundamental ability for privacy and transparency in our interactions with others, hence concealment and exposure (Nagel, 1998). This idea extends to oneself and information about that person. We have the ability to conceal parts of our lives, including our social relations. For example, people may want to keep the intimate and personal parts of their lives private. Concealment and willing exposure to various social circles greatly impact how people behave. The decision to expose certain parts of one's life to other individuals is a decision that, according to Nagel, should be kept by the individual (Nagel, 1998). The aspect of concealment allows us to foster different relations with different people. We can be professional with our coworkers and foster a professional workplace relationship with them. At home, we can have intimate relationships with our partner that do not involve work. It is only at our discretion that we might reveal our intimate relationships with others. Voluntary concealment gives us the opportunity to choose to whom we expose certain circles of our life. They may overlap, but they may stay discretely distinct. We might have a reason to keep many social relations private for legitimate reasons, with or without immoral intent. Whether those relations include their work, personal life, and family life, there is a need for the ability to regulate social circles.

Mass surveillance techniques have the potential to expose aspects of personal life to the public sphere and could do so without the regulation of the individual. For instance, when a candidate is running for political office, there is almost always media, sponsored by another party, that displays distasteful clips or quotes of the candidate in an

attempt to dissuade the public opinion. In some cases, the media uses information such as the status of intimate relationships to influence the vote (Grimaldi, 2012). If it is discovered that the political candidate is having an affair, the candidate's campaign can fall apart because of the revealed information, leading to public distrust. The intimate portions of political figure's life have not previously been of interest during the political race, rather the candidate's qualifications and platform have been more closely examined (Nagel, 1998). Though recently, the intimate portions of a candidate's life have been on display with the help of mass communication platforms and the media. Intimate matters in a candidate's life usually have little impact to their qualifications for a position, despite the media's pronounced interest in revealing all intimate matters of an individual's life (Nagel, 1998). It might be distasteful to have an affair, but is it the public media's role to disclose distasteful parts of individual's lives that have no affect to their ability to perform their job? Surveillance techniques have the capability to pick up these distasteful portions of lives and spread them for mass viewing. The affair we talked about with Thomson would also be hindered with Nagel's view. Information about this affair could again lead to harm from the spouse, or judgment and punishment by society. Information spread to others to whom the individual does not intend. In doing so, surveillance systems transgress the boundary that delineates the public and private spheres that an individual might intend to keep.

As mentioned in the previous section, a Muslim woman with the interest of keeping her body covered might not want her body exposed to others without a specific interest in doing so. The interest of keeping a right to bodily autonomy is highlighted by Thomson. Nagel intersects Thomson's view of bodily autonomy as a right through

willing concealment and exposure of the body. A public space surveillance system that unintentionally captures portions of the body that an individual keeps covered would be in violation of any individual's interest. If in the same moment, a nearby onlooker inadvertently catches a glimpse of the same portion of the individual's body, would the onlooker also be in violation of the individual's privacy? It can be argued yes, but that onlooker had no intention to view that portion of the individual's body. What would be a violation of the woman's privacy is if after, the individual disclosed information that the woman has some abnormality on that portion of their body. Surveillance systems in an 'always on' operation have the potential to capture things that a simple onlooker might miss or give the feeling of always being gazed upon. With improper data access protections, this information has the potential to be spread. Going back to the Ring camera (in Section 5.2), consider the potential for a camera to capture parts of an individual they were not thinking about covering because they were in their home. Then, this data is sent to the police for investigation of a crime in the neighborhood. Cameras that are 'always on' can pick up unintended imagery. Exposure of this unintentionally given information of that person can be seen as a violation of privacy by both Thomson's and Nagel's views on the matter. In the case of a surveillance system, it might capture portions of person's bodies or actions and store them in a database. This database has the potential to be exposed to unwanted individuals. We will revisit this type of behavior and how surveillance systems can be built with regulated access of personally identifiable information in Section 5.5.

5.4 Privacy and Punishment

Section 5.2 introduced the potential for violence rooted in the leaking of a surveillance video. In this section, we will talk about how violence is not the only form of punishment possible by the implementation of mass surveillance. As a consequence of private facts entering the public sphere, there is a potential for harm and undue punishment. Tunick highlights the potential criticism by society from video leaks (Tunick, 2013). Surveillance systems, to be effective as outlined in Section 4.4, have the potential to make information about an individual's past wrongdoings or behavior easily accessible (Tunick, 2013). Videos that collect information about people have the potential for leaks. When such a video is leaked, it can cause the person to be judged and potentially punished by society. Such a judgement can happen soon after such an event is revealed, or it can be carried with the person for the rest of their life. Ex-convicts are the most notable to carry such a weight of judgement and mistreatment by society. When applying for jobs, many that have a background check, whether formal or informal, ask if the individual has been convicted of a previous crime. The person then is left to lie about their past doings in hopes of receiving equal treatment if not exposed or left to tell the truth and face potential judgement by the employer. In the case of an ex-convict, many prisons have rehabilitation programs (Harding, 2014). While rehabilitation programs are aimed at deterring offenders from future crime by social change, the ex-convict still must face the judgement of the employer as an ex-convict rather than someone who did not commit a crime. Paralleling the example of the political candidate, someone who has the correct qualifications might be overlooked because of something revealed to the public space about their past (i.e. an open relationship). In these cases, their "punishment" is

