PAVAC: Privacy-Aware Vehicular Autonomous Computation

Philip Do
UCLA

Arshia Dabiran
UCLA

Vinh Nguyen
UCLA

Alexandre Hafemeister
UCLA

Charley Sanchez
UCLA

Advisor: Prof. Nader Sehatbakhsh
Lab: Secure Systems and Architectures at UCLA
Industry Partner: Nokia Bell Labs

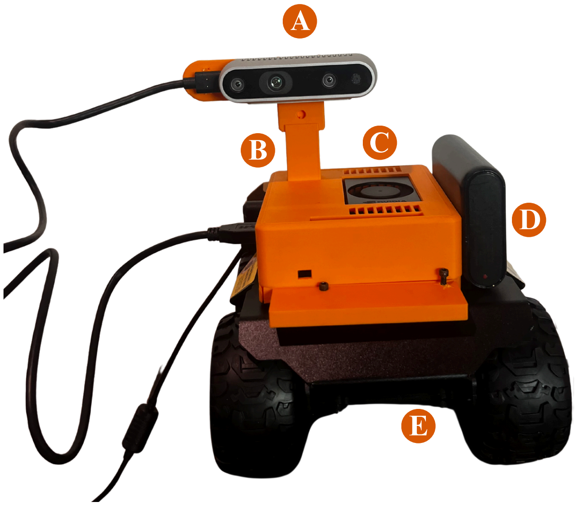

Figure 1: PAVAC System. **A** is a RealSense Depth Camera D435i, **B** is an orange 3D printed mount, **C** is a Jetson Orin Nano, **D** is a portable battery pack, and **E** is a Waveshare Wave Rover.

Abstract

The increasing prevalence of autonomous vehicles and their use of cameras for navigation will inevitably result in an increase in cameras recording us in public spaces. Concerns about being constantly recorded have led to the demand for built-in anonymization to protect individual privacy. Considering the security risks of wireless communication, our solution must be able to run entirely on hardware within the vehicle. We developed PAVAC to perform face detection, segmentation, and pixelation all within the real-time constraints of an autonomous vehicle that relies on the processed image for navigation. Comparing our model against Meta's Segment Anything Model showed an average dice score of 0.8274 and recall of 0.8971. System performance impact was also minimized with a low delay of 41 ms and decent FPS of 24. These results further demonstrate the system feasibility of our PoC.

1 Introduction

Volumetric videos provide an interactive experience for extended reality (XR) applications, including 3D telepresence, education, entertainment, and the metaverse. As XR technologies become more mainstream, privacy has become a significant challenge. Despite its growing importance, previous research on volumetric video streaming has primarily focused on system-level implementation issues, such as data movement and real-time processing. Moreover, existing works on privacy-preserving image processing have predominantly focused on single-camera, single-frame settings, making them inapplicable to complex multi-camera and 3D environments. As a result, there is a noticeable lack of studies aimed at achieving privacy-preserving volumetric video streaming while maintaining data processing efficiency.

This challenge also extends to autonomous systems, like self-driving cars, which rely on multi-view, high-resolution, continuously updated 3D data for perception and decision-making. These systems are vulnerable to privacy breaches, including sensitive data leaks such as pedestrian identities, license plates, or street-level activities [15]. Ensuring secure data transmission and storage is critical, especially when volumetric data is shared between vehicles or with central processing hubs [9]. Addressing these issues requires integrating advanced encryption, real-time anonymization, and decentralized data processing techniques to protect individual privacy without sacrificing data quality.

Our contribution is applying and evaluating real-time anonymization on a real-world camera feed, presenting a proof of concept that highlights the practicality of a privacy-aware, data-driven decision-making system called PAVAC: Privacy-Aware Vehicular Autonomous Computation (from a combination of the first initial each team member's given name). First, we ran anonymization algorithms on synthetic 3D environments using the CARLA simulator to test that our system is able to both quickly and accurately mask private information like faces in a cost-effective manner [7]. Then, we validated these synthetic results by adapting the same system pipeline to a camera-enabled rover with limited, on-vehicle computational capacity. By doing a combination of real and synthetic testing, we are able to show the practicality of system in the real world while also finding its theoretical limits. For example, the maximum number of faces or closest/furthest distance a face can be detected in an arbitrary scene. This required careful hardware selection for our requirements and resource constraints, consideration of environmental variables for testing scenarios, and final system integration of individual components.

2 Literature Review

2.1 Motivations for Autonomous Vehicle Privacy

As we move towards the future, the need for smart and sustainable infrastructure drives the need for society to adopt autonomous vehicles, as evidence has been shown for their potential for raising productivity, increasing road safety, and reducing overall fuel consumption [11]. However, a major obstacle to society's complete acceptance of autonomous vehicles is the fears and concerns of the general population regarding system incompetence, AI technology in general, and privacy [14]. In fact, surveys by Xiao et al. show that the number of civilians concerned about autonomous vehicles increased by approximately 30% between 2015 and 2019 [23]. These concerns cannot be ignored going forward, as the longer public fears about autonomous vehicles persist the more difficult it will be to receive legislative support and successfully integrate autonomous driving into society.

2.2 Data Vulnerability

In regards to autonomous vehicles as computing systems, there are various opportunities for autonomous vehicles to be attacked and have the sensitive information they manage be stolen. Hataba et al. explore how proposed methods of autonomous vehicle optimization such as platooning, traffic flow coordination, and cloud computing provide a potential opening for malicious actors to take advantage of these built-in routines to access user data without their consent [9]. A particular threat to data privacy is network vulnerabilities that allow untrusted outsiders to interrupt and receive the communication between an autonomous vehicle and an external entity [5]. Panda et al. describes how these outsiders have multiple methods for attacking these autonomous vehicles during their regular operations and how any system depending on communication with external entities runs the risk of falling prey to these attackers [15]. Considering the dangers of integrating wireless communications, for our goal of real-time masking we will show that our face detection and anonymization solution is capable of running only with onboard technology.

