SMARTPHONE-BASED METHOD FOR AUTOMATED SPEED ENFORCEMENT

Keya Li
Graduate Research Assistant
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
keya_li@utexas.edu

Jahnavi Malagavalli
Graduate Research Assistant
Department of Computer Science
The University of Texas at Austin
jahnavimalagavalli@utexas.edu

Lamha Goel
Graduate Research Assistant
Department of Computer Science
The University of Texas at Austin
lamhag@utexas.edu

Tong Wang
Graduate Research Assistant
Chandra Department of Electrical and Computer Engineering
The University of Texas at Austin
tong.wang@utexas.edu

Kara M. Kockelman, Ph.D., P.E. (corresponding author)
Dewitt Greer Centennial Professor in Engineering
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
kkockelm@mail.utexas.edu 512-471-0210

HIDDEN: Word Count = 4,906 words + 5 figures

Apr 30th, 2024 Submitted for 6th Bridging Transportation Researchers Conference.

ABSTRACT

Smartphone cameras and computer vision (CV) hold significant promise in assisting public agencies with enforcing traffic laws and enhancing road safety. This work designs and tests a smartphone-based method for automated speed estimation and vehicle identification (license plate, make/model, and color recognition) via an automated pipeline to assist enforcement agencies in reliably identifying speeders. The CV code accurately recognizes nearly half (46%) the license plates' text on 1,800 images from a Brazil open-source dataset, called UFPR-ALPR. Code tests on daytime recordings from hand-held smartphone videos (n=73) and roadside cameras (n = 42) in

Austin, Texas yield 60.8% accuracy for color detection (among all possible RGB color categories), 48.6% on vehicle make/manufacturer identification, and 16.89% on vehicle make and model identification. Prediction accuracy for speed estimation (within a 40% range), vehicle make (within the top 3 predictions), and license plate recognition (within the top 10 predictions) are 16.3%, 16.9%, and 29.7%, respectively. This paper also illuminates the legal, technological, and practical aspects of using smartphones for enforcement, including the potential use of private recordings for enforcement purposes, emphasizing the need to transform the potential of smartphone-based CV technologies into practical tools for vital information on traffic violations.

Keywords: smartphone cameras, computer vision, traffic safety, speed estimation, automated enforcement, vehicle identification

1. INTRODUCTION

Nearly 1.2 million people die each year globally in road traffic crashes (WHO, 2023), and almost 40,900 were killed on U.S. roadways in 2023 (NCSA, 2023). Speeding is a major contributor to U.S. crash counts and severities, with 28.7% of fatalities speeding-related (in 2021, and 29.3% in 2020) (Stewart, 2023). Active and automated enforcement of speed limits, road design strategies (like speed humps and purposeful use of red lights), speed governors on vehicles, and built-in tracking devices (like electronic logging devices) can significantly lower speeds, crash counts, and injuries (Sadeghi et al. 2016; Distefano and Leonardi, 2019; Rakesh, 2024). It is difficult and costly to enforce driving laws using police officers in real-time, and speeders may out-race police cars, sometimes delivering serious crashes (Rivara and Mack, 2004). Enforcement agencies have limited resources to allocate staff to identify violators and issue tickets on site. More and more communities are turning to automated enforcement techniques to prosecute traffic violations, but such applications remain very rare in the U.S.

Automation of fee collection is widely used in various transportation settings, including automated collection of road tolls (in the U.S., Singapore, London, Italy, and elsewhere), identification of illegally parked vehicles (in almost any developed-nation city setting) (Kashid and Pardeshi, 2014), identification of reported-as-stolen vehicles (in the U.K., U.S., China, and elsewhere) (Farr et al., 2020; Chang and Su, 2010), and enforcement of speed limits and red-light compliance (in nearly half of U.S. states, EU, China, and elsewhere) (Heiny et al., 2023; Gössel, 2015). Automated noise-limits enforcement was recently implemented in locations around Paris and New York, relying on radar detection (the radar device, composed of four microphones, measured noise levels every tenth of a second and triangulated the source of the sound) plus cameras for license plate reading (Moynihan and Esteban, 2019; NYC.gov, 2022). Presently, Russia may lead the world in speed enforcement deployments, with 18,413 speed cameras installed (Statista Research Department, 2023). As shown in Figure 1, 19 U.S. states and the District of Columbia permit the use of speed cameras (Governors Highway Safety Associate, 2024). These rely on radar waves and automatic number plate recognition (ANPR) programs for speed inference and license plate reading, either at single-camera stations (most common) or between cameras set miles apart (UK Department for Transport, 2007), and can be effective in reducing injuries and alleviating regulatory burdens. As of December 2021, thanks to NYC Automated Speed Enforcement Program, speeding at fixed camera locations had dropped, on average, 73 percent in 750 school zones on all weekdays between 6 AM and 10 PM (NYC, 2023). Seattle's Speed Safety Camera

Program reports 18% fewer pedestrian and bicyclist injury crashes at 17 camera sites/segments and 5% fewer along 100 adjacent segments (Heiny et al., 2023).