through the sense of a barrier created by their past actions that inhibits them from contributing to society. In some European countries, public disclosure of criminal records is considered degrading, and access is restricted (Tunick, 2013). After serving their sentence, Europe's policy dictates that the ex-convict has served their punishment for their crimes. Access to criminal records are not public after release which allows ex-convicts to have a chance to not be labeled as a convict and contribute to society after their sentence. Without the policy, the individual would be subject to unjust punishment by society. While this policy is beneficial for individuals seeking employment, other individuals might have an interest in knowing the past of an individual. Moving into a new neighborhood, the neighbors might have a legitimate interest in their child's safety if the new neighbor has been convicted of child molestation. In this case, disclosing such a criminal offense could be of the benefit of the current residents of the neighborhood. Though, the neighbors might subject the individual to the same judgement. After hearing about the past of the individual, there could be a tension in the neighborhood out of fear for the safety of their children or the neighbors might spread the word in the area to make sure others are aware. In both cases, the reaction of the neighbors might be reasonable, though they did not respect the isolated disclosure of the information. These acts of social punishment are due to the undesired dissemination of information. As discussed in the next section, all of the examples of privacy violations are rooted in the actions of people towards one another after being exposed to surveillance footage.

5.5 Privacy Standards for Surveillance Systems

As discussed in Sections 5.2-5.4, there are various reasons that surveillance systems and photography can violate an interest in privacy. The cases discussed are rooted in the manner that people respond to footage they have viewed. It is clear that people have a right to privacy that implementors of surveillance systems need to consider. So, throughout this section we will discuss the safeguards and considerations that have to be taken in order to ethically implement a mass surveillance system. These methods include regulated access, routine audit, and increased transparency.

There is a difference between private and public information or facts. Public facts are facts which one cannot reasonably expect privacy (Tunick, 2013). Most of what is known about an individual by the public is for reasons of our own doing. This could be through government forms, social media posts, academic articles, etc. There are also facts about us that are released to the public without our consent such as your name, and date of birth. These facts are recorded before we have cognizant control over ourselves during our infancy, but such facts are usually not considered something that we expect to keep private. On the other hand, a private fact a private fact is typically regarded as intimate details about someone's life that they have not disclosed to the public (Nagel, 1998). For example, an individual's medical record is something that is usually kept private. Additionally, some facts are private and could be relevant to the public. The proprietary algorithm of a business might be of interest to the public especially when the algorithm engages in legal matters.

As stated before, there are many reasons medical records in particular should be kept private, but technologies such as surveillance have the potential to expose private

information through prolonged observation. Privacy of medical records helps promote transparency between practitioners and patients (Gostin, Nass, & Levit, 2009). In some instances, technology enables doctors to quickly examine and share information with other experts in the field to provide assistance quickly and efficiently. Though, technology also allows one to video you entering, for example, a dentist's office. By elongated examination through a surveillance system, one could deduce that a political official visits the same oncologist every week and could deduce that the individual is suffering from a form of cancer. While they might not have video of you inside of the office, they could conclude that the official has a deteriorating health condition. With the spread of this information, there could be a societal backlash trying to replace the official as they are not seen fit for office. So, while medical practices are kept private, you are almost always in the public eye and under enough scrutiny, habits can be revealed. When reviewed over long periods, surveillance systems have the ability to capture habits that can be exposed to people the individual did not intend.

The habits surveillance system reveal can lead to social punishment as seen with the political official, but in some cases can lead to legal punishment. In a 1973 decision on Roe v. Wade, the US Supreme Court abolished virtually all abortion restrictions imposed on the state and local levels in the United States (Liben, 1995). The Roe v. Wade ruling became precedent allowing access to abortions and abortion related medicine, treatment, and contraceptives. In the summer of 2022, the Supreme Court overruled the almost 50-year precedent determining that the United States Constitution did not imply the right to an abortion (Library of Congress, 2022). With this ruling, the regulations surrounding abortion was put to the states. Some of which had "trigger laws"

which took effect immediately after the ruling. These "trigger laws" effectively made abortion illegal in the state in which they occurred. Despite those with "trigger laws", abortion remains legal in most states across the U.S. Even before the most recent ruling, someone trying to get an abortion in an area where abortion was frowned upon would have to face scrutiny from protestors standing outside of abortion clinics. A person attempting to get the procedure could have a privacy concern if they are trying to get this procedure when their close peers believe against the act. Mass surveillance systems could reveal someone routinely visiting a clinic for meetings. Perhaps these meetings take place in a state where the action is illegal, but they reside in a state where the act is legal. A mass surveillance system can pick up on the act that the person is trying to commit an act that is illegal in their locality. Upon return to their place of residence, information of this act could have been spread to the other residents or policing official in the locality. When returning to their place of residence, the individual could face social battery and potentially legal punishment as a result of prolonged observation.

Ultimately, cameras that are always watching can pick up actions an individual is trying to keep. While, in the latter example of the person traveling back into their home state, this thesis a stance on whether that action should be punishable by the state, I am claiming that someone, in either case, an individual might have a legitimate interest in keeping their actions private. They might want it to be private because of conflicting interests by their peers, their society, or many other reasons. As we discussed in Section 5.3, there is a privacy interest in informational boundaries that we can establish or allow lenience at our discretion. Surveillance systems have the ability to spread information about us without our intention. These actions of the individual might not be able to stay

private but should remain undisclosed and unpunished if not illegal. Throughout our previous discussion, most concerns of privacy have been due to the concern of what happens to a video when it is disseminated and judged. We will next discuss a general use case of a surveillance system.