2.3 Privacy Protection Insights

Privacy comes in many forms and understanding the type of privacy we are specifically aiming to protect will help guide our development of a solution. In relation to autonomous vehicles, Xie et al. categorizes privacy as individual (information linked to those interacting with the vehicle), population (info about nearby locations or infrastructure), and proprietary (info revealing the inner workings of the autonomous vehicle) [24]. The project goal of masking facial features is specifically targeting individual privacy protection, so there is no need for masking the surroundings for population privacy. Ravi et al. elaborated on the types of solutions for addressing digital image privacy to which our desired method would best be described as image filtering via pixelation, which is more broadly placed in the visual obfuscation category of privacy protection [17]. From their work, we can understand that while more advanced methods of image filtering would be more resistant to image manipulation to discover the original data, these algorithms tend to be more computationally expensive than pixelation and may not be able to meet the real-time processing requirements of an autonomous vehicle.

2.4 Impact on Data Utility

Within data science, there has long been discourse on the interaction between maintaining privacy and the effectiveness of the data itself. In order to preserve the privacy of the individual, data about them must be obscured, manipulated, or removed entirely, and this loss of information hinders the ability for the collected data to fully describe the subject matter. Alharbi et al. mentioned in their work how visually distorting images to preserve privacy makes it difficult for humans to annotate data to train computer vision models on [2]. This will make obtaining training data more difficult, but we also need to examine the efficacy of training data that has been anonymized. Fortunately, Lee et al. showed that when sufficiently trained to find people, object detection models trained on anonymized data only showed only a slight reduction in accuracy when used to detect people anonymized by various methods [10]. For our goal of real-time masking, we must consider what anonymization algorithm to use that will minimize the accuracy loss for models trained on our manipulated images. When testing object segmentation and pose estimation models, Stenger et al. discovered that techniques like blurring and pixelation had the least impact on model accuracy [22].

2.5 Retroactive Masking

A vast majority of research on the topic of video masking and anonymization is applied to existing or prerecorded datasets for training and testing purposes. This is the most controlled and easily testable means of verifying model accuracy, but within the scope of using masked camera footage for autonomous vehicle sensing and navigation, this is insufficient proof of viability. To prove that a masking model is capable of use within autonomous vehicles, it must be proven to operate fast enough for real-time object detection to alert the vehicle. Furthermore, many of these research experiments run these anonymization operations on servers much more powerful and capable than the on-board processing of modern vehicles, so real-time testing must be done on comparable hardware.

The work by Lundström et al. highlights many of the unique challenges of masking for autonomous vehicle footage [12]. The paper describes the research done for preserving privacy for stationary security camera footage by initially establishing a portion of the 3D space to be considered private and removed from the video feed entirely. To be able to interpret the 3D space being recorded, Lundström et al. use depth cameras to discern which objects are behind the privacy partition and need to be masked and which items are in front of this partition and can remain in the scene. In our research, we will similarly be using depth cameras for data collection and model accuracy verification, but our need for face object detection, segmentation and targeted anonymization mean that the 3D space partitioning method described in this paper cannot be applied to our research. Regardless, the documentation of the depth feed will be useful for our implementation, especially in future development when autonomous vehicle navigation would heavily depend on such data.

The work that most influenced our work was the previous software project by Shah et al. during development of their own face masking program [20]. This code had previously only been run on simulated or prerecorded data so it had not yet been optimized for our real-time needs, but their work guided much of our early development and inspired the initial concept for our research.

2.6 Real-Time Anonymization

A major hurdle in our goal for privacy preservation via face recognition and masking of autonomous vehicle footage is the need for real-time processing that will not hinder an autonomous vehicle's capacity for interpreting this visual data. Fortunately, Bentafat et al. have explored this in their concept for a privacy-preserving system for criminal face detection [4]. The intention behind their research was to design a framework for detecting criminals logged within a database by detecting either their face or corresponding license plate number while maintaining anonymity for all noncriminals recorded. Their design included a remote server communicating with camera systems that will feed them video data with which the server will perform face detection, identity matching, and anonymization prior to video storage. A similar setup is explored by Fitwi et al. for their design of a home security system that masks faces and windows captured by security cameras feeding their data to a remote cloud server for image processing and storage [8].

An important lesson to learn from these papers is that while these systems are capable of image anonymization with a sufficiently small total delay, they depend entirely on remote resources for processing and stable internet connection, which is especially dangerous if we are considering this implementation for an autonomous vehicle. When considering the security risks of this communication, such a setup would provide the risk of data breach by malicious parties that can access the stream of video footage before anonymization has been applied and endangering the privacy we sought to protect in the first place [9]. To avoid this potential risk, we must set our sights on developing an anonymization system that can run solely with the hardware onboard regardless of the stability of the wireless connection.

2.7 Masking Vehicle Footage

Computer vision research specifically designed around vehicle footage is most relevant to our goals, and exploring approaches for maintaining privacy while preserving data utility reveals unique insights for how to develop our solution. The most unique method of privacy protection was by Rezaei et al. who applied a vision to text encoder to convert vehicle or traffic camera footage to descriptive text [19]. This solution is excellent for detecting specific objects or actions that appear on the camera feed while preventing access to the visual identifiers that could be used to identify individuals without their consent. However, this loss of information is too substantial to be applied to autonomous vehicles and their exterior camera footage. Autonomous vehicles rely on both RGB data and depth data from their sensors in order to understand the relative position of everything in the surroundings and safely navigate, but the reduction of visual to textual information makes preserving this locational data much more difficult.

