# NATURAL EYE CONTACT USING CONVOLUTIONAL NEURAL NETWORKS

# LIMITATIONS OF VIDEO CONFERENCING AND ITS EFFECTS ON SOCIETY

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
David Chen

November 24, 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.

Signed Date: 11/24/2020

Approved: Date:
Mark Sherriff, Department of Computer Science

Approved: Date:
Catherine D. Baritaud, Department of Engineering and Society

Over the course of recent months, COVID-19 has caused a great decrease in the amount of in-person contact. Events are being moved virtual, and the use of video conferencing and video calls have increased greatly. Online video conferencing applications including Zoom, Microsoft Teams, and Google Hangouts, have seen a great growth in usage, since the beginning of the COVID-19 pandemic. Zoom in particular grew from averaging 56,000 daily downloads in January of 2020, to averaging 2.13 million daily downloads in March of 2020 (Iqbal, 2020). These applications, however, are limiting in several ways. One key limiting factor is that eye contact, a key aspect of nonverbal communication, is not accessible to users.

The STS topic will conduct research and analysis of data on the technology of video conferencing and its effects on society. This research will be tightly-coupled with the technical project by providing the motivation for the technical project. The technical project will attempt to come up with a revolutionary solution to one issue, lack of eye contact, discussed in the STS topic. My advisor for the STS research is Professor Catherine Baritaud of the Department of Engineering and Society.

## NATURAL EYE CONTACT USING CONVOLUTIONAL NEURAL NETWORKS

Video calling can be tiring and uninteresting, due to many factors. One of them is a lack of eye contact. The issue with current technology is that the camera is not located at the caller's eye that is on the screen. Instead, it is either at the top of the screen or the bottom. Because the distance between the camera and the eye exists, the camera, which is the other person's eye, won't provide an accurate view of what the person sees. Figure 1 (p.2) Shows a clear visualization of why this happens.

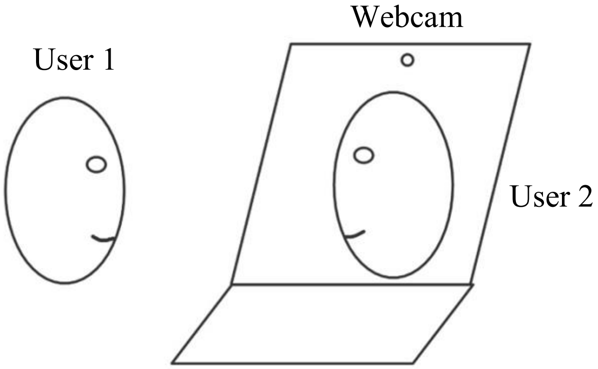

Figure 1: Visualization of Typical Video Call: There is a clear difference in the location of the camera and User 2's eyes (Chen, 2020)

While there have been several attempts to achieve eye contact in video calls, they all have shortcomings. Apple has recently released a feature that tries to correct eye contact during FaceTime calls. However, one user (Schukin, 2019) tweeted a video showing that when placing a straight object near his eyes, it caused FaceTime's new feature to bend it, making it look wavy. Intel also published a paper describing their process of correcting eye contact. Their approach was also to only edit the area around the eye (Fadelli, 2019). DIY Perks, a YouTube Channel, released a video that showed viewers how to build an apparatus that involved a second webcam and a two-way mirror (DIY Perks, 2020). This physical contraption would be too complicated for an average user to build, and too bulky to carry around.

The goal of this project will be to utilize machine learning to simulate eye contact. A Convolutional Neural Network (CNN) model will be designed and used to accomplish this. A CNN model is a type of Neural Network used in machine learning that is commonly used for learning involving images. The idea is to give the model an image taken from a real webcam at the top of the screen, and have it predict what the image would look like if the location of the camera was shifted down to where the caller's eyes are on the screen.

## APPROACH

The project will involve 4 major steps: data collection, data processing, training, and testing. First, data will be collected in the form of video of the faces of people in conversation that is recorded by 2 devices simultaneously at different angles. These angles can be seen in Figure 2.

