An Overview of Facial Recognition Technology

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Daniel McNamara

Spring, 2020.

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Worthy Martin, Department of Computer Science

An Overview of Facial Recognition Technology

A Literature Review Focusing on the Current Landscape and Future Potential of Facial Recognition

Daniel McNamara
University of Virginia
djm3am@virginia.edu

ABSTRACT

Facial recognition technology (FRT) is one of the most exciting and powerful tools for security, surveillance, and identification. While FRT's use in government, military, and social media opens massive debates about ethics and privacy, it is also important to consider the uncertainties surrounding its effectiveness and data security. These issues range from algorithmic racial bias to security of facial databases. This literature review will analyze existing implementations and inherent security risks of FRT in various fields. The reviewed literature includes technical papers on deep-learning FR algorithms, their uses, and their security weaknesses. Also reviewed is a United States Congressional report on current and future implementation of FRT by agencies of the Department of Homeland Security. The goals of the research are to identify strengths and shortcomings—in the design and security—of FRT systems in order to develop a better understanding for strategies to design better FRT algorithms and implement them in an effective, ethical, and secure manner.

INTRODUCTION

The development and implementation of facial recognition technology (FRT) presents an exciting, but difficult challenge to developers, governments, and private institutions. FRT promises potentially groundbreaking functionality in a diverse set of fields, from border security to medicine. There are also several uncertainties surrounding the widespread adoption of FRT, such as general inaccuracy, racial bias, and privacy violations. The future of FRT is heavily dependent on identifying and combating these issues.

This literature review primarily summarizes and analyzes several papers regarding current and future implementations, potential government regulation, and general concerns about accuracy and security. Reviewed materials include a congressional report on FRT from the United States Government Accountability Office (GAO), a technical report on a deep machine learning facial recognition algorithm, a study on model inversion attacks (MIA), and a review of regulation strategies for FRT in the private sector. The aim of this paper is to provide a broad overview of several topics related to FRT and its regulation, find connecting threads among the reviewed material, and synthesize a general outlook of the current state and potential future of the technology.

BACKGROUND

Facial recognition is a field that applies concepts from computer vision and machine learning to develop algorithms that can identify, categorize, or authenticate human faces by pinpointing and measuring features from an image. These algorithms are improved by using training databases of real human faces.

A convolutional neural network (CNN) is an example of a machine learning model that is often used to analyze images. CNNs are introduced and used to classify images in CS-4501 Intro to Computer Vision.

LITERATURE REVIEW

1 Deep Facial Recognition System Using Computational Intelligent Algorithms

The first paper, published in PloS ONE by Abdelminaam, Almansori, and Taha, proposes a deep CNN model for facial recognition. The proposed algorithm claims to be significantly faster than traditional CNNs at recognition, especially in images with challenging angles and lighting.

The proposed system first preprocesses the images by identifying regions of interest in the images, which are regions in the images that appear to be human faces. The images are then passed to a modified version of AlexNet, a GPU-implemented CNN, to extract significant facial

features from the image. A feature vector is then passed to the recognition phase and either identified or, if it is a new subject, registered to the system for future identification.

The preporcessing phase uses Viola-Jones detection to find faces in the images. The Viola-Jones detector is popular because it is highly accurate while working in real-time to detect objects quickly. It works by scanning the images with moving windows and using Haar filters to calculate feature values based on differences in pixel values. Based on these values, the detector determines the location of any face in the images.

After faces are detected, the images are then passed to the recognition phase. The proposed framework uses cloud and fog computing to distribute the computational workload. On average, the proposed system outperformed other popular approaches in precision, accuracy, and specificity.

2 Automatic Eye Disease Recognition System Using Machine Learning

Akram and Debnath proposed a novel use for FRT in the medical field in this paper published in the Turkish Journal of Electrical Engineering and Computer Science in 2019. Their approach involved applying a deep convolutional neural network (DCNN) model to identify eye diseases such as cataracts and trachoma. Many eye diseases are typically diagnosed by visual observation, but this is not viable in many remote areas around the world. This application allows more people to receive access to accurate diagnoses without needing to see an eye doctor in-person.

The system first detects facial features in the input image so that it can identify and isolate the eyes for classification. It uses histogram of oriented gradient (HOG) and a support-vector machine (SVM) classifier to achieve this. The image of the eye is then ready to be classified or used for training data for the DCNN.

After supervised training using a set of images featuring seven different eye diseases, the DCNN achieved an impressive 98.79% accuracy rate by classifying 334 of 351 images correctly. This accuracy outperformed SVM models significantly. The paper also calculates specificity and sensitivity, which consider proportions of true positives, true negatives, false positives, and false negatives. The model featured a 97% specificity and 99% sensitivity, which means that the was incredibly proficient at correctly identifying diseases without a high number of false positives or negatives.