A similar issue is found when examining the research by Mishra et al. who developed an in-cabin security camera system which anonymizes the faces of occupants in the autonomous vehicle and can re-identify anonymized individuals that have already been recorded in the database [13]. This paper not only includes extraneous features that would strain on an autonomous vehicle's onboard hardware, but it also falls into the same pitfall discussed with Lundström et al. in their security risks with remote server communication [5, 12]. Fortunately, since our goal does not include identifying or recording individuals we can circumvent the need for a remote database or server and avoid the security risks of wirelessly communicating video data prior to anonymization.

A much more general approach to privacy protection was taken by Adeboye et al. in their work on autonomous vehicle data privacy [1]. To preserve the privacy of passengers and their surroundings, Adeboye et al. developed DeepClean for generating synthetic representations of camera data that can prevent an outside entity from discerning the user's location by heavily reducing the detail of the images while still maintaining distinct object labels. This more blanket approach helps to ensure privacy at all times for more than just facial identifiers, but when considering how visual clarity may help the human operators of a vehicle (like with rear-view cameras), manipulating the entire image may be an excessive solution for preserving privacy.

3 Methods

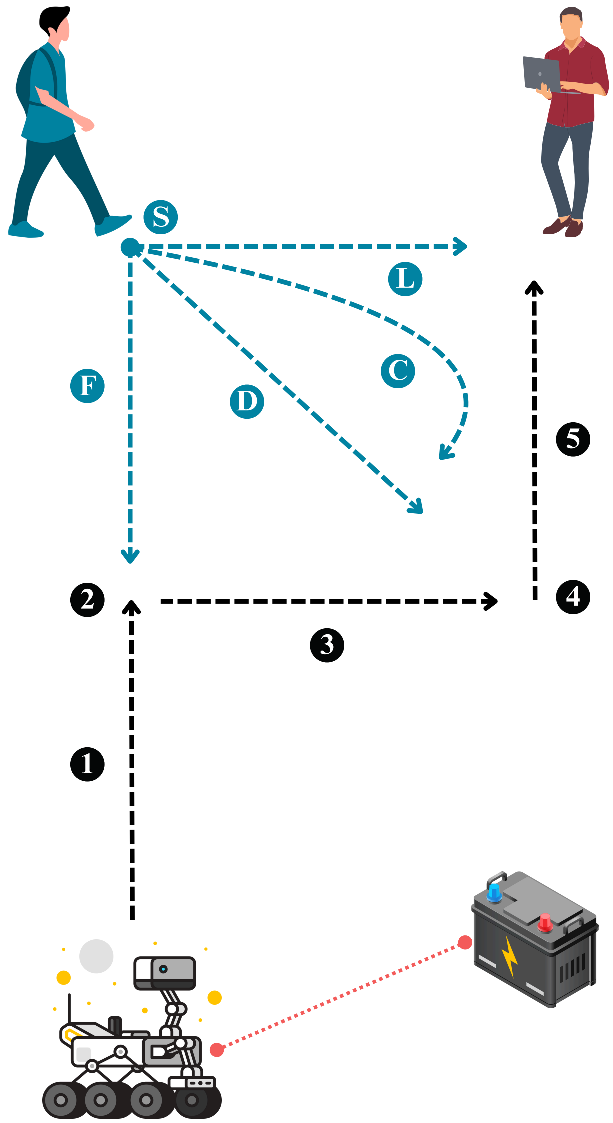

Figure 2: Real Test Scenarios Setup. The different movement paths for a person (in blue) are denoted by different letters, corresponding to Table 1. The rover's path remained the same for each test scenario. In scenarios with two people, one person (in red) remained stationary to control the rover's movement.

3.1 Research Approach

The goal of this research project is to create a proof of concept (PoC) that algorithmically applies automatic privacy masking to a video feed, enabling practical, privacy-aware control of a real-world robotic system. To achieve this, our research group split up into two teams, synthetic and real, which focused on each part of the research problem before integrating our work together towards the end. The synthetic team first worked on validating a previous experiment which used CARLA to simulate 3D environments, record the 3D scenes in video, perform masking on individual frames, and finally compile masked images into videos [7]. The real team focused on preparing the hardware of a real-world robotic vehicle with the sensors and computational resources required to collect data and run a machine learning model locally then stream that processed data to a remote user who can control the vehicle. The synthetic team tested their improved anonymization algorithm from prior work on simulated data using CARLA while the real team validated those results on the PAVAC system through metrics that evaluate performance in terms of speed and accuracy.

3.2 Assumptions and Constraints

We consider an end-to-end volumetric video streaming vehicular system that enables the sharing of video frames captured with a remote user who is able to control the vehicle. A real-life example of this would be the Coco delivery robots which can be controlled by human operator with the potential for AI enhancement [21]. To obtain a comprehensive view of the environment, multiple viewpoints of an environment are typically needed. Since the system requires generating a 3D representation of the scene for decision-making, the cameras should be capable of capturing both RGB and depth information (RGB-D). Many commercially available off-the-shelf cameras, such as (Intel) RealSense, offer this functionality. Further details of our specific setup are provided in the System Design section.

We assume that either the camera or an edge device connected to it has some computational capability to run a machine-learning model (e.g., a Jetson Orion Nano). However, the device is resource-constrained and cannot run complex tasks. As such, heavy-load tasks may need to be uploaded to a server.

The streamed environment is assumed to contain a variety of objects, including humans, trees, screens, and other items such as desks, chairs, and lamps. These objects may belong to one or multiple parties or simply be part of the background.

As with any privacy problem, it is important to thoroughly define exactly what the threat model is. Knowing who the adversary is allows us to determine which component of the overarching system is safe to trust and which agents must never receive private information. In our study, we assume that the local environment including components such as local cameras, edge computation devices, individuals, and the network connection are secure since they are fully managed by the PAVAC system while components outside this local environment such as remote users, the cloud, or external services, are considered untrusted. These untrusted components are assumed to be *honest-but-curious*, meaning they execute tasks correctly (e.g. vehicle driver will not purposely crash into a pedestrian) but may attempt to infer private information from the data they receive. With this assumption, private information must be removed locally before any data is shared with the remote user, and PAVAC cannot rely on the cloud for performing any privacy-related filtering or masking.