Let us now focus on public surveillance through a morning cup of coffee. As you arrive to work, you stop at the local coffee shop to grab a cup of coffee. Someone walking by filming in the area might catch you in frame and post the video online. You could be upset that they did not ask for your consent to post the video, but were they within their rights to post a video with you coincidentally in it? In most instances, yes. There is not much wrong inherently with a video having another individual in frame. Posting the video without the intent to slander/publicly shame individual is okay. Though if the individual consequentially captured you in the video and it showed evidence of an affair, the individual has the potential of receiving public backlash and undue punishment by society or the individual's partner. Through all of our previous examples we have established one cannot expect privacy to the full extent as in private quarters when in the public eye. However, one can also expect to be free from unwanted attention and harsh treatment (Tunick, 2013). So, for a mass surveillance system that is always on, it might be easy to be against such a system. But let us consider a situation where a surveillance system would benefit the community.

Consider the same example, you are walking on the street grabbing a morning cup of coffee. On your way to your building of work, you are pushed into the alley at gunpoint and demanded to hand over your phone, wallet, and keys. You comply in fear of personal harm, and after you hand over your valuables, the assailant hits you with the

base of the pistol which renders you unconscious as the assailant escapes back into the busy streets. When you regain consciousness, you go to work to talk with your security office which relays the incident to local law enforcement. You are then called into the local police station for a detailed account of the incident, but do not remember the assailant's face enough for a drawing. A couple of weeks after looking into the incident, the department follows up with you and tells you there is nothing they can act on to continue their investigation. No DNA, no witnesses, and CCTV footage came up empty. After learning this information, you must go on with your daily life, but wish there were something that could have been done. Maybe, if the department was better funded, there would be more police presence in the area, or more cameras to deter or witness the crime. Unfortunately, crime in the area remains because of monetary distribution to the local departments by the city. As time passes, you eventually move on, but a thought stays in the back of your head. What if there was more surveillance in the area? The police could have caught a glimpse of the assailant and brought them to justice, but the former situation could have also played out. The coffee shop camera could have even caught you in a presumed affair and for various intents the footage could have been leaked. If the police were to have access to such footage, it could be uncomfortable knowing that someone is looking at footage of you in an act you might feel ashamed about. Though, the footage could be used to bring an assailant to justice. The video can have extraneous sections removed if they are not pertinent to the case. If all parts of the footage must be presented, then the content should be kept confidential by members of the trial. The concern for privacy after exposure to such a clip is contingent on how people respond to viewing such a video. When used for the benefit of catching criminals, those who were

exposed through a trial would probably think nothing more of the footage. It is when the video is spread and view negatively that people can receive punishment from a video.

We have discussed a surveillance system that has the ability to capture actions at a large scale that one might have thought to be private. As the effectiveness of surveillance systems come from the knowledge of surveillance and plausible ramifications for wrong doings, a surveillance system would have to be implemented with countless cameras working together in a system. These cameras have the ability to pick you up getting your morning cup of coffee, or intimacy with a partner in public. It could be seen as unjustified for a surveillance system to always be watching as the consequence is capturing unintended actions. However, when used as a crime deterrence, the system has value. Though, the value of deterrence comes with the need for regulation to minimize privacy infringements. In society, the value of deterrence also comes with the sense of protection rather than targeted punishment.

Surveillance systems that are always watching could bring many a feeling of discomfort. You might be persuaded to act in a manner that is appropriate to your surroundings rather than a natural environment. The mass use of public surveillance cameras has the potential to create a "chilling effect" on public life (American Civil Liberties Union, 2022). Individuals aware that they are being surveyed might put more thought into how they look, dress, and might not try to draw attention to themselves in the way they act. One might change their way of life to not seem a robber, or a terrorist that could alarm an observer or an authority figure. Being watched by an authority figure such as a surveillance system or a systematically placed officer could both bring this effect. As mentioned in Chapter 4, there exists the problem of authority and governance

in society. There is the line between taking direction of authority and policing figures, and balancing the feeling of being sought out to be punished rather than protected by the policing bodies. The need for a balance between authority and fair treatment to keep the population willing to follow the rules and accept the policing actions as protection. This gives us the need to limit a built surveillance system to not actively punish individuals. ALPRs in traffic lights primarily serve as positive punishment for traffic violations. They have a negative effect towards the individual that is in violation. ALPRs cite a ticket, a court date, and directions to resolve the ticket. While it leaves the option to argue the ticket in traffic court, the evidence clearly shows a traffic violation through an image of the car, the traffic violation. However, what if the car had to move into an intersection to avoid a police car or ambulance and did not catch the emergency vehicle in the image? It is normally correct to move to let an emergency vehicle pass through an intersection.