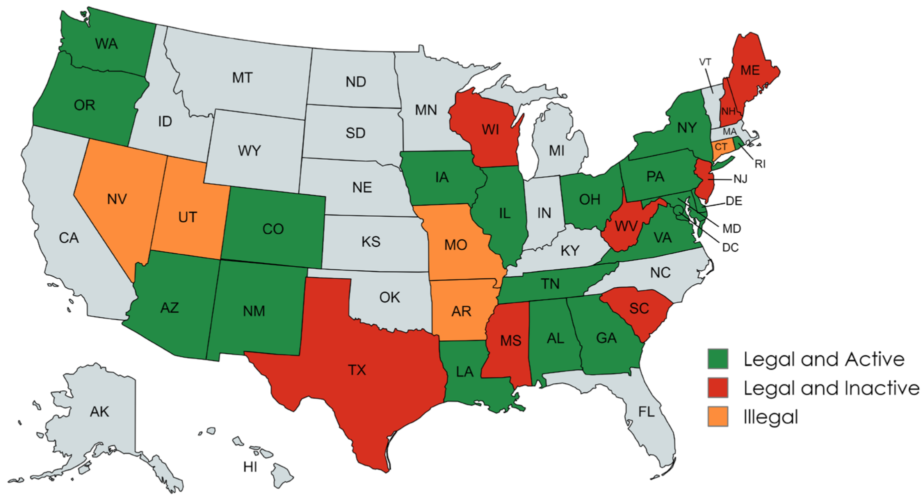

Figure 1: Speed Cameras Use across the U.S. (as of May, 2024)

Despite their road-safety and law-enforcement effectiveness, stationary cameras for reliable, automated traffic surveillance are expensive to purchase and maintain. According to New York City's Independent Budget Office (2016), speed camera costs averaged roughly $120,000 for hardware plus installation, and over $150,000 for 5 years of operation and maintenance. Given these costs, cities and states cannot afford to install stationary cameras along all road segments. Moreover, point cameras can cause drivers to slow down when being watched and then speed up downstream, a so-called 'kangaroo effect' (Chen et al., 2020) - which reduces effectiveness. Instead, allowing private citizens to share video of violations can assist in ensuring better driving at all times and in all settings. Citizens have been helping New York City officials enforce diesel-truck idling laws for several years; those submitting 3 minutes of video also receive 25 percent of any fine obtained from heavy-truck owners, which is close to $87.50 (Wilson, 2022).

This paper designs a smartphone-based method for speed enforcement, which includes automated/online speed inference and vehicle identification for automated reporting. This comprehensive framework allows for automatic detection of speeding vehicles, and can output speeds and other information (such as license plate, color, make, and model) for delivery of information to public enforcement agencies. Such practical and low-cost programs can bridge the widening violation-enforcement gap, by helping authorities identify regular offenders and take action (which may be warnings or directed conversations, ticketing and fees).

The following paper sections describe (1) prior work and existing speed inference, vehicle license plate, color, make, and model recognition techniques; (2) our methods for estimating speeds and identifying vehicles; (3) application results in Austin, Texas; (4) summary of survey findings on automated enforcement techniques in the U.S.; and (5) conclusions plus opportunities for future work.

2. SYNTHESIS OF RELATED WORK

Automated speed enforcement consists of two components: speed inference and vehicle identification. Speed inference serves as the basis for identifying potential speeders and provides estimates of speeds to determine if action needs to be taken. Once a vehicle is identified as speeding, vehicle identification is crucial for accurately and automatically identifying vehicles and extracting useful information for further use.

2.1 Speed Inference

Computer vision (CV) techniques to estimate vehicle speeds typically start by taking video recordings plus parameters (like image/size scaling factors) as inputs, and then use detection and tracking algorithms to estimate distances traveled a 2D (two-dimension/x-y) domain. Speeds are calculated with an estimated distance in the real world (calculated with a scale factor) and the time intervals between frames. Methods for distance calculation include homography-based (Kim et al., 2018), augmented intrusion line-based (Dahl and Javadi, 2019), pattern- or region-based, and image-based (Moazzam et al., 2019) techniques. Calibration plays a crucial role by helping calculate both intrinsic camera parameters (like sensor size, resolution, and focal length) and extrinsic parameters (such as location relative to the road surface). Vanishing points (VPs) (as shown in Figure 2, two VPs are accumulated separately by red and green edges) are commonly used for camera calibration and can be estimated using various algorithms, categorized into two main groups: The geometry-based methods leverage the fact that VPs occur at the intersection of straight lines. These methods estimate VPs by associating lines to VPs (Feng et al., 2010), clustering lines (Barinova et al., 2010) or searching within a Gaussian sphere (Collins and Weiss, 1990). The second methods group focuses on learning to infer VPs from large-scale datasets containing VP annotations. For example, Zhai et al. (2016) extracted global image context with a deep convolutional network to constrain the location of possible VPs while Chang et al. (2018) trained models on one million Google street-view images to detect VPs. Based on estimated VPs and assumptions that the camera is free of skew and the principal point is at the center of the frame, the camera's intrinsic and extrinsic parameters can be calculated. These parameters enable a transformation between the camera's coordinate system and the world coordinate system. However, these methods are developed for fixed traffic cameras, which need further analysis in the case for mobile cameras.

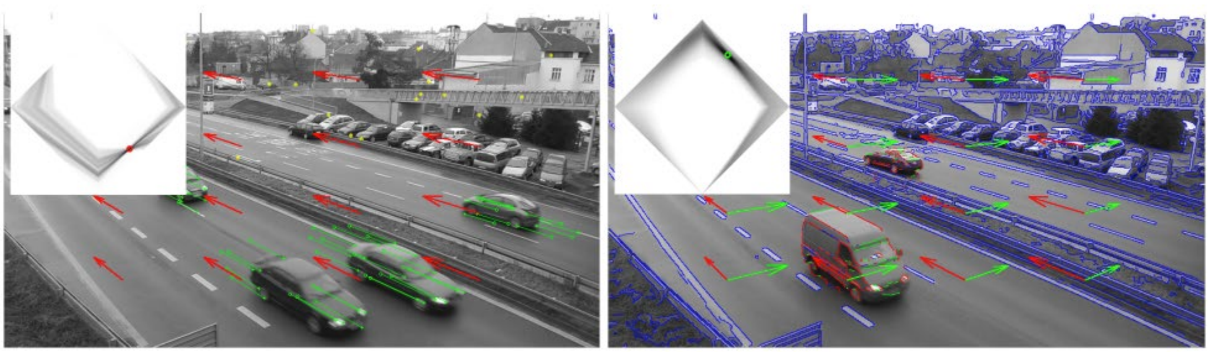

Figure 2: Vanishing Points (Source: Dubská et al., 2014)

