Investigation Into Facial Recognition Technology: Biases, Privacy, and Ethics

Emma Deckers, Fall 2021, MATH 380

Abstract: With advancements to technology and machine learning, facial recognition softwares have become a ubiquitous aspect of our daily lives. From your iphone, social media accounts, and surveillance cameras, facial recognition algorithms are using your face to learn and identify you in private and governmental databases. In this paper, we will discuss the mathematical process behind the development of a facial recognition algorithm and introduce the areas where biases may occur. We will also encounter questions regarding the ethics behind data-driven methods and their collection of photos. While computers are inherently unbiased, through human error, implicit and algorithmic biases may develop and produce algorithms that perform better on light skin tones and perform worse on darker skin tones. When used by governmental agencies, these biased algorithms may target specific groups by having greater potential to provide false positive identifications for which generally marginalized groups of people, African American, Native American, and women, are used as justification for persecution. We will further investigate how biased algorithms are affecting the daily lives of Americans and offer potential solutions.

Introduction

The development of machine learning technologies, especially facial recognition technology (FRT), has been in the special interest of engineers since the 1960's. With so many useful applications, FRT's are being used across the globe to identify individuals in databases containing hundreds of million images in a matter of seconds. This technology is frequently used by governmental agencies and private companies to identify criminals caught on security cameras. However, many of the facial recognition algorithms used by law enforcement agencies have the potential for implicit and algorithmic bias which, in turn, could disproportionately affect some groups in the United States more than others. In order to understand how biases arise, we will need to understand how facial recognition algorithms are developed.

In machine learning, there are two kinds of learning: supervised and unsupervised learning. With supervised learning, an algorithm is provided a base of information and told by the developers how to interpret the data. With enough examples, the algorithm is able to learn how to read new data and make predictions. Facial recognition is a branch of supervised machine learning where an algorithm learns to classify individuals and identify them as familiar or unfamiliar based on the information provided. Training a facial recognition algorithm relies heavily on the methods of linear analysis. To train an algorithm, a team of developers will need to gain access or create their own set of training images, usually including thousands of pictures of different faces. The most accurate algorithms will be the algorithms trained with a greater set of training images composed of hundreds of people of different ethnicities, age, and gender. We will outline the steps necessary to train a system (Turk).

Initialization Operations:

1. Collect a training set of face images taken from different angles with different schemes of lighting on various subjects. The base images are known as eigenpictures

2. Extract the relevant information in an eigenpicture by calculating the eigenfaces, only keeping the M images that correspond to the highest eigenvalues. These M images define the "face space"

3. Calculate the corresponding distribution in M-dimensional weight space for each eigenpicture by projecting each face onto the face space

Recognizing New Facial Images:

1. Calculate a set of weights based on the input image and M eigenfaces by projecting the input image onto each of the eigenfaces

2. Determine if the image is a face by checking to see if the image is similar to others in the face space

3. If the image is a face, classify the weight pattern as either known person or a new face

Calculating Eigenfaces

Let an eigenpicture (face image) I be a two-dimensional N by N array of 8-bit intensity values. An eigenface is a low-dimensional representation of an eigenpicture. In order to calculate an eigenface, one must use principal component analysis (PCA), a common linear algebra technique used to reduce dimensionality of large data sets. The images of faces for a training set will be similar in composition to one another, so reducing their dimensionality using PCA will preserve data in a lower dimensional space.

Let $x_1, x_2, x_3, ..., x_M$ represent M of the initial eigenpictures with the highest eigenvalues. The average face in the set is defined by $\bar{X} = \frac{1}{M}\sum_{i=1}^{M} x_i$, where each face differs from the average by $\Phi_i = x_i - \bar{X}_i$. To reduce the eigenpictures to a lesser dimensional space, we seek a set of M orthonormal vectors, u_m, which provide a basis to best maximize the variance of each projection. We chose M images with the highest eigenvalues because they best maintain

their variance when reduced. A single vector, u_k, in the set of orthonormal vectors is chosen such that

$$\lambda_k = \frac{1}{M}\sum_{i=1}^{M}(u_k^T\Phi_i)^2$$

is a maximum in the set of all λ_i. The vectors u_k and the scalars λ_k are the eigenvectors and eigenvalues of the covariance matrix

$$\begin{aligned} C &= \frac{1}{M}\sum_{i=1}^{M}\Phi_i\Phi_i^T \\ &= AA^T \end{aligned}$$