Since this project involves working with hardware under a strict timeline, our study was faced with multiple constraints we had to manage including time, personnel, and cost. Since we had about 10 weeks to complete the project, time was a major issue that severely limited the scope of our research since just selecting, ordering, and delivery of hardware required a substantial amount of time. The capstone occurring during the summer also limited the number of personnel we had available to collect real-world data, especially when campus is relatively empty with few students around. Finally, we also considered monetary cost when deciding which hardware components to go with since cheaper parts are more readily available, thus arriving faster, and are inherently more accessible to the general public.

3.3 System Design

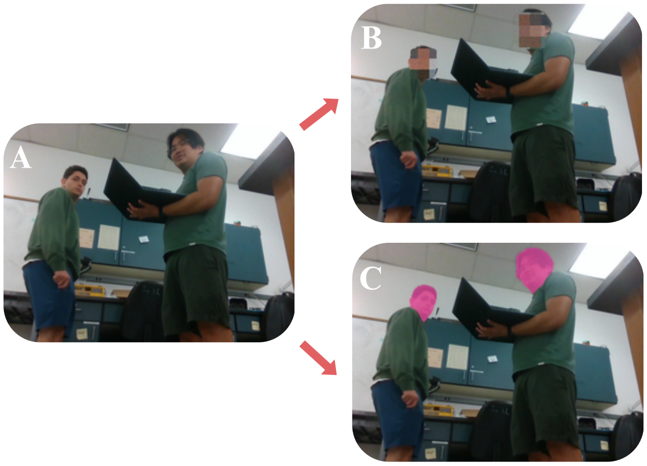

Figure 3: PAVAC Data Pipeline. **A** is the original image captured by a RealSense D435i, **B** is the anonymized image, and **C** is the ground truth mask generated by SAM.

To complement our real-world data collection, we leveraged the CARLA simulator as a controlled environment for evaluating the masking algorithm. The simulator offers several advantages: it eliminates the need for time-intensive data gathering in physical settings, ensures repeatability of scenarios, and allows us to systematically stress-test the system under extreme edge cases [7]. For instance, CARLA enables us to generate scenes with unrealistically high pedestrian densities or severe, unconventional camera angles. These conditions would be impractical or unsafe to reproduce in reality. By incorporating such synthetic scenarios into our evaluation, we extend the coverage of our test dataset and ensure that the system's performance is assessed not only in typical environments but also in challenging corner cases.

The masking pipeline itself consisted of three major improvements from the previous team's work: model selection, anonymization technique, and (Intel) RealSense camera integration.

Detection Backbone: Face-specific vs. Generic Person Detector

The prior pipeline used `fasterrcnn_resnet50_fpn` trained on COCO and filtered detections to the 'person' class [18]. While this approach performs reasonably well in general-purpose detection tasks, it is relatively heavy for embedded inference and not specialized for facial detection. More critically, because the model is designed for detecting whole persons rather than faces, it often fails in scenarios where only a face is visible without the body. This limitation leads to missed detections and incomplete anonymization. Furthermore, even when detections are produced, the bounding boxes tend to be less precise than those generated by models explicitly optimized for facial detection.

For this work, we switched to a face-specific detector (InsightFace/SCRFD/RetinaFace-500MF via FaceAnalysis [6]). This yields tighter, face-aligned boxes and removes the need to segment whole bodies. The chosen variant maintains high accuracy while achieving low latency on the Jetson Orin Nano (≈16 ms per frame in our setup), meeting real-time constraints. The face-specialized prior also reduces false positives on non-face objects that the generic "person" detector can produce in cluttered scenes

Mask Construction: Depth-guided and Box-only Options with Boundary Control

The prior pipeline constructed per-person masks by refining Faster R-CNN detections with depth information. Specifically, a depth profile was extracted around each bounding box center, and a global threshold on $|d-\mu|$ within the box determined which pixels were included in the mask. The individual person masks were then OR-combined across the frame. While this method provided reasonable coverage, it was sensitive to missing or sparse depth values and prone to over-segmentation when bounding boxes were excessively large.

In our system, we extend this design by introducing a dual-path strategy. When depth is available and reliable, we retain the depth-guided refinement for tighter segmentation. In scenarios where depth is missing or corrupted, we fall back to a box-only path that generates the mask directly from the face bounding boxes. This path supports two enhancements: (i) adjustable padding to capture peripheral facial regions such as the hairline and forehead, and (ii) elliptical fills to approximate natural facial contours more closely. To mitigate leakage at the mask boundary, both depth-guided and box-only masks are further refined through dilation with a structuring element. This dual-route design improves robustness across different sensors and lighting conditions while preserving a consistent anonymization interface.

Anonymization: Streamlined Pixelation with Noise

The prior pipeline anonymized detected regions through a cascade of operations (blur, pixelation, random color transforms, and additive noise). While this effectively removed facial identity, the multi-stage process introduced unnecessary latency, making it less suitable for real-time deployment on embedded hardware.

In this work, we adopt a simplified yet robust approach. The image is first pixelated once at reduced resolution and upsampled. Only the pixels within detected face masks are then replaced with these anonymized values. To further prevent reversibility, small stochastic noise is injected into the pixelated regions. This reduces computation to a few lightweight operations per frame, lowering average anonymization time to just a few milliseconds, while still ensuring that no identifiable features remain. The result shown in **B** of Figure 3 is a method that balances strong privacy guarantees with real-time performance.