2.2 Vehicle Identification

Vehicle detection algorithms are a type of object detection, and classified as one-stage detectors (such as You Only Look Once (YOLO) or Single Shot Detector (SSD)) or two-stage detectors (like Region with Convolutional Neural Network (R-CNN) and faster R-CNN). The latter use two neural networks to find regions of interest and classify regions, delivering better accuracy but longer processing times (Kim et al., 2020). YOLO is a popular method for efficiently detecting vehicles and traffic violations - like jumping red-light signals (Ravish et al., 2021). Wang et al. (2023) analyzed the performance of YOLOv7 in detecting objects at different frame rates and found that it outperformed two-stage detectors in terms of both time and accuracy. Meanwhile, DeepSort (Wojke et al., 2017) is often used to track vehicles by adopting two association matrices (for object velocity and appearance) to create downstream-frame boxes via Kalman filters and then predicting vehicle positions across video frames.

License plates are essential to vehicle identification. After detecting and tracking vehicles, precise license plates can guarantee delivering police tickets to specific vehicles. For instance, license plate recognition systems have been used for parking enforcement; they are installed on officer cars or at parking lot entrances and exits to scan and identify vehicles violating parking regulations. Automatic License Plate Recognition (ALPR) algorithms are the most common way to identify unique vehicles. It is a three-step process: first, the license plate is localized by either feature-based (Du et al., 2012) or deep learning-based (Laroca et al., 2019) methods, then character segmentation is done, and recognition techniques are applied to extract the text. Current techniques use separate YOLO models to extract vehicles and license plates. Text recognition on these license plates is accomplished through segmentation (a two-step process involving segmentation and a recognition model) or segmentation-free methods (a one-step process). There are several optical character recognition (OCR) techniques available (EasyOCR, 2021; Kuang et al., 2021; Pytesseract, 2022), which also pre-process images (de-skewing, smoothing edges, and converting images to black and white) to boost the chances of recognition (Karandish, 2019). ALPR algorithms are mainly hindered by poor image quality and low-resolution cameras. Much research has gone into improving image quality (Dong et al. 2015, Hamdi et al. 2021), and general adversarial networks (GANs) have proven successful in super-resolution reconstruction (Hamdi et al. 2021). While the entire pipeline used for ALPR on fixed camera videos (Silva and Jung, 2020; Zhang et al., 2021), including drone-recorded videos (Kaimkhani et al., 2022) is included in many publications, the accuracy and applicability of ALPR algorithms haven’t been validated for use with mobile phone video recordings.

License plate recognition may fail due to dark (nighttime or shade) conditions, occlusion by heavy rain or other vehicles, fake or missing plates, camera lens quality, and zoom level. In cases where a license plate is illegible, vehicle color, make, and model information can serve as alternative means to narrow the possibilities of the vehicles involved in unlawful driving situations (Lee et al., 2019). Changing a vehicle’s plate to commit crimes or avoid enforcement is relatively easy, but that is not the case for color, and especially not for make and model features. Proprietary tools are available for recognizing vehicle makes and models using traffic cameras installed, but no such system exists for general phone cameras. Conversely, open-source approaches, especially application programming interfaces (APIs), are accessible and low-priced to help institutions and communities worldwide reduce incidents of dangerous driving, death, and other losses. For example, PlateRecognizer (2024) advertises vehicle classification (including sedans, sports cars, pickup trucks, SUVS, etc.) across over 9,000 makes and models and is used in over 50 countries. RapidAPI (2024) detects vehicle color, make model, generation, and orientation for more than 3,000 models common in the U.S. In terms of color detection, Baek et al. (2007) proposed a SVM (Support Vector Machines) method for color classification, and the implementation achieved a success rate of 94.92% for 500 outdoor vehicles with five colors (black, white, red, yellow, and blue). Tilakaratna et al. (2017) employed a SVM-based method with six features and provided a wide range of 13 colors for classification. Their method performs with an accuracy of 87.52% over 2,500 images.

3. VEHICLE SPEED DETECTION AND IDENTIFICATION

This paper assumes that mobile phones are held stationary while recording videos. Since videos are analyzed frame by frame, inclination angles and phone movements can be neglected in short time intervals.

3.1 Obtaining VPs and Estimating Speeds

In this work, VPs are obtained automatically in the first frame using Lu et al.'s (2017) detection algorithm. This algorithm iteratively and randomly selects two straight line segments. It uses their intersection point as the first vanishing point (V_1) and then uniformly samples a second vanishing point V_2 on the great circle or equivalent sphere of V_1, as shown in Figure 3. Starting from each VP, tangent lines of vehicle "blobs" (a group of pixels in a frame of a video that represents a vehicle) are found, enabling construction of 3D bounding boxes (Dubská et al., 2014). Using these two VPs, two lines are extended to intersect with the points inside the frame. Four intersection points from these extended lines provide a rectangle (with lines selected to avoid including at least one VP in the rectangle). Assuming the vehicles are moving towards one of the VPs, the perspective transformation could be constructed to rectify this rectangle so only the vertical (or horizontal) movement of vehicles is preserved.