Now consider a mass surveillance system works in a manner of positive punishment. A system such as surveillance has a much broader impact than a simple ALPR. A surveillance system can catch what an ALPR does and more. For example, imagine a mass surveillance system that catches an individual shooting someone and is programmed to cite a ticket, and a court date. In the case of apparent murder, it might even trigger measures to bring the person to jail for sentencing before a trial date. A system as such might not have caught the whole situation where the shooting was actually an act of self-defense. Or it could have caught the full interaction, and punished the individual after incorrectly determining it was an act of self-defense rather than unprovoked violence. With the advancements of AI systems has come the advancement of action. Some are able to correctly identify position, whether or not an individual is

walking or jogging. Consider a group of people that are wearing dark hoodies and standing in a circle on the corner of the street. They might be interpreted as selling/consuming drugs, scheming an act of malicious intent. On the other hand, an officer that observed the situation might be intrigued and investigate the meeting rather than instantly citing a ticket. Now imagine an AI system that can interpret a scene as effectively as a human and is used to alert authorities as soon as an illegal act occurs. Individuals might be dissuaded from going into public spaces to prevent exposure to such a system. An ALPR system, if built to cite tickets and put into motion the needs for instant arrest without considering all the factors involving the situation, could be detrimental to the relationship between society and law enforcement. A surveillance system that seeks to punish society would surely tip the line of protection and punishment in favor of punishment and not be received well by society. Individuals would fear that even the smallest legal infractions will give them a court date, a ticket in the mail, or an instantaneous police response. So, an AI powered surveillance system should serve as a tool for police departments and federal entities to act in a manner appropriate within the law to deter crime instead of a policing figure. Being such a powerful system, the primary focus of surveillance systems should be piecing together criminal investigations and as a crime deterrent. Though, there needs to be a balance in focus between major organized crime, national security issues, and petty crime. This balance will keep surveillance serving as a crime deterrent to maintain social order, while preventing people from being improperly punished. Automation to such a system will only result in unjustified human punishment. Human involvement in the system provides a level of 'humanity' to the system that technology has not yet reached.

The majority of the concerns we have addressed with surveillance systems have to do with how they are built, and how they are used. Many uses of the system we have found problems which we can call situations of minor abuse. Minor abuse is use of the system by an agent of the state for nefarious purposes (Taylor, 2005). For such abuses, there are fairly simple safeguards that can be implemented to address these abuses. Before we address the safeguards that are required for an ethical surveillance system to be utilized, we need to first address a concern of major abuse. Major abuse can be recalled through our discussion in Section 4.5 where a corrupt state is the root of the abuse caused by the system (Taylor, 2005). In such cases, we should address the legitimacy of the state rather than the methods used (Taylor, 2005). In a corrupt state, there are many mediums in which the state can abuse its citizens and surveillance efforts are only one. This thesis only defends that an uncorrupt state can ethically implement a mass surveillance system. For our purposes, we can assume an uncorrupt state is one that only uses surveillance technology when it is morally justifiable. While it is important to keep technologies (such as mass surveillance) that have a mass effect on a population regulated and out of the hands of corrupt states, it is given that the use of any such technology by a corrupt governing body is indefensible. So, for the remainder of our discussion, we will establish safeguards to reduce the potential for major abuse by the state and minor abuse by agents of the state, all of which must be in place for implementing a mass surveillance system. To ensure a surveillance system is fundamentally useful, the system must be implemented with safeguards considering access, transparency, and audit to ensure privacy concerns and minor abuses are minimized.

Surveillance systems need strictly regulated access. To regulate access, the footage collected in a surveillance system must first be stored and protected. Though, we first need to establish who is accessing such a system. A state-owned mass surveillance system can collect mass amounts of data. Keeping bulk data can help build a timeline of individuals and become useful in the future (Brayne, 2020). While this thesis concerns state-owned and operated surveillance systems, privately owned and operated surveillance systems must follow the same ethical guidelines and be subject to the same audits. As discussed earlier, a timeline of someone's day can be deduced simply through human observation. The data from years of observation can be analyzed and can help create timelines which may serve useful in the future. The drawback of large data collection and storage is that it would require massive data centers to keep, store, and process all this data. With the advancements in quantum computing, the processing of the video in the near future would be minimal compared to the amount of data that would need to be stored (Mengoni, Incudini, & Di Pierro, 2021). Those who need access to data from the system should only be the officers pursuing an investigation, lawyers for case evidence, and auditors (engineers) to maintain the system. Regulated and recorded access to the system keeps users accountable for their actions. Credentials for each user can be tracked through a system of logs which show who access what information and when. This logged data can be useful for routine audits to determine proper use and access of the system. It is a given that these systems be accessed with legitimate purpose which protects privacy (Macnish, 2014; Taylor, 2005). It can be established that the system is accessible only after a warrant is provided. Which in such a case, it is typically regarded as justifiable that the information requested through the warrant is viewed in the case of a

legal investigation. Surveillance data, when requested by warrant, falls under the same scrutiny of the law of any other pertinent case information.

The access can be considered as a harm to an individual's privacy but should not. When a surveillance system is operated, strictly regulated access with justifiable means should be the foundation for a system that considers privacy. When considering privacy as control of information, mass surveillance systems prevent individual control of the information collected. And as discussed earlier, unregulated information can cause a multitude of subsequent harms. However, when considering privacy as a when someone accesses information, a surveillance system only violates that access when the data is accessed with a legitimate purpose. Also, as stated before, such information should not be the foundation of a case. In an any legal system, video of an action might not always tell the whole story (Tunick, 2013). As discussed with the examples of the emergency vehicle and the apparent murder, actions in a video only reveal a glimpse into the context of what happened before the situation and shows almost no light into the reasonings behind an action. While some believe that a surveillance system can replace traditional witness and jury relations in court, this thesis defends that the conviction of someone should not rely solely on a video surveillance system (Taylor, 2005). When surveillance information is only accessed and used in legal proceedings, it protects an individual's privacy (open access and information leaks do not). Any legal investigation could contain many more potential privacy infringements but are a necessity to prove innocence or guilt. So, when surveillance information is regulated to strictly legitimate uses (legal proceedings), a surveillance system can be an aid to criminal investigations.