Hardware Selection and Integration

A critical aspect of a successful PoC is moving away from simulations to the real world which has many tradeoffs, physical constraints, and less room for error. As such, hardware selection is a crucial component for proving the feasibility of an idea. As seen in Figure 1, the PAVAC system consists of RealSense Depth Camera D435i to capture RGB-D images, a Jetson Orin Nano for anonymization and other processing, a Waveshare Wave Rover as the vehicle platform, portable battery pack for power, and an orange 3D printed mount to integrate all of these individual components together. Outside of this, we used other computers such as laptops and phones to establish a wireless AP link to the system and control its video feed capture and vehicle movement. As some of us have had experience with the RealSense D435i and there are some units readily available in the lab, it made the most sense as our choice for depth camera. The Jetson Orin Nano was also a simple choice due to development experience and high availability of these relatively powerful edge devices. The choice of rover was most interesting with us opting to go for the cheapest option (≈$100) to purchase multiple rovers for future convoy experiments. The cheaper option also has the added benefit of being the most readily available with the fastest shipping time as more expensive rovers would require approved purchase orders and we did not hear back from potential corporate donors of autonomous robots. A 3D printed mount to perfectly integrate our components into a cohesive system was also a sound decision to fit our custom needs. Note that due to small battery constraints, we performed testing with the Jetson and sometimes rover connected to an Uninterruptible Power Supply (UPS) with extension cords. These had a minor impact on the rover's movement but allowed us to essentially test in any environment without worrying about constantly replacing dying batteries.

Table 1: High-level Overview of Scenario Variables.

Independent Scenario Variables	
Background Environments	Laboratory (L) Inside Hallway (H) Outside Campus (C)
Camera Elevation Angles	0 Degrees (Low) 25 Degrees (High)
Number of People	[0, 1, 2]
Movement Paths of People	N/A (0 People) Stationary (S) Forward (F) Lateral (L) Diagonal (D) Combination (C)

3.4 Data Collection

To thoroughly evaluate the robustness of PAVAC, we collected real-world data across different scenarios with a wide range of various factors including the environment, camera angle, number of people, and people movement path. The independent scenario variables are summarized with corresponding reference abbreviations in Table 1. We kept the rover's path the same for each scenario but made it sufficiently complex to ensure we tested multiple different vehicle actions: stationary, moving, and turning.

The background environments we selected were chosen to mimic what an autonomous system like an urban food delivery robot or rescue robot in a building might see. The laboratory presents a common indoor environment with numerous and varied objects common in any office or lab space. The inside hallway again presents another relatively controlled space with minimal background object variation. The outside campus represents our most difficult background environment due to the large amount of uncontrolled variables such as harsh sunlight, wind, and unknown people. It is also a very common location that delivery robots actually traverse.

Camera elevation angles were another variable that we realized should be a varied scenario factor since the initial straight-facing angle of the video feed is very low as the robot is likewise very low to the ground (less than a 0.3 meters in height). We were able to quickly adjust it to a higher angle (specifically 25 degrees) with the help of a camera attachment as seen in Figure 5. This also helped the video feed get a better view of more faces, our main privacy concern.

The number of people in each scenario was heavily constrained by the number of available researchers due to the lack of availability in the summer which is why were only able to test with one or two people in a scenario and used no people as a baseline scenario that checks for false positives (masking when there are no faces).

To collect data, we first ran a script which begins video feed capture from the RGB-D camera and saves both the original and anonymized video frames to the file system denoted by the date and timestamp the data capture began. Once video capture has started, we had experiment participants partake in one of five possible movement paths as shown in Figure 2: stationary (not moving), forward (moving forward towards the robot's initial movement), lateral (moving laterally towards the right hand side of the robot moving forward, diagonal (moving both forward and laterally as described in **F** and **L**, and finally combination which is a catchall for a stochastic movement path meant to push the anonymization algorithm to its limit. These movement paths were chosen to be a diverse and increasingly difficult but plausible set of actions that a typical pedestrian may take.

Also shown in Figure 2 is the rover's entire trajectory for every scenario denoted by the numbers 1-5 which was the path followed for each test. The path is described by the procedure: 1) move forward 2 ft, 2) pause for 1 second then turn clockwise 90 degrees, 3) move forward 2 ft, 4) pause for 1 second then turn counterclockwise 90 degrees, 5) move forward 2 ft. We kept this fixed for all scenarios to speed up data collection but ensured that the rover's path was complex enough so that it captures the typical actions that a robotic vehicle may perform on a mission. In particular, we made sure that that we had the rover stay stationary (baseline for camera feed with minimal vibrations), move forward (significantly increased vibrations), and turn (even more vibrations with significant scenery changes).

3.5 Data Analysis Techniques

To analyze the data, we first need the ground truth to know exactly which pixels belonging to person's face should be masked out. To figure out this ground truth, we used Meta's Segment Anything Model (SAM) which isolates pixels belonging to a particular object class and produces a corresponding mask which we can use as a baseline standard for what pixels in frame need to be anonymized [16]. SAM is perfect for our testing purposes as shown in **C** of Figure 3 because it was capable of continually updating its object segmentation throughout the course of sequential frames and it could maintain multiple segmented objects simultaneously throughout a video.

To achieve segmentation, SAM required us to manually select positive labeled pixels for each face object and and negative labeled pixels outside of the bounds of each face object. These positive labels taught the predictor model what patterns to identify the subjects' faces by while the negative labels were used to indicate similar patterns that SAM may be confused by and incorrectly segment. The model is not capable of propagating object segmentation backwards through a sequence of frames, so after selecting initial pixel label prompts, SAM will only produce accurate segmentation masks starting from that initially selected frame. Thus, comparisons between our masking model and ground truth can only begin from that frame.