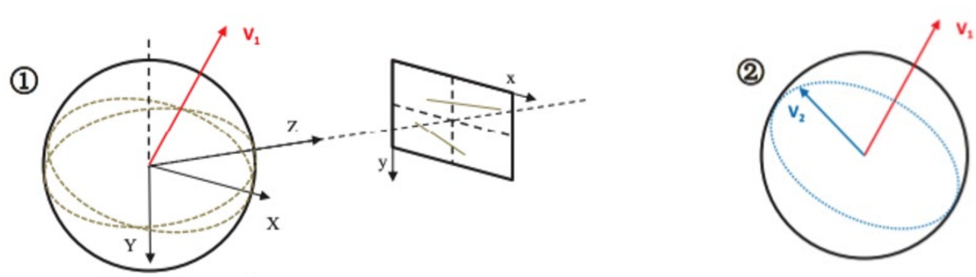

Figure 3: Procedures of generating two VPs (Source: Lu et al., 2017)

Denote two points at both ends of the 3D bounding box as $A=[a_x,a_y]^T$ and $B=[b_x,b_y]^T$ in the former frame, and $A' =[a_x',a_y']^T$ and $B' =[b_x',b_y']^T$ in the next frame. Taking vehicle length (L) as a reference (assumed here as the median lenght of U.S. passenger vehicles: 4.5 meters (Ibiknle, 2024)), the actual moving distance x would be $||A\text{-}A'||\cdot L\ /\ ||A\text{-}B||$. Vehicle speed estimate is then that distance (x) multiplied by frame rate, which is 30 frames/second (fps) for most smartphones.

3.2 Training Data for Vehicle Make and Model

Several datasets have been used to train models for automated make and model detection. For example, Yang et al.'s (2015) CompCar dataset consists of 136,727 internet vehicle images plus 44,481 surveillance-camera vehicle images across 153 car makes and 1,716 car models. Tafazzoli et al.'s (2017) Vehicle Make, Model Recognition Dataset (VMMRDb) was compiled across websites and contains 291,752 images for 9,170 distinct vehicle classes, but ended with the 2016 model year. The average lifes8 pan of U.S. passenger vehicles is roughly 16 years (Parekh and Campau, 2022), and this paper first identified the nation's 100 most popular vehicles from the 2017 National Household Travel Survey's (NHTS's) 220,430 million trip records (based on total vehicle-miles traveled by make/model). We scraped the Internet for 15,639 make/model images to use as a training dataset (alongside 300 images of those 100 most-used passenger-vehicle fronts, sides, and backs), as shown in Table 1.

Table 1: Vehicle Make and Model Training Data

Dataset	Training Data
# Images in total	15,639 web-scraped images + 300 manually-collected images
# Images for each vehicle make/model	100-200 images per make/model (including front, back, and side views in different colors and settings)
Method	Collected automatically via web scraping & combed manually to remove irrelevant images.
Example images	 7 of 174 images for Ford F-Series

3.3 Overall System Implementation

This paper relies on a series of deep-learning programs (as shown in Figure 4) for speed estimation and vehicle identification, incorporating object detection, object tracking, license plate recognition, make, model, and color detection to infer information from videos recorded via a mobile device. Vehicle bodies are first detected in each video frame using YOLOv8 code, and then tracked/connected (across frames) using DeepSort and StrongSort (Du et al., 2023). Speeds are estimated via vehicle bounding boxes and VPs (as described above, in Section 3.1). Each cropped image of the tracked vehicles is sent to a fine-tuned YOLO v7 model for license plate detection (Anpr-Org, 2023). The detected and extracted license plate images are passed to a Super-Resolution Model (by Wang et al., 2018) and an Easy-OCR (optical character recognition (EasyOCR, 2021)) model to infer and output license plate characters. The ColorDetect (2024) package and istogram and Özlü's (2018) histogram and K-nearest neighbors (KNN) techniques are then used for color inference. The KNN method compares the bounding box image to 8 base colors (white, black, red, green, blue, orange, yellow and violet) and outputs the closest color match. Meanwhile, ColorDetect compares it to all possible RGB colors and provides the fraction of color present in the vehicle bounding box.

Meanwhile, the cropped image is sent to a Resnet-50 architecture model for make/model inference (He et al., 2016). This Convolutional Neural Network (CNN) model computes the dot product between two matrices (this is accomplished by multiplying the corresponding values and adding the results to get a single scalar value in parallel (Taye, 2023)) - one representing features of images and another representing the convolutional 'kernel,' which helps preserve the spatial structures of images. It is large enough to capture variations in vehicle makes and models while also lowering computing time. In this work, the model is initially pre-trained on the VMMRDb dataset (Tafazzoli et al., 2017). Following pre-training, the last layer of the model is replaced with a fully connected layer with 100 nodes, as to detect 100 top U.S. vehicle makes and models. The training dataset collected in Section 3.2 is then used to fine-tune and re-train the last layer. Freezing the earlier layers helps the model retain its learning from the VMMRDb dataset and the amount of data used for fine-tuning is relatively small compared to the amount CNNs usually need, so only the last layer is re-trained. In addition to using the VMMRDb dataset to pre-train the model, data augmentation is employed during fine-tuning to help increase the amount of data the model detects. Four transformations are used in this process, see Table 2. These transformations also deal with real-world issues like tilted videos, blurry recordings, dark environments, bad weather conditions, etc.