3 Analysis of Model Inversion Attack on a Convolutional Neural Network

Published in the Korean Society for Internet Information's journal Transactions on Internet and Information Systems by Khosravy et al., this paper analyzes the vulnerabilities of cloud-based, deep-learning convolutional neural networks (CNN) to model inversion attacks (MIA). Deep-learning networks such as CNN require training using a large set of images. Although the fully-trained network does not contain the original images, it could potentially be vulnerable to a MIA. This type of attack works by reversing the model from its output and attempting to estimate its corresponding input. In the case of a facial recognition CNN, the training data contains images of real people which could potentially be used for malicious purposes. This poses an important challenge to developers of deep-learning systems to focus on security of their model and datasets.

The paper considers MIA under a gray-box scenario, which assumes that the attacker has some information regarding the structure and parameters of the CNN model. The attack works by providing a seed image to the recognition model which calculates confidence scores and returns the class label with the highest confidence. The end goal of the MIA used in this study is to produce a clone of the seed image by defining a loss function and using an iterative gradient descent process, as shown in Figure 1.

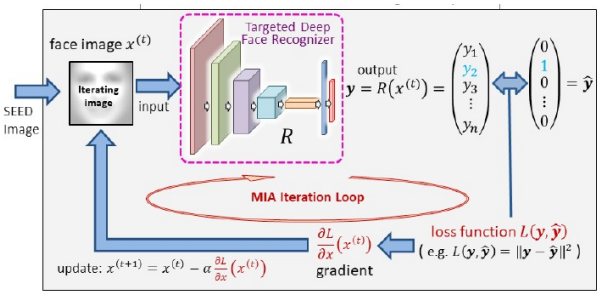

Figure 1: Visualization of model inversion attack in gray-box scenario (Khosravy et al., 2021).

For the purposes of this study, the researchers identified five images from a large database and used them as seed images for MIA. The attack yielded several clone images for each seed image that were close approximations of the seed images. The sets of clone images were judged

subjectively, by a numeric evaluation program, and objectively, by a group of humans. While only two of the sets were scored very highly by the numeric evaluation, all of the clone image sets show clearly recognizable facial features to the human eye. This is important because the danger of an MIA is that the attacker may discover the identity of an image's subject and use that information maliciously.

Despite using a relatively simple MIA approach, the researchers were able to reverse a complex CNN with multiple convolutional layers to produce reasonably recognizable approximations of the seed images. This highlights the vulnerabilities that are present even in deep learning models such as CNNs, especially for a more complex, deeper MIA. This is a serious threat that must be considered in the development and implementation of facial recognition models.

4 FRT Regulation in the United States Private Sector

Departing from the technical papers on FRT, the next paper is an article published in the Stanford Technology Law Review in 2020 that broadly considers the benefits and concerns of FRT and makes recommendations for federal regulation.

Examples of FRT in the private sector include systems that track workers hours or identify common shopping patterns. The common concern for many of these uses is data security and privacy. Companies address this by encrypting private biometric data and storing them on secure devices and servers. This is vital because people often are not asked for consent and cannot request that their photographs be removed.

Though focused on the private sector, the paper considers the United States government as a consumer of FRT, especially in law enforcement. The paper identifies Clearview AI as a major stakeholder in FRT regulation. Clearview scrapes images from several popular social media sites and stores them for use in its own FRT program, which is used by hundreds of law enforcement offices. Privacy activists contest Clearview's right to scrape and store these images and have sued them in a number of states.

On top of data security and privacy concerns, detractors of FRT will point to a number of cases of bias and general inaccuracies in the algorithms. Many algorithms struggle to correctly identify faces of women, children, and racial minorities. This could be due to imperfections in the training process or the methods used to identify features in the images. Amazon's algorithm for identifying qualified job candidates was found to favor men based on outside historical trends. This is a clear example of real-world bias translating into algorithmic bias.

The Department of Homeland Security (DHS) found similar issues in its border security FRT system, which is discussed in more detail in the next piece of reviewed literature. In a hearing regarding this system, one congress member asked if the benefits of FRT outweigh its potential for "automated discrimination." When accurate, FRT allows companies and government agencies to operate much more effectively and efficiently, but it is not clear that these systems are currently consistent and fair enough for use.

Any regulation must consider the trade-off between the convenience and efficiency of FRT with the need for accuracy and fairness. The paper does not recommend a broad, one-size-fits-all approach to regulation. Instead it suggests precise guidelines for companies too follow, using trade secrecy guidelines as a reference.

5 Congressional Report from USGAO

In 2020, the United States Government Accountability Office (GAO) was charged by the Senate to develop a report on the implementation of FRT by Customs and Border Patrol (CBP) and Transportation Security Administration (TSA). The 102-page report assesses the deployment of FRT by the two agencies and specifically focuses on privacy compliance and tracking the accuracy of the systems.

In 2017, CBP began implementing FRT as part of its entry-exit tracking program in order to verify identities of foreign nationals as they enter and exit the United States. As travelers arrive or depart, airports take their photographs along with other biometric and biographic information. The facial images are stored in the Department of Homeland Security (DHS) Office of Biometric Identity Management's Automated Biometric Identification System. CBP intends to work toward expanding its FRT systems to sea and land ports as well. In 2018, TSA partnered with CBP to integrate FRT into its pre-flight security protocols.