Trust in the legal system relies on the sense of a fair punishment. Innocent until proven guilty should be the standard in which it operates, and the same logic should be applied to AI driven systems. Much of our trust in the institutions that run our country are due to policies of transparency and accountability (Meijer, 2014). Trust is enabled through transparency. Without transparency, it is difficult to hold entities accountable and establish trust. As seen on many government websites such as the FBI, CIA, NSA, they contain sections on their purpose, and methods of operation (Central Intelligence Agency, n.d.; Federal Bureau of Investigation, n.d.; National Security Agency, n.d.; National Security Agency, n.d.). While their respective methods are not completely explained, they are stated and can be researched to a limited extent. Government operation is almost always under scrutiny, so transparency is a need for trust and so the public can hold government bodies accountable. A mass surveillance system operating under a government entity can work to the protection of the public while being almost completely transparent in operation. While surveillance systems that utilize facial recognition are not immune to false positive predictions, surveillance systems based on CNNs can achieve near perfect accuracy on facial predictions. As discussed in Chapter 3, a human-in-the-loop architecture can help mediate false positives by required human confirmation when accepting a prediction by the system. Additionally, transparency into the system's usage, access, and deficiencies can be a means to enhance public trust. The strictly regulated access of a system can easily be reviewed during an audit. With the presumption that the data is only accessed by legitimate persons in the context of a case, the information accessed, as well as the pertinence to the case can be reviewed to evaluate proper usage.

While the system and raw data should stay private for protection of individuals, the architecture and algorithms can remain public knowledge so they can be rigor tested by private and public audits. Conferences (i.e. ACCV, CVPR, ECCV, ICCV, IEEE, ILSVRC) in image recognition tasks already have their participants work to test and suggest optimizations for the algorithms and architectures used in computer vision applications. The conferences above have proven to show state-of-the-art results in image detection and recognition problems. Such conferences can even use anonymized data from the surveillance system to fine tune the models and increase accuracy when they are deployed. Computer vision conferences are also known for creating an environment for critical analysis by the research community. Research that is performed throughout these conferences have a high degree of understandability, hence, the inner workings of such systems are transparent. In the artificial intelligence umbrella, there is a subfield called Explainable AI (XAI). XAI focuses solely on creating machine learning techniques that produce more explainable models that are human interpretable to increase levels of trust and understanding. The most common techniques for understanding a CNN include plotting kernels, activation maps, and higher feature activation maps. This sort of analysis allows researchers to see what the CNN is using as a 'feature' and can show if the CNN is learning the correct parts of an image. The power of audit and required regulated access and transparent function give some of the power back to society that the seemingly one-sided surveillance program causes. The ability for society, engineers and ethicists, to audit the system for bias, its usage, and abuse forces the state to operate in a justifiable manner. The routine audits can ensure the system's algorithms are fair and that the program is used for legitimate purposes.

Throughout our discussion, we have established that surveillance can violate interests of privacy. Though, we have established the necessary functions, and preventions that need to be in place for such a system to be justifiably used. Surveillance holds individuals accountable. Many of us have heard the phrase "act as if someone is always watching" in reference to always behaving in a proper manner. In our case, someone might not be always watching, but something is watching. The system would work to deter crime in a similar manner to that of heightened police presence (Weisburd, 2014). As discussed throughout Section 4.4, surveillance systems have been proven to have a positive effect in crime deterrence (Gómez, 2021; Weisburd 2014; La Vigne et al., 2011). The surveillance system holds society accountable for an individual's actions by the possibility for review and legal punishment. The legal punishment behind the system is through the rules of the society that the society established through a representative democracy as seen in the United States. The check for these rules enforced in the U.S. are already established through the 3 branches of the government. We have established that a mass surveillance system can be built with due regard to privacy considerations with the following stipulations:

1. Regulated access
2. Increased transparency
3. Routine audit

A surveillance system as discussed, can influence a society to sustain social order. The privacy issues created by mass surveillance are greatly minimized socially when people treat each other in good faith and are mindful of privacy interests. These issues are also minimized when the state uses the system with justifiable intent.

Ultimately, isolated bits of information should not be spread about individuals lives, and increased scrutiny needs to be justifiable and have the correct intent. While the spread of information is protected under the 1st amendment in the U.S., but people need have more regard to privacy and respect for another not to violate interests in privacy (The United States Government, 2021). When any implementation of a mass surveillance is made, it must consider the ethical concerns discussed throughout this thesis. If so, a mass surveillance system presents an opportunity for social accountability and reducing criminal activities.

CHAPTER 6

Conclusion

6.1 Efficacy of Mass Surveillance Systems

Throughout this thesis we have discussed the efficacy of surveillance systems for policing agents and determined they are effective. While surveillance systems do pose some privacy infringements, those infringements are not to be of concern when the system is regulated with the concerns addressed in this thesis. In Chapter 2, we explored the history of convolutional neural networks and their ability to perform facial recognition tasks. Chapter 3 demonstrated that state-of-the-art CNNs can achieve highly accurate results when applied to facial detection and facial recognition. Throughout Chapter 4, we learned that data used in predictive policing systems is highly prone to biases due to disproportionate representation. This results in higher scrutiny for individuals and communities by policing agents. However, we also determined that video surveillance systems are a better solution than other predictive policing systems. Throughout Chapter 4, we discussed the potential for privacy violations by surveillance systems. In Chapter 5, we determined that state implementation of a mass surveillance system must have safeguards in place to minimize privacy violations and the potential for abuse. This can be accomplished through regulated access, transparency, and routine audits. Access can be regulated in a mass surveillance system through detailed access logs where each user has a unique access key and their navigation through the system is logged for review. Transparency in a surveillance system as described in Section 3.5 can

be achieved by routine audit and disclosure of the use of the system. High levels of transparency about the proper uses and the performance of the system can lead to increased trust that the system is used for the protection rather than punishment of society. The findings of this thesis show that once technical and ethical concerns are addressed, a mass surveillance system can be implemented effectively by law enforcement.