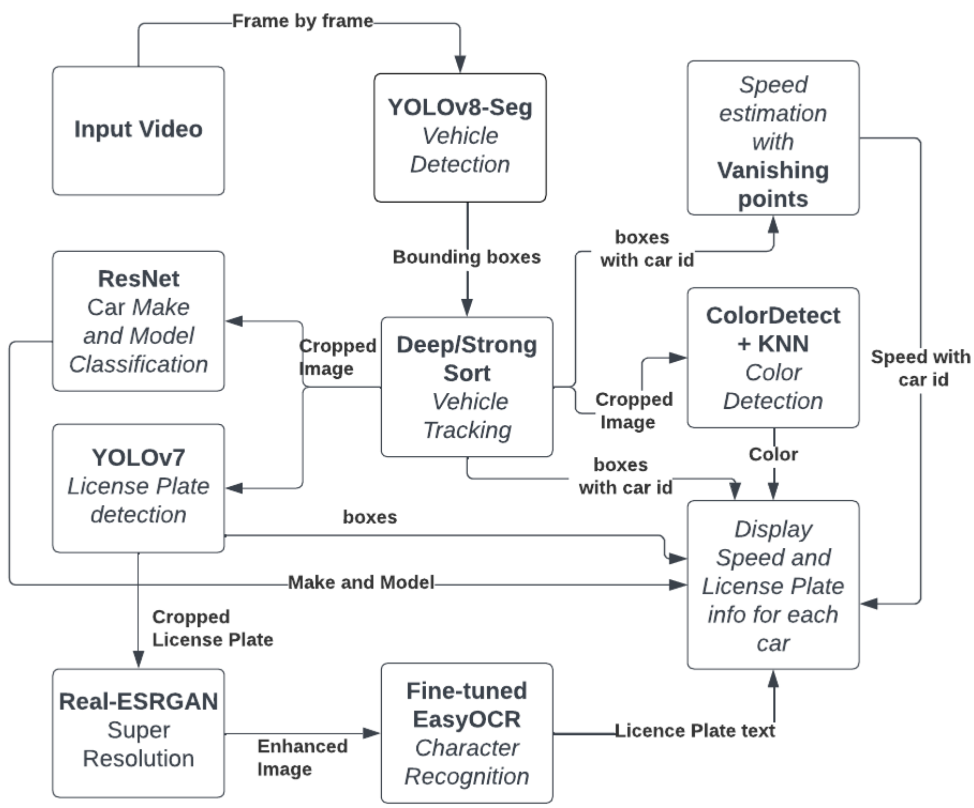

Figure 4: Flowchart of Vehicle Speed Estimation and Identification System

Table 2: Four Transformations used for Data Augmentation

Transformation	Purpose
#1: Horizontal Flip	Remove the bias towards the vehicle direction.
#2: Random Rotation	Enable the model to see vehicles from different angles.
#3: Gaussian Blur	Provide the model with different levels of blurred images.
#4: Color Jitter	Reduce the color bias by changing different aspects of the color (brightness, hue, saturation, etc.).

The crude results for each frame in the video include multiple features such as frame ID, vehicle ID, vehicle class, bounding box, color, make, model, license plate bounding box, license plate text, and their respective probabilities. To streamline the analysis, a Python script is developed to process these results and generate a consolidated output for the entire video. The processed output includes a timestamp indicating when predictions are generated, vehicle ID for tracking all vehicles in the video, the most frequent vehicle class for each vehicle ID along with the mean probability, the mean speed, the most frequent color, the top 3 prevalent color with their portion in the image, the top 3 frequent makes and models with their mean probabilities, and the top 10 frequent license plates. A sample output is displayed Table 3.

Table 3. Sample Output of Vehicle Speed Estimation and Identification System

Output Image	Feature	Output
	Timestamp	2024-03-25 12:15:38
	Vehicle ID	1
	Vehicle Class	Car
	Vehicle Class Probability (%)	91%
	Speed (mi/hr)	88.36
	Most Frequent Color	Black
	Top 3 Prevalent Colors + Shares	DarkSlateGray: 30.3% Black: 25.6% Gray: 16.5%
	Top 3 Makes & Models	Honda Accord: 20% Ford Edge: 17% Toyota Corolla: 13%
	Top 10 Plate Estimates	SLL[1], TOA, SLL##35[2], SII ##35, TCAW

Note: [1]Incomplete license plate estimates (like "SLL") are original predictions. [2] Hashtags (#) are to obscure actual values for photo anonymity. (did it read those ## values CORRECTLY? What are they?) how do those 5 plates give us your top 10 plate estimates? And what are the probabilities or % shares? Are there really 4 different SLL's and 3 SII's?

4. RESULTS

4.1 Performance of License Plate Recognition Model

The ANPR model was first tested on 1,800 images from the UFPR-ALPR dataset: a publicly available and commonly-used set of over 30,000 license plate characters from 150 vehicles (?? How can you have 30k license plates from must 150 vehicles?? captured in real-world scenarios in Brazil, where both camera (is this camera from a hand-held phone? What kind of camera? Iphone 7? Samsung XX? What kind of frame frequency? 60 Hz?) and vehicle are moving. Table 4 presents performance metrics for the YOLOv7 detection model in combination with either the Easy-OCR, Super Resolution (Real-ESRGAN) + Easy-OCR, or Super Resolution (Real-ESRGAN) + Fine-tuned Easy-OCR text recognition model. Figure 5 illustrates improvement of the Super Resolution technique. The Easy-OCR model's output is simply and 100% incorrectly 'EE', while Super Resolution predicts 'IU B6t5O62' - with XX (how many?) characters identified correctly. Key reasons for low accuracy of the license plate text recognition model are the lack of clarity of extracted images and the fact that the Easy-OCR model is not specifically trained to recognize license plate characters. To further increase accuracy, the Easy-OCR model is fine-tuned on a small subset of UFPR license plates and synthetic data, reaching up to a 47.06% accuracy rate.

Table 4: Performance of License Plate Recognition Model

Model		Model Output	Criteria	# Correct Images	Accuracy
License Plate Detection	YOLOv7	License plate bounding box.	The predicted bounding box covers more than 70% area of the true one.	1413 out of 1800	78.50%
License Plate Text Recognition	Easy-OCR	License plate characters.	The predicted license plate is the same as the true one.	252/1800	14.00%
	Super Resolution (Real-ESRGAN) + Easy-OCR			407/1800	22.61%
	Super Resolution (Real-ESRGAN) + Fine-tuned Easy-OCR			847/1800	47.06%

(a) Output of Easy-OCR (b) Output of Super Resolution + Easy-OCR

Figure 5: ANPR Improvement using Super Resolution (? These are not ANPR outputs... they are simply images. What are the alphanumeric outputs?)

