

Thesis Portfolio

CLIP-tology: Classifying Images of Unseen Categories with Zero-Shot Learning
(Technical Report)

An Analysis of Facial Recognition in Black Lives Matter Protests
(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Nikash Sethi
Spring, 2021

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

CLIP-tology: Classifying Images of Unseen Categories with Zero-Shot Learning

An Analysis of Facial Recognition in Black Lives Matter Protests

Thesis Prospectus

Sociotechnical Synthesis

In the past few decades, innovations in artificial intelligence have revolutionized the field of computer vision. A particularly interesting problem within this field is zero-shot learning, in which models are tasked with classifying categories of images that they have never seen before. Perhaps the most influential application of this style of computer vision is facial recognition. Facial recognition has recently come under heavy scrutiny as police departments around the country used the technology to track down activists participating in Black Lives Matter protests. My technical research on innovating zero-shot learning and my STS research on analyzing the controversy of facial recognition in policing together serve to ethically advance applications of computer vision.

The technical portion of my thesis focuses on improving the performance of a state-of-the-art model for zero-shot learning called CLIP. This revolutionary model learns from diverse text and images all across the web and is therefore far more robust in its applications than most previous solutions for zero-shot learning. CLIP classifies images by matching them to a set of textual prompts describing different categories that the images could belong to. My research improves the performance and accuracy of CLIP by exploring novel ways to provide more context and information to the model. With this work, I advance CLIP to be even more applicable to a diverse range of classification tasks, ranging from identifying cars and pedestrians for self-driving to recognizing individuals with facial recognition.

In my STS research, I analyze the use of facial recognition in policing through an Actor-Network Theory framework to explore the complex and intricate relationships between various actors and entities involved. Given recent events surrounding the Black Lives Matter movement, the relationships between police departments, activists, legislators, and the technology

companies that develop facial recognition algorithms have grown increasingly complicated, which demands the need for an in-depth analysis of this network. In the following figure, I identify the actors involved, their primary goals, and their support for or opposition to the use of facial recognition technologies in policing.

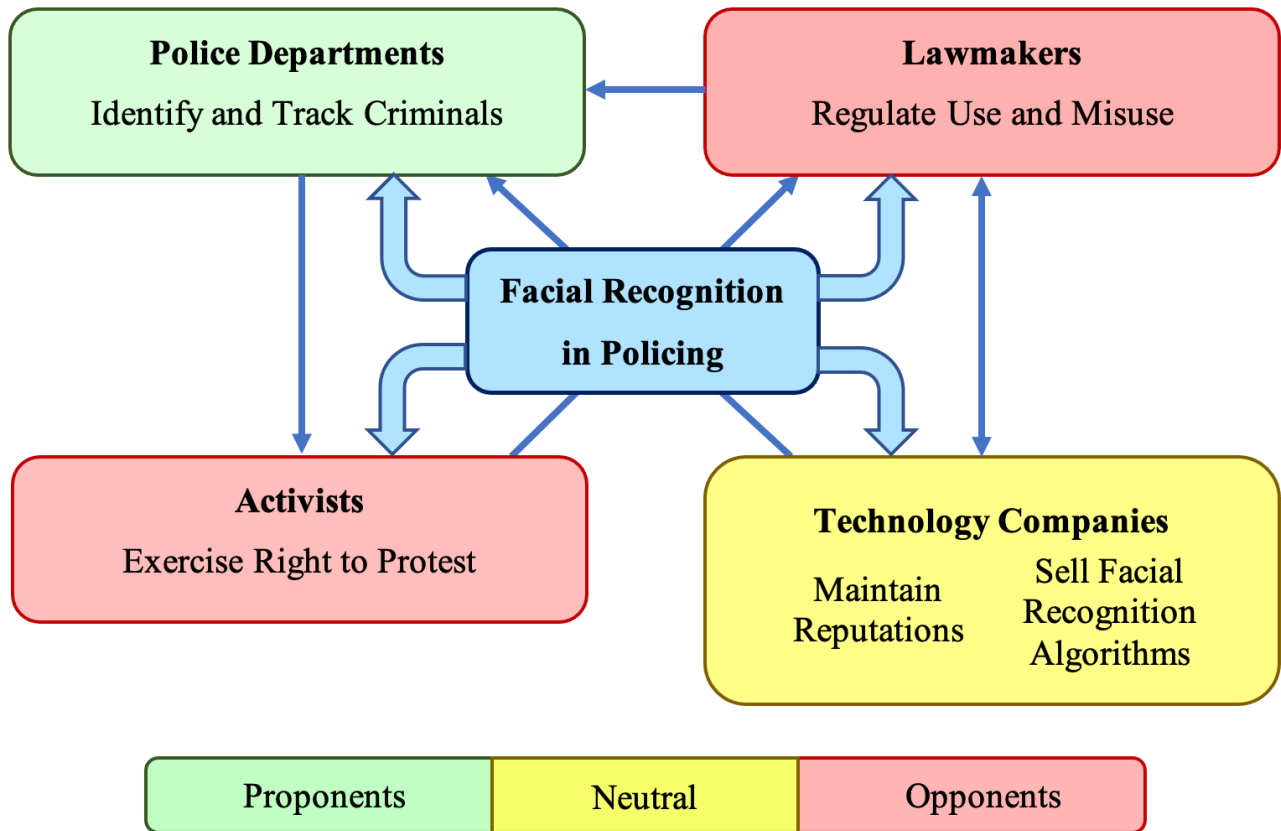


Figure 1: Actor Perspectives on Facial Recognition. Various actors have conflicting motivations and main goals related to facial recognition technologies in policing, which has deterred regulation on this practice from being widely accepted.

After breaking down each of the actors, my thesis discusses potential solutions or regulations we can apply on facial recognition that benefit society as a whole. Through my analysis, I propose a legislative approach to regulating facial recognition that is not all-or-nothing, following a case study of Massachusetts’s new police reform bills. Although such an approach still allows police departments to use potentially biased facial recognition algorithms, it

is an important step toward mitigating the disproportionate targeting of minority groups currently impacted by facial recognition in policing.

The field of zero-shot learning has led to large-scale social controversies in the application of facial recognition in policing, and as an engineer I cannot disregard the non-technical aspects and implications of my research. The tight coupling between my computer vision research in zero-shot learning and my STS research in facial recognition applications in policing serves as a testament to conscious engineering and allows me to innovate in the right way.

I would like to thank Professor Vicente Ordonez-Roman of the Computer Science Department at UVA for advising me and providing guidance in my technical research project. Additionally, I would like to thank Professor Richard Jacques for assisting me in my STS research and my thesis portfolio.