The findings of this thesis provide important insight into the use of mass surveillance systems by policing agents. Potential biases in predictive policing systems highlight the need for privacy protections and can be used as a starting point for future research into reducing bias in surveillance systems. Future research might benefit from focusing on improving accuracy of facial recognition by use of 3-d models. Humans are great at identifying people through past outfits, articles of clothing, and off-axis identification because we are able to analyze the full body of an individual. Recognition based on factors additional to facial embeddings could prove to be the future of recognition system.

Ultimately, surveillance systems provide an opportunity for social accountability, but it is essential that they are implemented thoughtfully and with proper safeguards. With increased cooperation of individuals and communities, policing agents can be better equipped to serve society while making it safer and more enjoyable for everyone. It is my hope that the insights provided by this thesis will contribute to the ongoing conversation about the use and role of surveillance systems in law enforcement and lead to more ethical and effective policing strategies.

References

ACLU. (2013). *The War on Marijuana in Black and White*. American Civil Liberties Union. https://www.aclu.org/report/report-war-marijuana-black-and-white

Alexandrie, G. (2017). Surveillance cameras and crime: a review of randomized and natural experiments. *Journal of Scandinavian Studies in Criminology and Crime Prevention,* 210-222. doi:10.1080/14043858.2017.1387410

American Civil Liberties Union. (2022). *What's wrong with public video surveillance*. Americal Civil Liberties Union. https://www.aclu.org/other/whats-wrong-public-video-surveillance

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Asian Conference on Computer Vision. (2022). ACCV. https://www.accv2022.org/en/

Brackey, A. (2015). *Analysis of Racial Bias in Northpointe's COMPAS Algorithm* [Master's Thesis, Tulane University]. Tulane University Libraries. https://digitallibrary.tulane.edu/islandora/object/tulane%3A92018/datastream/PDF/view

Braga, A., & Bond, J. (2018). *Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial*. Wiley Online Library. https://doi.org/10.1111/j.1745-9125.2008.00124.x

Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review*, 977-108. https://doi.org/10.1177/0003122417725865

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* New York: Oxford University Press.

Cao, Q., Shen, L., Xie, W., Parkhi, O., & Zisserman, A.(2018). VGGFace2: A Dataset for Recognizing Faces across Pose and Age. In 13th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 67-74). Xi'an, China: IEEE.

Caplan, R. (2018). *Content or Context Moderation?* New York: Data & Society. https://datasociety.net/wpcontent/uploads/2018/11/DS_Content_or_Context_Moderation.pdf

Cayford, M., & Pieters, W. (2018). *The effectiveness of surveillance technology: What intelligence officials are saying.* doi:10.1080/01972243.2017.1414721

Central Intelligence Agency. (n.d.). *About.* Central Intelligence Agency. https://www.cia.gov/about/

Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird B., Grusky, D., & Leskovec, J. (2021). Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. *Nature*, 82-87. https://doi.org/10.1038/s41586-020-2923-3

Bailey, C. (2019). *The Lingering Trauma of Stasi Surveillance*. The Atlantic. https://www.theatlantic.com/international/archive/2019/11/lingering-trauma-east-german-police-state/601669/

Chawla, M. (2022, February 23). *Compas case study: Investigating algorithmic fairness of Predictive Policing*. Medium. https://mallika-chawla.medium.com/compas-

case-study-investigating-algorithmic-fairness-of-predictive-policing-339fe6e5dd72

Chollet et al. (2015). *Keras*. GitHub. https://github.com/fchollet/keras

Computer Vision and Pattern Recognition Conference. (2023). CVPR. https://cvpr2023.thecvf.com/

Ciresan, Dan & Meier, Ueli & Masci, Jonathan & Gambardella, Luca Maria & Schmidhuber, Jürgen. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification. International Joint Conference on Artificial Intelligence IJCAI-2011. 1237-1242. 10.5591/978-1-57735-516-8/IJCAI11-210.

Crimes and Criminal Procedure, 18 U.S.C §1111 (2021). https://www.govinfo.gov/app/details/USCODE-2021-title18/USCODE-2021-title18-partI-chap51-sec1111

Davies, D. (2020, May 20). *Journalist Who Helped Break Snowden's Story Reflects On His High-Stakes Reporting*. National Public Radio. https://npr.org/2020/05/20/859376407/journalist-who-helped-break-snowdens-story-reflects-on-his-high-stakes-reporting

Dean, J., Corrado, G.S., Monga, R., Chen, K., Devin, M., Le, Q.V., Mao, M.Z., Ranzato, M., Senior, A.W., Tucker, P.A., Yang, K., & Ng, A. (2012). Large Scale Distributed Deep Networks. *NIPS*.

Drug Policy Alliance. (n.d.). *Race and the Drug War*. Drug Policy. https://drugpolicy.org/issues/race-and-drug-war

Gong, S., Liu, X., & Jain, Anil. (2020). Mitigating Face Recognition Bias via Group Adaptive Classifier.