4.2 Performance of Overall System

To access the overall system's performance in estimating speeds and identifying plate, make, model and color, this work collected 73 smartphone-recorded videos (4 to 5 second durations each) and 42 traffic-camera recordings (2 to 3 seconds each, at Austin intersections). These 115 videos contained reasonable imagery of 148 separate vehicles, during daytime on roadways close to (how close to? Please say "within X.X miles of") the University of Texas at Austin campus, where accurate make, model, color and plate information could be obtained by eye (human/manual review of the videos) or from images of slowed vehicles downstream at a red signal light (yes? What are the exact locations of these 115 videos?). "True" speeds for these 148 vehicles were determined using speed radar guns or image-frame-by-frame review. Manual frame review was also used to provide make, model, color, and plate number. Table 6 displays accuracies for each feature, with color identification around 60.8% accuracy. The combined color codes excelled in distinguishing gray and black vehicles (but tended to confuse other paint/body colors, perhaps because it's assuming tires & rims & trip are part of the body? What training images did you use for color training? Someone else's color code for non-vehicles/things that don't have tires, rims, & trim?). Vehicle manufacturer (model) identification was 48.6% accurate, and model was just 16.9% accurate - when using the Top 3 model estimates. Speed (within 20% of "true" speed) and license plate (within how many characters??) 16.3% and 29.7% accurate, respectively, due to factors like parked cars and handheld/moving or otherwise (?) blurry phone-camera images.

Table 6. Performance Results of Overall System

Features	Criteria	#Correct/ Sample Size	% Correct
Color	Predicted color(s) is correct. (Out of 8 colors? What does it mean to be "correct"? You like both of the first colors listed by the 2 different codes, for the paint color of the vehicles?)	90/148	60.81%
Make	Actual make is in Top 3 predictions.	72/148	48.65%
Make & Model	Predicted make + model occur among Top 3 predictions.	25/148	16.89%
Speed	Predicted speed within 20% of actual.	24/147	16.33%
License Plate	Actual license plate is among top 10 predictions.	30/101[1]	29.70%

Note: [1]License plates are unreadable in 47 testing vehicles.

5. SURVEY FINDINGS

To supplement these numeric results, an online survey was distributed to over ??? How many? US law enforcement agency officers across all 50 (? Ok, Jahnavi?) US states. The survey asked for participants' thoughts regarding 1) major challenges for automated enforcement application inside the US, 2) best applications they have seen for automated enforcement (anywhere in the world), 3) use of individuals' smartphones to assist US law enforcement practices, and 4) automated enforcement accompanied by automated ticketing of other/non-speeding behaviors (like illegal parking).

Their responses highlight the effectiveness of automated enforcement systems (in the U.S. and elsewhere), with Europe's time-over-distance (average speed) camera systems and the U.S.'s speed + red-light cameras proving effective and defensible. Table 7 shows responses relating to top challenges, with public perception, privacy, safety, and practicality listed as top concerns.

Table 7. Concerns + Challenges in U.S.-based Automated Enforcement

Area	Challenges
Public Perception	• Public may be unaware of automated enforcement's benefits. • Automated enforcement got off on the wrong foot in the US and looked too much like a money grab by local governments and the automated industry. It needs to be revenue neutral, focused on safety, with industry kept on a short leash.
Privacy and Related Topics	• Privacy concerns. (like what?? • Courtroom admissibility issues and potential bias concerns, from private citizens recording or submitting data. These 3 red items are not about automated enforcement really. They're about giving citizens a voice at the table by allowing their submiossions to carry some weight. All the other items in this table are about true AE, as deployed by public agencies. (Note: Public agencies are not going to issue tickets based on submission of cell phone videos. So we won't really have AE for that. It'll always be reviewed by a human first [after using CV to filter through it & provide estimates].) • Emergence of public vigilantes. (please explain what the concerns really is)

	• Lack of room for officer discretion. (how does this relate to privacy or safety? And why are privacy and safety together? They're very different things... are respondents worried about the safety of the person submititng the info b/c the bad drivers are going to hurt them? We can't figure this out...)
Practicality in Application	• Location (of what?) is challenged; law enforcement agencies need to involve communities in site selection and support the locations by being transparent with data. • Officers conducting speed enforcement will eventually have to testify to their training and calibration of equipment used. • Emergency vehicles should be exempt. • In most driving situations, speed naturally increases downhill and decreases uphill; decreases in congested traffic, increases in the absence of traffic, and so on. If the driver is paying too much attention to the speedometer, he may be failing to pay attention to the road ahead, causing more accidents than are prevented with speed enforcement technology. • Possible vulnerability that the system may be filled with unnecessary submissions.

Most of the XX (how many?) respondents seemed hesitant or against the use of smartphones for automated enforcement. (? Most of the items in the table are NOT about smartphone use. Please read what you've written very carefully and rewrite it so it reads 100% correct. Are they against AE (which means using agency cameras for auto-ticketing) or are they against allowing us to submit our videos for some review by officers, to maybe issue warnings or follow up with vehicle owners? They can be against one and not the other, and vice versa. Most may be against issuing tickets on the basis of private video, but most may be supportive of collecting such data...) While using computer vision with smartphone images to assist in making roadways safer, via follow-up enforcement, appears very promising and natural. (Much like anyone calling 911 or other emergency hotlines to report what they see with their own eyes, but much better – since safety officers can now review the footage themselves.) Several participants recommended working to obtain public buy-in, and making such video submissions part of a larger safety campaign, where the objective is not revenue but safety. For example, privately-provided images may simply be used to increase police patrol of certain locations at certain times of the week or year, as should be done when individuals leave messages with 911 and 311 operators in the US, every day.