where the columns of the matrix A are represented by the mean-normalized vectors $A = [\Phi_1\Phi_2...\Phi_M]$. However, the matrix C, is of dimension N^2 by N^2, which for typical image sizes, is too large to efficiently perform operations on. To form a new matrix L, of smaller dimension M by M, we construct L such that $L = A^TA$, where $L_{mi} = \Phi_m^T\Phi_i$, and find the M eigenvectors, v_l, of L. These vectors determine linear combinations of the M training eigenpictures to form the eigenfaces u_l where

$$u_l = \sum_{k=1}^{M} v_{lk}\Phi_k,$$

for $l = 1, ..., M$. With this new dimensionality, time spent on calculations is greatly reduced. The new eigenvalues of L allow us to categorize and rank the associated eigenvectors in terms of usefulness in determining variation among the images. The highest ranking eigenvectors and associated eigenfaces form the basis for the future use of categorizing, recognizing, and describing face images. This basis of only the highest quality eigenfaces is called the face space, and each of the eigenfaces in the face space is called a face class. The face space is used to compare images that are not in the training set to determine if they are familiar faces, a new person, or something else.

Using The Face Space to Classify New Images

Now that we have formed a basis for describing faces, we will want to use them to compare

new images to the face space. The new images might be pictures of the same people used in the training set taken from a different angle or in different lighting, or they may be new faces that we want to add to our collection of face classes. Take a new face image, z. We compare this image to the others by projecting the new face onto the face space using the equation

$$\omega_k = u_k^T(z - \bar{X})$$

for $k = 1, 2, ..., M$. This describes a set of operations in which to compare the new image to the face space and to break down the new image into its own eigenvalues, eigenvectors, and associated eigenface. The calculated weights create a vector $\Omega = [\omega_1, \omega_2, ..., \omega_M]$ that describes the contribution of each eigenface of the face space in representation of the new face image. This vector may then be used in a pattern recognition algorithm to determine which, if any, of the pre-identified faces (the eigenfaces in the face space) best describes the new face. The simplest way to determine which of the face classes best compares to the new face is to find which of the k eigenfaces minimizes the Euclidean distance $e_k = min(\Omega - \Omega_k)^2$ between the unknown photo and the face space.

There are four possible cases for an input image and its pattern vector: (1) near a face space and near a face class, (2) near the face space but not near a face class, (3) distant from the face space but near a face class, (4) distant from the face space and not near a face class. In the first case, an individual is identified and recognized. In the second case, the photo is identified as an unknown person. Cases three and four indicate a picture of something that is not a face, but may potentially be identified as false positives.

Problems In Training Facial Recognition Algorithms

Now that we have a better understanding as to how facial recognition algorithms are programmed, we may look into the possible complications of creating such an algorithm. Facial recognition algorithms are created via data-driven methods. This implies that they rely

solely on the information initially provided to them in order to learn and make educated guesses. Similar to humans, facial recognition algorithms can only identify faces based on the information available; they cannot produce results from information that doesn't exist. Therefore, algorithmic biases may arise in the training stage when they are provided training datasets with a lack of diversity. Humans as a species are inherently very diverse; their faces reflect aspects of age, gender, race, ethnicity, culture, and many other qualities that make an individual unique. Thus, it is imperative that identification algorithms are given a wide array of different faces during the training stage in order to produce the most accurate results (Merler et. al).

During the programming stage, another type of bias, implicit bias, may also arise. Implicit bias in the programming stage could occur if a team of programmers have a negative opinion of certain groups of people and choose to involve their personal biases in the development of the algorithm. They may purposefully choose to exclude photos of those groups or only include very few of those people in the training set in comparison to the number of examples of other faces. While we would hope that the programmers would leave their biases aside and include an appropriate number of photos of every ethnicity, gender, etc. in their training set, many of the facial recognition algorithms used today are commercially produced with little to no regulation, so it is impossible to monitor the development stage of every software company. Implicit bias may also occur when companies purposefully choose not to test their algorithms for biases before selling the technology.