We quantified the end-to-end (E2E) latency of the anonymization pipeline using Python's high-resolution timer `time.perf_counter()`. For each frame, timing was initiated at image read and terminated upon completion of anonymization. Session wall-clock time was defined as the interval from the first image read to the last anonymized frame. The effective throughput, expressed in frames per second (FPS), was calculated as

$$\text{E2E FPS} = \frac{N}{T_{\text{wall}}},$$

where N denotes the number of frames processed and T_{wall} the corresponding wall-clock duration. To characterize performance variability, latency distributions were further summarized by mean, median (p50), and upper quartiles (p90, p99).

3.6 Ethical Considerations

All test participants were members of the research team and gave their informed consent to the experiments that took place. There is minimal risk of physical harm when collecting data since the PAVAC rover is small and data is collected using a camera (no intended physical contact). Extra care was taken when collecting data outside on the public campus to ensure that only scenes with either no people or the two designated team members were saved. Data is stored locally on the Jetson when capturing video and only accessible through a wired connection or a password-protected, private SSH connection known only to research team members. Note that due to the system's low height, there was the risk of capturing personal private areas when using the high camera angle facing upwards. We thought about whether this was a ethical research topic we wanted to dive deeper into but opted out of further examination due to the lack of time and need of collecting undesirable data to make a new, specific detector for masking. All test participants are aware that undesirable but not revealing images may have been captured as part of the dataset. However, we note that this would be similar to what Apple has already does for object removal using their "Clean Up" tool [3].

4 Data and Analysis

4.1 Figures of Merit

We define several Figures of Merit (FoM) to measure the impact of our pipeline on user experience and privacy:

- **End-to-End Latency (E2E latency)**: Time from initial frame capture to the final anonymized image. A lower E2E latency suggests near-real-time performance and better user experience.
- **Frames Per Second (FPS)**: Throughput of anonymized frames. A higher FPS indicates smoother interactivity and is preferred.
- **Dice Score**:

$$\text{Dice} = \frac{2\,|A \cap B|}{|A| + |B|},$$

  measuring overlap between the predicted private-object region (A) and ground truth (B). A higher value (near 1) is better.
- **Recall**: Ratio of ground-truth private-object pixels correctly detected. A higher recall indicates fewer missed private regions. Note that the key difference between recall and dice score is that recall is a direct measure of privacy, while dice score considers both privacy and utility (i.e., overmasking increases recall but might reduce dice score). We chose to show both as they can highlight different aspects of PAVAC.

4.2 Results

The main results of our experiments are shown in Table 2 which shows the average dice and recall scores for different environments. Note that 0 faces does not have a score since SAM needs an object to segment so there is no ground truth needed as there are no pixels to mask. Our average E2E latency across all tested scenarios is approximately 41.1 ms with a 50th percentile E2E latency of 40.9 ms, 90th percentile E2E latency of 43.1 ms, and 99th percentile E2E latency of 44.6. The average FPS across all tested scenarios is approximately 24.3 frames per second.

Table 2: Overview of Dice and Recall Scores.

Aggregate Dice and Recall Scores		
Environment	**Average Dice**	**Average Recall**
Outside Campus	0.7547	0.8674
Laboratory	0.8377	0.9115
Inside Hallway	0.8964	0.9241
Overall	0.8274	0.8971

4.3 Interpretation and Discussion

The aggregate results for each environment and the overall aggregate dice and recall scores in Table 2 line up with what we expected of our implementation. Obfuscation of faces was incredibly successful as faces were kept private at the pixel-level with a very high recall score of nearly 0.9 on average. Even in the worst case when outside with harsh lighting from the sun and wind that increased the amount of perturbations in the video feed due to more vibrations, the model at worst performed 0.03 points worse than the average. In both indoor environments, the system performed above 0.9 which again indicates a high degree of success. However, this was more expected due to the lack of confusing objects in the hallway environment along good lighting and no wind in both scenes. This high of a score indicates that faces are almost always successfully detected even in edge cases (e.g. looking sideways, slightly obfuscated) which means they are also successfully anonymized. The relatively high dice score which is also only 0.07 lower on average than the recall score at 0.83 which indicates that the PAVAC system is very precise with its anonymization, rarely overmasking pixels that do not need to be hidden. This indicates that the remote user will have a video feed with useful information to perform their tasks, even if the faces are privatized. Note that the gap between dice and recall scores is higher in more difficult environments as expected. This means that while the system struggles more in harsher environmental conditions, the results are still sufficient such that a vast majority of face pixels are properly masked out of the video feed. Since we already have such a high score for privacy preservation at the pixel level, it is also likely that only pixels belonging to the edge of faces is revealed but further testing using a face detector on the anonymized images would need to be run to validate this conjecture.

Using Meta's Segment Anything Model (SAM) brought about several errors and obstacles in our pursuit of acquiring ground truth for comparison to our face detection model. The greatest recurring flaw of SAM was how all segmentation needed to begin on a single frame, even if there are multiple faces in need of segmenting. Our attempts to circumvent this flaw by marking each face on different frames would often result in a critical error preventing SAM from functioning.