As of enforcing speed compliance, most U.S. fleet owners place speed governors on heavy-duty trucks to ensure safety.

6. CONCLUSIONS AND FUTURE WORK

This study demonstrates the potential and practicality of a smartphone-based method in the context of automated speed enforcement to improve road safety. The license plate number recognition model detected 78.5% of license plates and then accurately recognized 47.1% of license plates' text when tested on 1,800 images from Brazil's UFPR-ALPR dataset. The entire system achieves 16.33% accuracy in estimating speeds, with errors staying with 20% range, and 29.70% in recognizing license plates. As a supplement to identifying vehicles, it can reach up to 60.81% and 48.65% correctness in detecting vehicle colors and makes. This research aims to envision further individual engagement in regulating traffic laws and the autonomous technologies involvement in this process. It is evident that these technologies can play a pivotal role in enhancing road safety and traffic management, and additional research will be key to realizing these goals. This work also investigates the common challenges of automated enforcement and future huddles and recommendations of the practical use of private recordings for enforcement purposes.

To improve speed estimation accuracy, using the specific length of each vehicle (by make/model) should be used, instead of a single average or median assumption as currently used (especially for very long or unusually short vehicles). The model for color detection can be modified to focus on specific parts of the vehicle, such as the hood and trunk, rather than considering the entire image within the bounding box. And the entire system can be more accurate by training the model with moving camera data, collecting and labeling more data, and working to eliminate noise from nearby vehicles. Identification of a vehicle's make, model, year, and color will prove useful when license plates are obstructed or missing (or falsified), increasing the likelihood of successful law enforcement for safer roadways. Mobile camera properties, like aperture size and shutter speed, can be experimented with to improve video recordings without motion blur.

Directions for future research include extending the analysis to more complex scenarios such as nighttime videos (in lighted and unlighted settings) when speed and plate inference will probably prove more difficult and with moving cameras (as is common with hand-held devices and/or when inside nearby vehicles). Another extension is developing a mobile smartphone application for regular or automated submission of flagged video segments with precise position/location details (during actual recording rather than user-estimated values). The scalability of the presented idea has to be explored to see how it will perform for a dense observation environment, such as, expressways and city center.

Another endeavor is building comprehensive maps for relevant enforcement agency response. Encouraging enforcement agencies to adopt private-phone video for enforcement support may be challenging due to data privacy concerns - and the potential for fake video submissions. But automakers like GM are already surveiling and sharing such driving behavior with insurance companies. Currently, many US states do not allow the use of traffic cameras or speed cameras for law enforcement purposes, but other nations rely heavily on and benefit greatly (in safety, effort and cost) from automated enforcement. From a system implementation standpoint, an end-to-end system that can optimize the current system is preferred. Automating the entire system decreases human involvement and manual costs.

ACKNOWLEDGEMENTS

We are grateful to UT Austin's Good Systems research program for sponsorship, Eva Natinsky and Samridhi Roshan for helping generate license plate results, and undergraduates Connor Moening and Austin Turgeon for trimming and labeling smartphone videos. We thank undergraduate Eric McDonnell for his early review of state laws, Tong Wang and Xuxi Chen for help with speed estimation model, Aditi Bhaskar and Lucas Allison for their valuable edits, and Sheila Lalwani and Dr. Mike Murphy for information on courtroom admissibility of phone-recorded videos.

REFERENCES

Anpr-Org (Accessed March 2024). GitHub - ANPR-ORG/ANPR-using-YOLOV7-EasyOCR. https://github.com/ANPR-ORG/ANPR-using-YOLOV7-EasyOCR

Baek, N., Park, S. M., Kim, K. J., & Park, S. B. (2007). Vehicle Color Classification Based on the Support Vector Machine Method. In *Advanced Intelligent Computing Theories and Applications. With Aspects of Contemporary Intelligent Computing Techniques: Third International Conference on Intelligent Computing, ICIC 2007, Qingdao, China, August 21-24, 2007. Proceedings 3* (pp. 1133-1139). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74282-1_127

Barinova, O., Lempitsky, V., Tretiak, E., & Kohli, P. (2010). Geometric Image Parsing in Man-made Environments. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part II 11* (pp. 57-70). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15552-9_5

Chang, C. K., Zhao, J., & Itti, L. (2018, May). Deepvp: Deep Learning for Vanishing Point Detection on 1 Million Street View Images. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4496-4503). IEEE. https://doi.org/10.1109/ICRA.2018.8460499

Chang, W., & Su, C. (2010, June). Design and Deployment of A National Detecting Stolen Vehicles Network System. In *Pacific-Asia Workshop on Intelligence and Security Informatics* (pp. 22-27). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13601-6_3

Chen, T., Sze, N. N., Saxena, S., Pinjari, A. R., Bhat, C. R., & Bai, L. (2020). Evaluation of Penalty and Enforcement Strategies to Combat Speeding Offences among Professional Drivers: A Hong Kong Stated Preference Experiment. *Accident Analysis & Prevention*, *135*, 105366. https://doi.org/10.1016/j.aap.2019.105366

Collins, R. T., & Weiss, R. S. (1990, December). Vanishing Point Calculation as A Statistical Inference on the Unit Sphere. In *ICCV* (Vol. 90, pp. 400-403). https://doi.org/10.1109/ICCV.1990.139560

ColorDetect. (Accessed March 2024). ColorDetect. https://colordetect.readthedocs.io/en/latest/

Dahl, M., & Javadi, S. (2019). Analytical Modeling for A Video-based Vehicle Speed Measurement Framework. *Sensors*, *20*(1), 160. https://doi.org/10.3390/s20010160

Distefano, N., & Leonardi, S. (2019). Evaluation of The Benefits of Traffic Calming on Vehicle Speed Reduction. *Civil Engineering and Architecture* 7(4), 200-214. https://www.hrpub.org/journals/article_info.php?aid=8083