Databases Available: Commercial Databases

Most of the facial recognition softwares widely used today are produced commercially for public and private use. Due to the Federal Acquisition Streamlining Act of 1994, which "strongly" stated preference for buying commercial items rather than purchasing government-unique items, the United States government tends to rely on already-made commercialized facial recognition softwares instead of procuring softwares created exclusively for govern-

ment use (United States). Current federal law enforcement agencies, like the DEA, U.S. Customs, and other sections of the Department of Justice for example, use commercially developed facial recognition softwares to identify criminals by pictures of their faces caught on surveillance cameras (Fleischer). While the technology is, for the most part, very helpful in criminal investigations, when left unregulated and untested for bias, some groups may find themselves at a higher risk of being falsely identified by a biased algorithm.

FRT Accuracy On Women And People Of Color

Despite having the funds to access millions of diverse photos, numerous commercially-produced facial recognition softwares have been exposed by researchers for containing significant racial biases. A report conducted in 2011 that researched the effectiveness of facial recognition softwares in Western and East Asian countries discovered a positive relation between race of the developers of the algorithms and accuracy in identifying people of the same race (Phillips et. al). The report found that a combination of eight separate Western algorithms were much more effective in identifying Caucasian people than identifying East Asian people, and the East Asian algorithms were more accurate in identifying Asian faces than Caucasian faces. Although this report was conducted some time ago, it still exposes a concerning aspect of machine engineering: the accuracy of your algorithm on certain groups of people may be dependent on who programs the algorithm.This aspect is particularly concerning when we consider the fact that most of the software engineers in the United States are males, specifically Caucasian males, which is why testing these algorithms before selling them is vital in the realm of facial recognition. We will look further into some newer studies regarding accuracy of FRT.

In 2019, the National Institute for Standards and Technology (NIST) conducted research into the effects of race, age, and sex on facial recognition softwares. The study researched 189 (mostly commercial) algorithms from 99 developers from across the world and researched effectiveness in one-to-one matching, confirming the identity of one person in the database

by comparing the image to another image of the same person, and one-to-many matching, using one image of a person and finding another picture of the same person in the database (Grother et al.). The database contained a total of 18.27 million images of 8.49 million people taken from databases provided by several governmental agencies. The study found that for one-to-one matching, there were higher rates of false positives for Asian and African American faces relative to images of Caucasians, differing from factors of 10 to 100 times depending on the algorithm. It was also found that among U.S. developed algorithms, there were high rates of false positives in one-to-one matching for Asians, African Americans, and native groups with the American Indian demographic having the highest rates of false positives. On the contrary, in the Chinese developed algorithms, false positives for East Asian faces were low, and false positives for Caucasian faces were much higher. This verifies that the information found in the 2011 study is still relevant. For one-to-many matching, African American women were falsely identified more than any other demographic group. A similar investigation by the MIT Media Lab also verified that commercial facial recognition algorithms performed up to 20% better on identifying Caucasian male faces, were up to 19% more accurate identifying lighter skinned faces, and performed up to 35% less accurately on African American female faces (Buolamwini).

In conclusion, all of the information provided by these three resources tells us that racial and gender biases are still very prominent in not only American facial recognition algorithms, but also in algorithms developed in other countries. Many of the algorithms examined above are the same algorithms that are being sold to governmental agencies for law enforcement. With the current state of facial recognition technology, historically marginalized groups will continue to face higher rates of arrest and incarceration due to higher rates of false positives among darker, non-Caucasian skin tones. We will look into a case where an innocent black man was arrested for faulty facial recognition technology.

Case Study: Robert Julian-Borchak Williams

On a Thursday afternoon in January 2020, an African American man named Robert Julian-Borchak Williams was handcuffed on his front lawn in front of his wife and daughters by the Detroit Police Department (Hill). After a blurry still image was taken from a surveillance tape from 2018, Detroit police used a facial recognition algorithm to falsely identify the man who had stolen over $\$3,800$ worth of jewelry from a Midtown boutique. Although the image was blurry, it was clear upon the detainment of Williams that the image was not him. The software provided to the state, DataWorks, in combination with softwares developed by the Japanese company NEC and Rank One Computing based in Colorado, used the still image to search for possible matches in their database of over 49 million images. Both of the softwares from NEC and Rank One Computing were included in the NIST study that found that African American people were at risk of being falsely identified by over 100 times. The report from the face search resulted in a list of potential identification along with their respective confidence scores, Williams, happening to be at the top of the list. The report was given to the Detroit Police with the disclaimer that "it is an investigative lead only and is not probable cause for arrest" printed at the top. The Detroit Police proceeded to use the report to locate Williams and unlawfully arrest him at his house.