European Conference on Computer Vision. (2022). ECCV. https://eccv2022.ecva.net/

Federal Bureau of Investigation. (n.d.). *Mission & Priorities*. Federal Bureau of Investigation. https://www.fbi.gov/about/mission

Feng, Y., Yu, S., Peng, H., Li, Y.-R., & Zhang, J. (2022). Detect faces efficiently: A survey and evaluations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1), 1–18. https://doi.org/10.1109/TBIOM.2021.3120412

Ferguson, A. G. (2017). Policing Predictive Policing. *Washington University School of Law*.

Gómez, S. M. (2021). The Deterrent Effect of Surveillance Cameras on Crime. *Journal of Policy Analysis and Management*, 5530571.

Gostin, L., Nass, S., & Levit, L. (2009). Beyond the HIPPA Privacy Rule. *National Library of Medicine*. https://www.ncbi.nlm.nih.gov/books/NBK9579/

Grimaldi, J. V. (2012). *Marianne Gringrich, Newt's ex-wife, says he wanted an 'open marriage'*. The Washington Post. https://www.washingtonpost.com/politics/marianne-gingrich-newts-ex-wife-says-he-wanted-open-marriage/2012/01/19/gIQAJzgwAQ_story.html

Gupta, V. (2022, December 20). *Face detection – dlib, opencv, and Deep Learning (C++ / python)*. LearnOpenCV. Retrieved February 23, 2023, from https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/

Harding, R. (2014). Rehabilitation and prison social climate: Do 'What Works' rehabilitation programs work better in prisons that have a positive social climate?

Australian & New Zeland Journal of Criminology, 47(2), 163-175. https://doi.org/10.1177/0004865813518543

Hardyns, W., & Rummens, A. (2018). Predictive Policing as a New Tool for Law Enforcement? Recent Developments and Challenges. *European Journal on Criminal Policy and Research*. 24. doi:10.1007/s10610-017-9361-2

Harwell, D. (2019). *Doorbell-camera firm Ring has partnered with 400 police forces, extending surveillance concerns*. The Washington Post. https://www.washingtonpost.com/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach/

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep residual learning for image recognition*. arXiv. https://doi.org/10.48550/arXiv.1512.03385

Hilbert, M. (2012). How much information is there in the "information society"? *Oxford University Press*, 9: 8-12. https://doi.org/10.1111/j.1740-9713.2012.00584.x

King, D. E. (2015). *Max-margin object detection*. arXiv. https://doi.org/10.48550/arXiv.1502.00046

Huang, G., Mattar, M., Lee, H., & Learned-miller, E. (2012). Learning to Align from Scratch. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/d81f9c1be2e08964bf9f24b15f0e4900-Paper.pdf

Hung, T.-W., & Yen, C.-P. (2021). On the person-based predictive policing of AI. *Ethics and Information Technology*, 23. doi:10.1007/s10676-020-09539-x

Hunt, P., Saunders, J., & Hollywood, J. (2014). Evaluation of the Shreveport Predictive Policing Experiment. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR531.html.

International Conference on Computer Vision. (2023). ICCV. https://iccv2023.thecvf.com/

Kleck, G., & Barnes, J. C. (2014). Do More Police Lead to More Crime Deterrence? *Crime & Delinquency*, 60(5), 716–738. https://doi.org/10.1177/0011128710382263

Krizhevsky, A., Sutskever, I., & Hinton G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*. 25. doi:10.1145/3065386.

La Vigne, N., Lowry, S., Markman, J., & Dwyer, A. (2011). Evaluating the Use of Public Surveillance Cameras for Crime Control and Prevention - A Summary. *The Urban Institute*. https://www.urban.org/sites/default/files/publication/27556/412403-evaluating-the-use-of-public-surveillance-cameras-for-crime-control-and-prevention_1.pdf

Langton, L. (2007). Aviation Units in Large Law Enforcement Agenceis. *Bureau of Justic Statistics*.

Le, K. (2019). *A Study of Face Embedding in Face Recognition*. [Master's Thesis, California Polytechnic State University]. Digital Commons at California Polytechnic.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *IEEEE*, 86. 2278-2324. doi:10.1109/5.726791

Liben, P. (1995). What the U.S. Constitution says. The law and abortion. *Freedom Review*, 26(5), 20-21.

Library of Congress. (2022). S*upreme Court Rules No Constitutional Right to Abortion in Dobbs v. Jackson Women's Health Organization*. Congressional Research Service. https://crsreports.congress.gov/product/pdf/LSB/LSB10768

Long, Y. Z. (2017). Face Recognition with A Small Occluded Training Set Using Spatial and Statistical Pooling. *Information Sciences*, 430-431. doi:10.1016/j.ins.2017.10.042

Macnish, Kevin (2014). Just Surveillance? Towards a Normative Theory of Surveillance. *Surveillance and Society* 12 (1):142-153. https://philarchive.org/rec/MACJST-2

Manfredi, L. (2022). *Amazon's ring dinged for handing over footage to law enforcement*. New York Post. https://nypost.com/2022/07/16/amazons-ring-dinged-for-handing-over-footage-to-law-enforcement/

Mayson, S. G. (2019). Bias In, Bias Out. *The Yale Law Journal*, 128(8), 2218-2300. http://www.jstor.org/stable/45098041

McCulloch, W. S., & Pitts, W. H. (1990). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 52, 99-155.

Mehta, N., & Shukla, S. (2022). Pandemic Analytics: How Countries are Leveraging Big Data Analytics and Artificial Intelligence to Fight COVID-19? *SN Computer Science*, 3(1), 54.