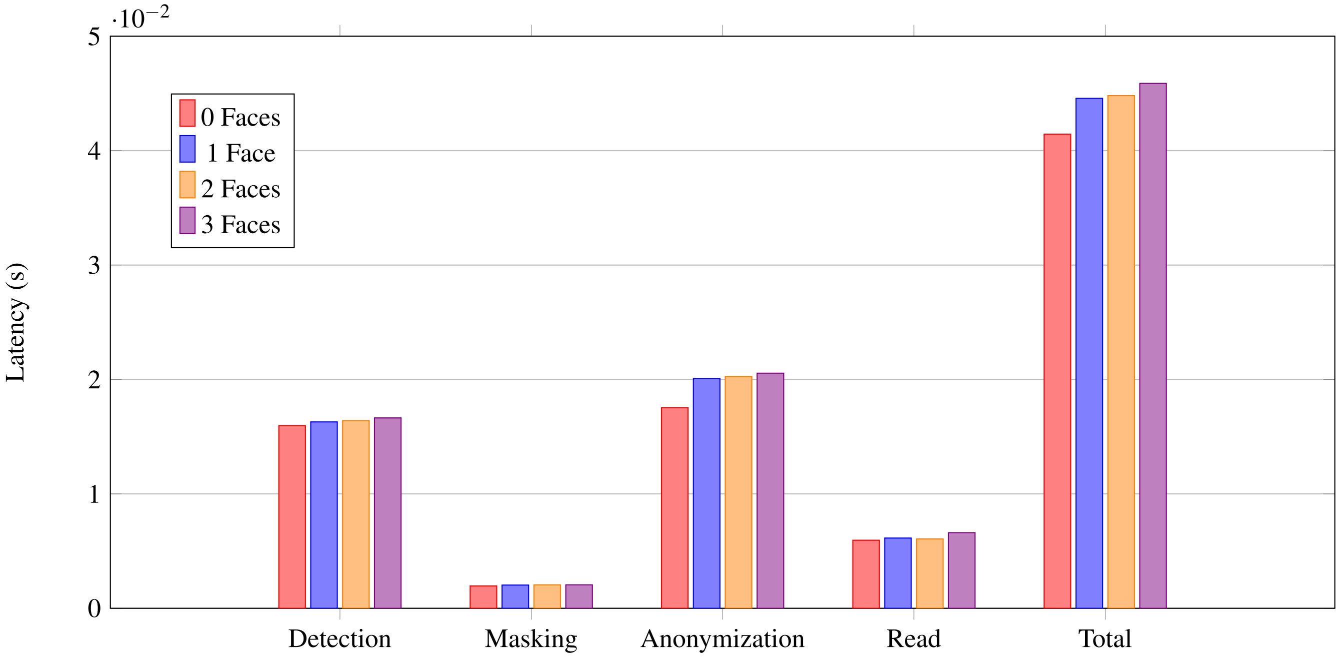

Figure 4: Latency at Different PAVAC Pipeline Stages.

This regularly led to having to wait until the video feed included all faces in the seen before being able to begin SAM segmentation, which often meant losing data due to ignoring a sizable portion of the available video when more than 1 person was involved in the scene.

A more minor issue with SAM occurred early on when we were still learning how to fully utilize SAM. The SAM model was regularly misidentifying hands and arms as faces due to having a similar skin tone as the person's face. This was partially remedied by learning to set more negative labeled points for the initial segmentation prompt in order to explicitly define arms and legs as non-examples but even with this new insight there were occasional errors in SAM's face detection. This was especially problematic in scenes recorded outdoors, as the harsh sunlight and shade from the trees often made identifying faces difficult. There were some recording sessions faces where no ones faces could be seen which again caused more segmentation issues.

As noted in the Results section, E2E latency is excellent with facial anonymization only requiring approximately 41 ms to completely pixelize faces. This latency would not even be noticed compared to the delay of a standard network connection. In Figure 4, we can see the breakdown for each part of the PAVAC pipeline and notice that detection and anonymization make up the majority of the computation time so improving these already fast times would be key to an even faster system. There is also a very minor but expected trend that increasing the number of faces in frame slows down the model, but only by at most 5 ms. Note that the scenario of 3 faces only occurred as a false positive at a rate of 0.005541% (3 out of 54,141 frames). However, we do note that the frame rate could leave more to be desired as approximately 24 FPS would be just acceptable for viewing be not great for real-time tasks that require interactivity, like controlling a vehicle. These positive results indicate a significant improvement upon prior works discussed above.

5 Concluding Thoughts

5.1 Findings

Through our data collection, we found that the overall performance of the system surpassed our expectations. As expected, we found the lowest average scores outdoors; however with an average dice score of about .7547 and recall of .8674, the algorithm is sufficiently masking the users' faces enough to prevent identification. As there were only a maximum of two faces in the frame at all times, it is difficult to determine how the amount of faces will affect the masking accuracy, however from the data it seems the masking performs better when two people are within the frame, as opposed to a single person.

5.2 Conclusions

Overall, this project validated the feasibility of a real-time anonymization algorithm that could be run entirely locally on an edge device to protect individuals' privacy while preserving data quality. The PAVAC system's computational power and efficiency were demonstrated through its ability to perform real-time masking with minimal latency. This validates that there would be virtually no impact on performance in real-world applications while reliably safeguarding privacy in autonomous systems.

5.3 Implications

In today's quickly changing world, there is often a push to release products and updates as fast as possible. Often times, this means skipping over security and privacy guarantees to release a product quicker and figure the rest out later. However, our work shows that privacy-by-design can be non-intrusively integrated into computer vision algorithms in a real-world system without heavily reducing the performance of the main task. Ultimately, this can protect the users of the product from unwanted identification and can also prevent litigation against the organization that created the product if a data breach were to occur. Our work shows that privacy can be a valuable feature of a system, not just an afterthought.

5.4 Future Work

Due to the limited time constraints imposed on this capstone project, there are many potential areas of exploration that can improve our system and produce more convincing results. First, it would be interesting to figure out what are the practical limits of PAVAC through increased real-world testing in more unique environments and particularly with a much greater volume of people. Second, we would like to include more synthetic scenarios for testing to find the theoretical limits of our system (e.g. at what point is there too many faces or movement paths become too complex). Third, we can implement more masking of other private objects in addition to faces such as license plates. Fourth, we can invite research study participants to conduct a qualitative control tests to see if controlling a rover using an anonymized video feed is any more difficult than controlling one using the original video feed. Next, we can run facial detection again on the anonymized frames to confirm at a high-level if faces are ever detected once it has passed through our system. Last, but potentially most interesting, we can develop an end-to-end system that is autonomous with an AI agent controlling the rover and making its own decisions to validate if an anonymized video feed is feasible for a fully autonomous system.