While in detainment, several detectives directly compared the surveillance footage to Williams' face and both concluded that the identification was incorrect and that "the computer got it wrong". Despite the false identification, the police continued to keep Williams in custody and only released him 30 hours later upon payment of a $1000 bond. This case study illuminates the frightening fact that innocent people of color are at risk of being falsely identified and accused of a crime they did not commit; if it can happen to Williams, it can easily happen to others too.

Now that we have looked further into the direct effects of biased FRT on law-abiding citizens, we will turn the discussion to another topic relating to concerns about the use of FRT by public and private entities.

Privacy Invasions: Policing

The use of facial recognition by the government is far from unusual; in fact, United States police use their database of license and ID photos to verify the identities of hundreds of Americans every day. However, a facial recognition verification search conducted on someone who has been legally stopped or arrested is very different from using continuous real-time scans of unsuspecting, law-abiding citizens' faces as they unknowingly walk past surveillance cameras. A study from Georgetown Law's Center on Privacy and Technology revealed that many major police departments in cities across America are exploring these real-time facial recognition technologies (Garvie et. al). The study found that more than 117 million Americans' faces have been captured and are being used in a "virtual perpetual lineup", meaning that law enforcement agencies across the United States can scan their photos at any time and use unregulated softwares to track their photos in government datasets and identify them by surveillance footage. Historically, these government datasets consisted of information taken from criminal arrests or investigations. But now, one in two American adult faces have been captured in an FBI biometric network composed mostly of law-abiding Americans and can be accessed at any time without the knowledge or consent of the individual being searched. Specifically, the Pinellas County Sheriff's Office runs 8,000 monthly searches on the faces of the drivers in Florida, most of the time without requiring reasonable suspicion before running a search.

Privacy concerns for Americans and people living in cities where police use these real-time surveillance scanners are very real. A police officer could be running a search on you for a crime you didn't commit, at any given moment, without reasonable suspicion or without your knowledge. The use of FRT by the government is out of control, and very little is being done to protect the rights of American citizens. When we have developed technologies this advanced, we must ask ourselves where the line stands: between appropriate use of FRT to prevent crime and protect citizens, and where people are deprived of their amendment given

right to privacy.

Taking Surveillance Too Far: Clearview AI

Clearview AI is a commercially produced American facial recognition company that sells its software for, which it claims, only governmental use. Some of the customers of Clearview AI include: the FBI, the DEA, US Secret Service, and many major police departments across the country (Ng). However, after a data breach in early 2020, it was revealed that not only was the company selling to non-governmental agencies, but it was also scraping digital images from public websites on numerous social media platforms including Twitter, Facebook, YouTube, Instagram, and others. Upon the leak, several countries' federal agencies began investigation into the company and found that Clearview AI did not attempt to obtain consent from the websites before taking the images for their collection. It was also revealed that the use of the stolen images was completely illegitimate, as the images were being used to expand their database for use by private entities as well as several large commercial companies for mass identification of individuals through surveillance cameras. Since then, many of the targeted websites have sent cease-and-desist letters to Clearview AI requesting that the information be permanently deleted as it violated their terms and services agreement.

If you've walked into a department store in the last two years, chances are you've already been pre-identified by a real-time surveillance camera that was programmed off pictures of you stolen from your social media pages. Not only are law enforcement agencies abusing the powers of FRT for illegitimate use, but the companies who create the technology do too, and they're building the algorithms off illegally collected data. Privacy for people living in the United States no longer exists; it is extremely likely that your photos are stored in a database, accessible to the government and commercial companies at any time, and you have little to no say in how your information is being used.

Conclusions

While it is certainly true that the fascinating science of machine learning and facial recognition has the potential to protect citizens and identify criminals caught on camera, we have seen that when left unregulated, we risk false identifications and privacy invasions. Now we must ask ourselves: how do agencies utilize these helpful technologies without violating your right to privacy? Here, I propose two potential options for regulating facial recognition algorithms.