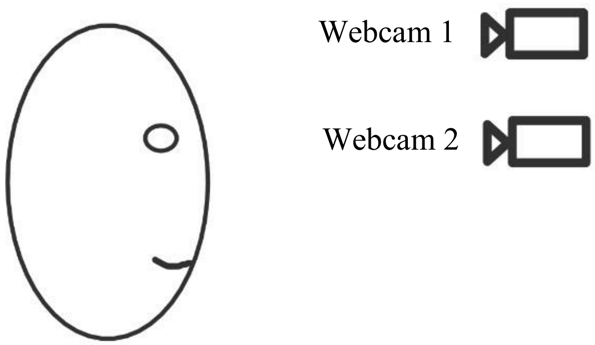

Figure 2: Visualization of Camera Setup: Webcam setup for recording data (Chen, 2020)

The first angle will represent the location of a webcam used for video calling. The second video will be taken from a point near the center of the screen, where the eyes of a caller will most likely be during a video call. Then, the videos will be synced and the frames will be paired to use for training and testing. Then, using Python and TensorFlow, a machine learning library in Python, the data will be fed to a CNN model to train the parameters. Once the model is trained, a second portion of the data will be used to test and evaluate the model. This step will involve measuring how similar the generated images are compared to the real images taken by the second webcam.

In order to make the project more reasonable in scale so that it can be completed within a single semester, it will be a proof of concept project. This means that the results of this project will either validate or invalidate this approach so that it can give more understanding to whether or not this approach should be used on a larger scale project. I will use my laptop and 2 webcams for recording data on myself. This data will be processed and used to train a model on a server using computing resources such as Google Colab or Amazon Web Services (AWS).

## DESIRED OUTCOMES

The anticipated outcome for this project is to come up with a CNN model that when given an image it has never seen before, it could output the image that is nearly identical to the image taken from the second webcam. This would validate the approach of using CNN models to artificially adjust the location of the camera so that it would appear as if the camera was at the center of the screen. It would then open up doors for future research where data taken from various people, in various rooms, with various backgrounds can be used to train a model that can be refined and used commercially.

This project will be completed over the course of the fall semester of 2020 under the supervision of Mark Sherriff, a professor in the Computer Science Department at the University of Virginia. Upon completion of the project, I will write a proof of concept paper explaining the fine details of the methods I used, as well as describing the options of next steps available to continue in the field of research of using CNN models to improve video calls.

## LIMITATIONS OF VIDEO CONFERENCING AND ITS EFFECTS ON SOCIETY

Throughout quarantine, video conferencing has exploded in usage. Meetings and classes have moved online, and video calling has never been more widely used. However, it is critical to realize that online video calling will not go away anytime soon. Even though it may seem that video conferencing came from the COVID-19 pandemic, it would be foolish to say that it will disappear when the pandemic is over. Video conferencing existed before and will continue to exist because of the society we are in. People have been able to form connections and maintain relationships using the internet before quarantine and will continue to do so after. Because of the digitalization of society, the number of meetings done through video conferencing will continue to grow.

It is generally accepted that nonverbal communication, including tone of voice and body language, is far more important than verbal communication. While it is impossible to assign a quantitative value that accurately describes the qualitative importance of nonverbal communication, a study at UCLA created the 7% rule. The study claimed that only 7% of communication is verbal, and that 38% was through tone of voice, and 55% was body language (Strain, 2020). Eye contact is a crucial part of nonverbal communication. One way to tell if someone is lying is to look at their eyes. If the person is lying, that person will likely break eye contact, or make too much eye contact (Brown, 2012). Eye contact is also critical to detect the other person's level of interest and engagement (Segal, 2019). Eye contact is very necessary for conversations because of the amount of body language it can communicate. It is one of the most important factors that make conversations more intimate when they are in-person. Murphy (2020) concludes in her article "Why Zoom is terrible," by saying that "no facial cues are better than faulty ones." Since there are so many things that hinder communication in video calls, Murphy proposes that everyone should just stick to regular phone calls.

## OBJECTIVE

The objective of this research project will be to discover and bring light to the culture that exists because of the way video conferencing currently exists. Understanding the effects of video conferencing on meetings will make way for seeing how it influences society.

Using the Actor Network Theory (ANT) model, we will analyze the dynamics within video conferencing. This model is a form of analyzing technologies where each actor has changings relationships with other actors within a given network (Law and Callon, 1988). One main actor would be the hosts, which can be teachers, team leaders, speakers, and friends that take initiative. Another category of actors are the participants, which include students, team members, and other friends. Other actors that influence the technology of video conferencing include the company or institution that decides on using the software, the family, or the people who share the WiFi of the users, as well as the engineers and developers who created the software. This network can be visualized in Figure 3.