Meijer, A. (2014). Transparency. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford Handbook of Public Accountability.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641253.013.0043

Mengoni, R., Incudini, M., & Di Pierro, A. (2021). *Facial expression recognition on a quantum computer*. Quantum Machine Intelligence. https://link.springer.com/article/10.1007/s42484-020-00035-5

Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press.

Nagel, T. (1998). Concealment and Exposure. *Philosophy & Public Affairs*, 27(1), 3-30.

National Security Agency. (n.d.). *About.* National Security Agency. https://www.nsa.gov/about/

Nelson, P. (2022). *OpenCV face detection: Cascade Classifier vs. YuNet*. OpenCV. https://opencv.org/opencv-face-detection-cascade-classifier-vs-yunet/

Pappas, S. (2016, March 18). *How Big is the Internet, Really?* Live Science. https://www.livescience.com/54094-how-big-is-the-internet.html

Pearl, B., & Perez, M. (2018). *Ending the War on Drugs*. Center for American Progress. https://www.americanprogress.org/article/ending-war-drugs-numbers/

Qiu, H., Gong, D., Li, Z., Liu, W., & Tao, D. (2021). End2end occluded face recognition by masking corrupted features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. https://doi.org/10.1109/TPAMI.2021.3098962

Rahmatulla, A. (2017). *The War on Drugs has Failed. What's Next?* Ford Foundation. https://www.fordfoundation.org/news-and-stories/stories/posts/the-war-on-drugs-has-failed-what-s-next/

Rahmad, C., Asmara, R. A., Putra, D. R. H., Dharma, I., Darmono, H., & Muhiqqin, I. (2020). Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection. *IOP Conference Series: Materials*

Science and Engineering, *732*(1), 012038. https://doi.org/10.1088/1757-899X/732/1/012038

Rachels, J. (1975). Why Privacy is Important. *Philosophy & Public Affairs*, *4*(4), 323–333. http://www.jstor.org/stable/2265077

Rosebrock, A. (2021, April 17). *Face detection with dlib (Hog and CNN)*. PyImageSearch. Retrieved February 23, 2023, from https://pyimagesearch.com/2021/04/19/face-detection-with-dlib-hog-and-cnn/

Rosenblatt, F. (1957). The Perceptron: A Perceiving and Recogniting Automaton. *Cornell Aeronautical Laboratory*.

Rosenbrock, A. (2021). *Face detection with dlib (HOG and CNN)*. PyImageSearch. https://pyimagesearch.com/2021/04/19/face-detection-with-dlib-hog-and-cnn/

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Zhiheng, H., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 211-252.

Scheuerman, W. E. (2014). Whistleblowing as Civil Disobedience. *Philosopy & Social Criticism*, 40(7), 609-628. https://doi.org/10.1177/0191453714537263

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). *Overfeat: Integrated recognition, localization and detection using convolutional networks*. Paper presented at 2nd International Conference on Learning Representations, ICLR 2014, Banff, Canada.

Sherman, Lawrence. (1995). Hot Spots of Crime and Criminal Careers of Places. Crime and Place. 4.

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv. https://doi.org/10.48550/arXiv.1409.1556

Taylor, J. S. (2005). In Praise of Big Brother: Why We Should Learn to Stop Worrying and Love Government Surveillance. *Public Affairs Quarterly*, 19(3), 227–246. http://www.jstor.org/stable/40441413

The United States Government. (2021). *The Constitution*. The White House. https://www.whitehouse.gov/about-the-white-house/our-government/the-constitution/

Thernstrom, A., & Thernstrom, S. (2022). Black progress: How far we've come, and how far we have to go. *Brookings*. https://www.brookings.edu/articles/black-progress-how-far-weve-come-and-how-far-we-have-to-go/

Thomson, J. J. (1975). The RIght to Privacy. *Philosophy & Public Affairs*, 295-314.

Toomey, P. C., & Gorski, A. (2016). *Unprecedented and unlawful: The NSA's "upstream" surveillance.* Just Security. https://www.justsecurity.org/33044/unprecedented-unlawful-nsas-upstream-surveillance/

Tunick, M. (2013). Privacy and Punishment. *Social Theory and Practice*, 643-688.

USA FREEDOM Act of 2015, H.R. 2048, 114th Cong. (2015). Congress. https://www.congress.gov/bill/114th-congress/house-bill/2048

USA PATRIOT Act of 2001, H.R. 3262, 107th Cong. (2001). Congress. https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf

Wachter-Boettcher, S. (2018). *Technically Wrong.* New York: W. W. Norton & Company.

Wang, G. (2019). *Humans in the Loop: The Design of Interactive AI Systems.* Human-Centered Artificial Intelligence Stanford University. https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems.

Weisburd, D., & Telep, C. (2014). Hot Spots Policing. *Journal of Contemporary Criminal Justice*. doi:10.1177/1043986214525083

Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* [Doctoral Dissertation, Harvard University]. National Science Foundation. https://www.researchgate.net/publication/279233597_Beyond_Regression_New_Tools_for_Prediction_and_Analysis_in_the_Behavioral_Science_Thesis_Ph_D_Appl_Math_Harvard_University

Young, E. (2018). *A Popular Algorithm is no Better at Predicting Crimes than Random People*. The Atlantic. https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

Zeiler, M.D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In: Fleet, D., Pajdla, T., Schiele, B., & Tuytelaars, T. (Eds.) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53