Testing for Algorithmic Bias

The Georgetown investigation into how police are using facial recognition came upon an important conclusion: facial recognition algorithms are not being tested for bias, and the engineers that develop the algorithms have no idea how to test them. When programming an algorithm, especially for use by law enforcement, it is critical that the algorithms are being tested for racial bias to ensure that no one group has more or less accurate results over another. To prevent more instances like the Robert Williams instance from occurring, it is suggested that any producers of facial recognition algorithms must first pass a bias test performed by the NIST. The algorithms must pass the test before being sold to any agencies, and the results of the test should be made publicly available for potential buyers to see. Then, buyers are able to see the accuracy of the algorithms before purchasing, and any algorithms that do not pass a fairness test will not be able to be sold. Law enforcement and federal agencies should be obligated to ensure that public funds are not being used to perpetuate an unjust system that targets marginalized groups. Regulating the softwares used by law enforcement will help prevent more people from being penalized for crimes they did not commit. Europe is in the process of passing legislation that requires prior assessment of facial recognition systems, including real-time identification, before replication. The United States should also move towards these regulations.

Passing Legislation To Limit How Facial Recognition Is Used

Even after FRT's are tested for bias, we still encounter the issue of unjustified use of the technology. Thus, we must also ensure that FRT's are being used within reason, and punishing users who abuse them. Proposed legislation from the Georgetown report suggests that law enforcement may only access their database of images only if they possess reasonable suspicion that a person has either committed a crime or could commit a crime. Other legislation proposed by Fleischer would suggest that within a reasonable amount of time, a senior officer of a law enforcement agency must provide information relating to the justification of a search and state explicitly how the information is being used in the criminal investigation. The information must be posted to a public website for anyone to access. These suggested laws are intended to prevent unjustified use of FRT on law-abiding citizens.

As pertaining to the Clearview AI incident, legislation must be passed in order to protect people's images from being scraped off the internet without permission. Today, Clearview AI has faced several lawsuits in the states and has been fined over 20 million dollars by the UK's Information Commissioner's Office with more penalties to come. While Clearview AI has faced legal punishment, incidents like these happen frequently, and the companies rarely face severe legal action. In order to fight for our own personal right to privacy, we must demand that not only the United States government pass more legislation relating to the legal procurement of photos online for commercial use, but we must also demand proper punishment of the companies that violate these rights in the future.

Facial recognition technologies are powerful tools that can be used for many good purposes, but when we let the people who create or use them go without regulation, they can be used by the same agencies that claim to protect our personal freedoms to violate them. We must push the United States government for stricter regulation of these technologies in order to obtain unbiased algorithms that will not perpetuate an unjust system that targets marginalized groups and to protect our government-given right to privacy.

Bibliography

Turk, Matthew, and Alex Pentland. "Eigenfaces for Recognition." Journal of Cognitive Neuroscience 3, no. 1 (1991): 71-86.

Merler, Michele, Nalini Ratha, Rogerio S Feris, and John R Smith. Diversity in Faces, 2019.

United States, Congress, Senate. Federal Acquisition Streamlining Act of 1994.

Fleischer, Rachel S. "BIAS IN, BIAS OUT: WHY LEGISLATION PLACING REQUIREMENTS ON THE PROCUREMENT OF COMMERCIALIZED FACIAL RECOGNITION TECHNOLOGY MUST BE PASSED TO PROTECT PEOPLE OF COLOR." Public Contract Law Journal 50, no. 1 (2020): 63-89.

Phillips Jonathon , Jiang Fang, Narvekar Abhijit, Ayyad Julianne, and O'Toole Alice. 2011. An other-race effect for face recognition algorithms. ACM Trans. Appl. Percept. 8, 2, Article 14 (January 2011), 11 pages. DOI:https://doi.org/10.1145/1870076.1870082

Grother, Patrick, et al. NISTIR 8280, National Institute of Standards and Technology, Face Recognition Vendor Test, Part 3: Demographic Effects.

Buolamwini Joy and Gebru Timnit, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, 81 Proc. of MACHINE LEARNING RES. 1, 8 (2018).

Hill, Kashmir. "Wrongfully Accused by an Algorithm." Chicagotribune.com, 2 July 2020, https://www.chicagotribune.com/featured/sns-nyt-wrongfully-accused-by-algorithm-20200702-

exziolvlhfeb5hn5yhg4vpiblu-story.html.

Garvie, Clare, et al. "THE PERPETUAL LINE-UP: UNREGULATED POLICE FACE RECOGNITION IN AMERICA." Center on Privacy & Technology at Georgetown Law, 18 Oct. 2016.

Ng, Alfred. "Clearview AI Hit with Cease-and-Desist from Google, Facebook over Facial Recognition Collection." CNET, CNET, 6 Feb. 2020, https://www.cnet.com/tech/services-and-software/clearview-ai-hit-with-cease-and-desist-from-google-over-facial-recognition-collection/.