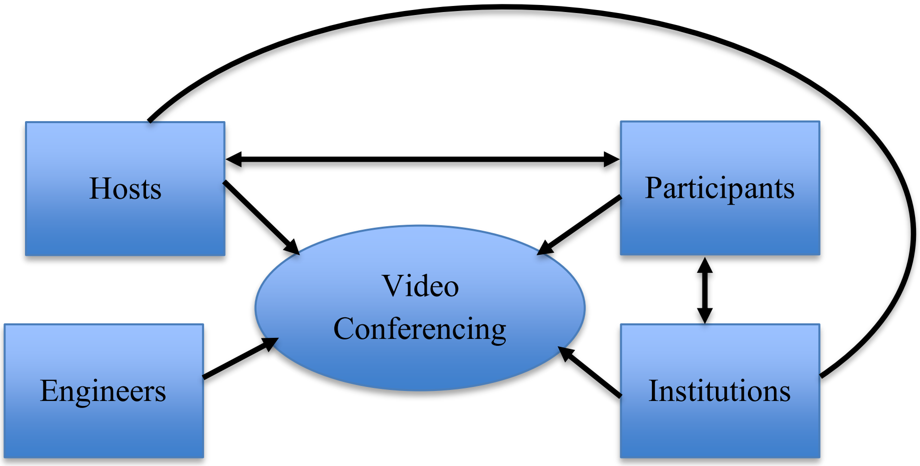

Figure 3: Visual Representation of the Video Conferencing Network: The interactions between different actors within the Video Conferencing network (Chen, 2020).

The hope is that by using ANT, it can bring deeper understanding to the technology of video conferencing and its effects on society. This further understanding will allow for finding different ways to improve on the technology, and help bring more answers to the question of how the technology of video conferencing influenced society, and how it will continue to influence society.

The paper written at the end of this research will be a scholarly article that breaks down the different uses of video conferencing, analyzing the different actors and the relationships between them, and revealing the ways video conferencing and its stakeholders are evolving and adapting to each other.

## WORKS CITED

Brown, J. (2012, August 15). Nonverbal communication: The importance of eye contact. Retrieved September 24, 2020, from https://www.thelanguagelab.ca/posts/nonverbal-communication-the-importance-of-eye-contact/

Chen, C. (2020). *Natural eye contact using convolutional neural networks* [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chen, C. (2020). *Natural eye contact using convolutional neural networks* [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chen, C. (2020*). Limitations of video conferencing and its effects on society* [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

DIY Perks. (2020, May 27). *Weird webcam mod that enables eye-contact conversation* [Video]. YouTube. https://www.youtube.com/watch?v=2AecAXinars

Fadelli, I. (2019, June 26). Intel researchers develop an eye contact correction system for video chats. Retrieved from https://techxplore.com/news/2019-06-intel-eye-contact-video-chats.html

Iqbal, M. (2020, July 20). Zoom revenue and usage statistics (2020). Retrieved September 24, 2020, from https://www.businessofapps.com/data/zoom-statistics/

Isikdogan, L. F., Gerasimow, T., & Michael, G. (2020). Eye contact correction using deep neural networks. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). doi:10.1109/wacv45572.2020.9093554

Murphy, K. (2020, April 29). Why Zoom is terrible. *The New York Times*. Retrieved from http://www.nytimes.com

Protalinski, E. (2019, October 03). Microsoft's AI-powered eye gaze tech is exclusive to the Surface Pro X. Retrieved October 08, 2020, from https://venturebeat.com/2019/10/03/microsofts-ai-powered-eye-gaze-tech-is-exclusive-to-the-surface-pro-x/

Schukin, D. [schukin]. (2019, July 3). How iOS 13 FaceTime Attention Correction works: it simply uses ARKit to grab a depth map/position of your face, and adjusts the eyes accordingly. Notice the warping of the line across both the eyes and nose [Tweet]. Retrieved from https://twitter.com/schukin/status/1146359923158089728?s=21

Segal, J., Smith, M., Robinson, L., &amp; Boose, G. (2019, June). Nonverbal communication.

Retrieved from https://www.helpguide.org/articles/relationships-communication/nonverbal-communication.htm

Strain, K. (2020, March 30). How Much of Communication is Really Nonverbal? Retrieved September 24, 2020, from https://www.pgi.com/blog/2020/03/how-much-of-communication-is-really-nonverbal/

Tung, L. (2020, July 23). Windows 10: Microsoft finally lets you make eye contact on video calls with this new preview. Retrieved October 08, 2020, from https://www.zdnet.com/article/windows-10-microsoft-finally-lets-you-make-eye-contact-on-video-calls-with-this-new-preview/

Law, J. & Callon, M. (1988). Engineering and sociology in a military aircraft project: a network analysis of technological change. *Social Problems* 35(3), 284-297. doi: 10.2307/800623