CBP's FRT system, known as the Traveler Verification Service (TVS) makes one-to-one and one-to-many comparisons to verify traveler identity. It works by capturing live photographs of travelers as they enter or exit the country and converting the images into mathematical representations, known as templates. TVS compares the template to a previous image of the traveler—which is transmitted from the DHS database or scanned from travel documents—to verify their identity. CBP claims that the current system has correctly matched over 90 percent of traveler photographs, which falls short of the performance goal of 97 percent. Similar pilot programs are in progress at several pedestrian, vehicle, and sea entry-points with similar accuracy numbers.

The primary area of concern addressed in the report is privacy protection and auditing of the systems. According to its self-imposed guidelines, CBP is required to clearly provide subjects with information, including how to opt out, in locations where FRT is used. CBP and its partnered airlines have failed to consistently follow these guidelines and, furthermore, the CBP only audited an airline partner for compliance on one occasion. It is particularly concerning that, despite claiming high accuracy performance, CBP's monitoring process did not notify their officials when the performance fell below minimum guidelines. The report recommends that the CBP focus on improving compliance to its privacy guidelines and ensuring that its audit process is expanded.

Another particular area of concern with FRT systems such as TVS is the storage and transmission of private data, such as photographs and biographic information. Photographs taken at entry-exit points are encrypted before transmission to TVS where they are permanently converted into templates. The templates are stored for up to 12 hours for U.S. citizens and up to 95 years for foreign nationals. The report found that a full cybersecurity evaluation for the system was not possible at the time.

One of the biggest weaknesses and most controversial topics in FRT systems is racial bias among ethnic minorities. This is especially relevant for CBP, since a large portion of travelers profiled by TVS are non-white. The report includes an appendix detailing findings from the National Institute of Science and Technology (NIST) regarding bias in several facial recognition algorithms. There was varying accuracy among the algorithms depending on sex, age, and country of origin, with the latter presenting the greatest discrepancies. For false positives, the algorithms were found to be "10 to 100 times less accurate for some demographics. Specifically for false positive rates, algorithms were less accurate for West and East African, American Indian, African American, and Asian populations, and more accurate for Eastern European populations," (pg. 76). These differences in accuracy are typically explained by issues with training data or photograph quality. Some level of inaccuracy is to be expected with any burgeoning recognition technique, but CBP needs to consider this bias as it continues to expand and develop its system.

TAXONOMY

The reviewed papers each give a different perspective on the general area of FRT. The first paper, on a facial recognition system using a deep CNN, covers a similar area to the paper on automated eye disease detection using a deep CNN. Both papers illustrate the benefits of applying machine learning and neural networks to facial recognition algorithms. Because these approaches are computationally expensive, they are often performed in distributed environments such as cloud computing. Thus, these models can be vulnerable to a number of attacks, such as the model inversion attacks discussed in the third paper. So while the first two papers demonstrate the incredible potential for these systems, the third paper emphasizes the importance of privacy and data security in FRT. These three technical papers are juxtaposed with two non-technical papers that provide insight into the real-world implementation and regulation of FRT in both the private and public sector.

CONCLUSION

Collectively, the reviewed literature provides a basic overview of the field of facial recognition, from benefits to security concerns to real-world implementation. The materials covered convolutional neural networks and other concepts that are discussed in CS 4501 Intro to Computer Vision. The security concerns in cloud-based models, especially the model inversion attack, is closely related to the attacks studied in CS XXXX Defense Against the Dark Arts.

It is difficult to discuss or summarize FRT without considering real-world implications, so two non-technical papers were included in the literature review. These, along with technical descriptions of FRT systems, help to provide a more complete picture of the landscape of facial recognition development, implementation, and regulation.

The potential good that FRT can provide in areas such as medicine is contrasted with troubling concerns about racial bias, privacy, and data security. It is indisputable that these novel approaches to facial recognition using machine learning and neural networks are an immensely powerful tool, but the nature of biometrics makes the FRT question a nuanced and complicated issue.

REFERENCES

[1] Diaa Salama Abdelminaam et al. 2020. A Deep Facial Recognition System Using Computational Intelligent Algorithms. *PloS ONE*, vol. 15, no. 12, 3 Dec. 2020, pp. 1 – 27.

[2] Ashrafi Akram and Rameswar Debnath. 2020. An Automated Eye Disease Recognition System from Visual Content of Facial Images Using Machine Learning Techniques. *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 28, no. 2, 1 Feb. 2020, pp. 917 – 932.

[3] Mahdi Khosravy, Kazuaki Nakamura, Yuki Hirose, Naoko Nitta, and Noboru Babaguchi. 2021. Model Inversion Attack: Analysis under Gray-box Scenario on Deep Learning based Face Recognition System. *KSII Transactions on Internet & Information Systems*, vol. 15, no. 3, 1 Mar. 2021, pp. 1100 – 1118.

[4] Elizabeth A. Rowe. 2020. Regulating Facial Recognition Technology in the Private Sector. *Stanford Technology Law Review*, vol. 24, no. 1, 1 Oct. 2020, pp. 1 – 54.

[5] United States Government Accountability Office. 2020. Report to Congressional Requesters: Facial Recognition. *GAO-20-568*. 1 Sep. 2020, pp. 1 – 101.