References

[1] Adeboye, O., Dargahi, T., Babaie, M., Saraee, M., and Yu, C.-M. Deepclean: A robust deep learning technique for autonomous vehicle camera data privacy. *IEEE Access 10* (2022), 124534–124544.

[2] Alharbi, R., Tolba, M., Petito, L. C., Hester, J., and Alshurafa, N. To mask or not to mask? balancing privacy with visual confirmation utility in activity-oriented wearable cameras. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 3*, 3 (2019), 1–29.

[3] Apple. Use apple intelligence in photos on iphone.

[4] Bentafat, E., Rathore, M. M., and Bakiras, S. Towards real-time privacy-preserving video surveillance. *Computer Communications 180* (2021), 97–108.

[5] Chah, B., Lombard, A., Bkakria, A., Yaich, R., Abbas-Turki, A., and Galland, S. Privacy threat analysis for connected and autonomous vehicles. *Procedia Computer Science 210* (2022), 36–44. The 13th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN) / The 12th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2022) / Affiliated Workshops.

[6] Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. In *CVPR* (2020).

[7] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. Carla: An open urban driving simulator, 2017.

[8] Fitwi, A., Chen, Y., Zhu, S., Blasch, E., and Chen, G. Privacy-preserving surveillance as an edge service based on lightweight video protection schemes using face de-identification and window masking. *Electronics 10*, 3 (2021).

[9] Hataba, M., Sherif, A., Mahmoud, M., Abdallah, M., and Alasmary, W. Security and privacy issues in autonomous vehicles: A layer-based survey. *IEEE Open Journal of the Communications Society 3* (2022), 811–829.

[10] Lee, J. H., and You, S. J. Balancing privacy and accuracy: Exploring the impact of data anonymization on deep learning models in computer vision. *IEEE Access 12* (2024), 8346–8358.

[11] Lim, H. S. M., and Taeihagh, A. Autonomous vehicles for smart and sustainable cities: An in-depth exploration of privacy and cybersecurity implications. *Energies 11*, 5 (2018).

[12] Lundström, M., and Pettersson, J. 3d privacy masking using monocular depth estimation. *Tryckeriet i E-huset* (2022).

[13] Mishra, A., Cha, J., and Kim, S. Privacy-preserved in-cabin monitoring system for autonomous vehicles. *Computational Intelligence and Neuroscience 2022*, 1 (2022), 5389359.

[14] Nordhoff, S. Resistance towards autonomous vehicles (avs). *Transportation Research Interdisciplinary Perspectives 26* (2024), 101117.

[15] Panda, S., Panaousis, E., Loukas, G., and Kentrotis, K. Privacy impact assessment of cyber attacks on connected and autonomous vehicles. In *ARES '23: Proceedings of the 18th International Conference on Availability, Reliability and Security* (New York, NY, USA, 2023), ARES '23, Association for Computing Machinery.

[16] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., and Feichtenhofer, C. Sam 2: Segment anything in images and videos, 2024.

[17] Ravi, S., Climent-Pérez, P., and Florez-Revuelta, F. A review on visual privacy preservation techniques for active and assisted living. *Multimedia Tools and Applications 83*, 5 (2024), 14715–14755.

[18] Ren, S., He, K., Girshick, R., and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[19] Rezaei, A., Sookhak, M., and Patooghy, A. Privacy-preserving in connected and autonomous vehicles through vision to text transformation, 2025.

[20] Shah, K., and Vora, D. Avprivacy, 2025.

[21] Smith, A. The human-operated sidewalk robots (with added smiles) - coco ceo, zach rash, Aug 2023.

[22] Stenger, R., Busse, S., Sander, J., Eisenbarth, T., and Fudickar, S. Evaluating the impact of face anonymization methods on computer vision tasks: A trade-off between privacy and utility. *IEEE Access 13* (2025), 11070–11079.

[23] XIAO, J., AND GOULIAS, K. G. How public interest and concerns about autonomous vehicles change over time: A study of repeated cross-sectional travel survey data of the puget sound region in the northwest united states. *Transportation Research Part C: Emerging Technologies 133* (2021), 103446.

[24] XIE, C., CAO, Z., LONG, Y., YANG, D., ZHAO, D., AND LI, B. Privacy of autonomous vehicles: Risks, protection methods, and future directions. *arXiv preprint arXiv:2209.04022* (2022).

6 Appendix

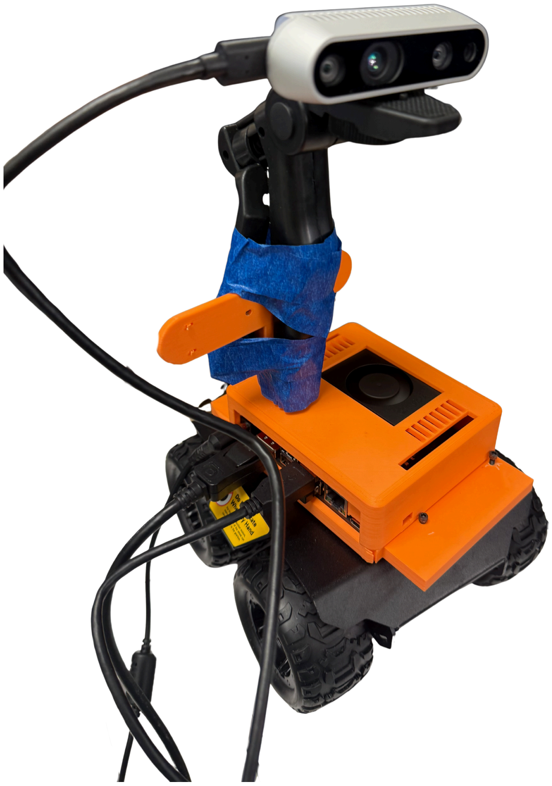

Figure 5: PAVAC System with tripod attachment that raises the camera height and allows for quick elevation angle changes.