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image Super-resolution Using Deep Convolutional Networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295-307. https://doi.org/10.1109/TPAMI.2015.2439281

Du, S., Ibrahim, M., Shehata, M., & Badawy, W. (2012). Automatic License Plate Recognition (ALPR): A State-of-the-art Review. *IEEE Transactions on circuits and systems for video technology*, *23*(2), 311-325. https://doi.org/10.1109/TCSVT.2012.2203741

Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H. (2023). Strongsort: Make Deepsort Great Again. *IEEE Transactions on Multimedia*. https://doi.org/10.1109/TMM.2023.3240881

Dubská, M., Herout, A., & Sochor, J. (2014, September). Automatic Camera Calibration for Traffic Understanding. In *BMVC* (Vol. 4, No. 6, p. 8). https://www.fit.vutbr.cz/~herout/papers/2014-BMVC-VehicleBoxes.pdf

EasyOCR. (Accessed March 2024). Jaided AI. https://www.jaided.ai/easyocr

Feng, C., Deng, F., & Kamat, V. R. (2010). Semi-automatic 3d Reconstruction of Piecewise Planar Building Models from Single Image. *CONVR (Sendai:)*, *2*(5), 6. https://doi.org/10.3390/s20010160

Fernandez Llorca, D., Hernandez Martinez, A., & Garcia Daza, I. (2021). Vision-based Vehicle Speed Estimation: A Survey. *IET Intelligent Transport Systems*, *15*(8), 987-1005. https://doi.org/10.1049/itr2.12079

Gössel, B. (2015). *Cross-border Traffic Police Enforcement: A Descriptive and Explanatory Cross-sectional Study on the Role of the EU's Fight Against the Three Main Killers' on EU Roads in the Joint Control Operations of the Police Forces of Lower Saxony (GER) and Oost-Nederland (NL)* (Bachelor's thesis, University of Twente). https://purl.utwente.nl/essays/67158

Governors Highway Safety Associate. (Accessed March 2024). *Speed and Red Light Cameras*. https://www.ghsa.org/state-laws/issues/speed%20and%20red%20light%20cameras

Hamdi, A., Chan, Y. K., & Koo, V. C. (2021). A New Image Enhancement and Super Resolution technique for License Plate Recognition. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08341

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Heiny, S., Lan, B., Vann, M., Sleeman, E., Proulx, F., Hintze, M., Thomas, L., & Carter, D. (2023). *Exploring the Impact of Select Speed-reducing Countermeasures on Pedestrian and Bicyclist Safety* (Report No. DOT HS 813 446). National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/67640

Hill Kashmir. (2023). Automakers Are Sharing Consumers' Driving Behavior With Insurance Companies. The New York Times. https://www.nytimes.com/2024/03/11/technology/carmakers-driver-tracking-insurance.html

Ibiknle, D. (2024). Average Car Sizes: Length, Width, and Height. *Neighbour Blog*. https://www.neighbor.com/storage-blog/average-car-sizes-dimensions

Kaimkhani, N. A. K., Noman, M., Rahim, S., & Liaqat, H. B. (2022). UAV with Vision to Recognise Vehicle Number Plates. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/7655833

Karandish, F. (2019). The Comprehensive Guide to Optical Character Recognition (OCR), Moov AI. https://moov.ai/en/blog/optical-character-recognition-ocr/

Kashid, S. G., & Pardeshi, S. A. (2014, April). Detection and Identification of Illegally Parked Vehicles at No Parking Area. In *2014 International Conference on Communication and Signal Processing* (pp. 1025-1029). IEEE. http://doi.org/10.1109/ICCSP.2014.6950002

Kim, J. A., Sung, J. Y., & Park, S. H. (2020, November). Comparison of Faster-RCNN, YOLO, and SSD for Real-time Vehicle Type Recognition. In *2020 IEEE international conference on consumer electronics-Asia (ICCE-Asia)* (pp. 1-4). IEEE. https://doi.org/10.1109/icce-asia49877.2020.9277040

Kim, J. H., Oh, W. T., Choi, J. H., & Park, J. C. (2018). Reliability Verification of Vehicle Speed Estimate Method in Forensic Videos. *Forensic science international*, *287*, 195-206. https://doi.org/10.1016/j.forsciint.2018.04.002

Kuang, Z., Sun, H., Li, Z., Yue, X., Lin, T. H., Chen, J., ... & Lin, D. (2021, October). MMOCR: A Comprehensive Toolbox for Text Detection, Recognition and Understanding. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 3791-3794). https://doi.org/10.1145/3474085.3478328

Laroca, R., Zanlorensi, L. A., Gonçalves, G. R., Todt, E., Schwartz, W. R., & Menotti, D. (2021). An Efficient and Layout-independent Automatic License Plate Recognition System Based on the YOLO Detector. *IET Intelligent Transport Systems*, *15*(4), 483-503. https://doi.org/10.1049/itr2.12030

Lee, H. J., Ullah, I., Wan, W., Gao, Y., & Fang, Z. (2019). Real-time Vehicle Make and Model Recognition with the Residual SqueezeNet Architecture. Sensors, 19(5), 982. https://doi.org/10.3390/s19050982

Lu, X., Yaoy, J., Li, H., Liu, Y., & Zhang, X. (2017, March). 2-line Exhaustive Searching for Real-time Vanishing Point Estimation in Manhattan World. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 345-353). IEEE. https://doi.org/10.1109/WACV.2017.45

Moazzam, M. G., Haque, M. R., & Uddin, M. S. (2019). Image-based Vehicle Speed Estimation. *Journal of Computer and Communications*, *7*(6), 1-5. https://doi.org/10.4236/jcc.2019.76001

Moynihan, Q., & Esteban, C.F. (2019, September). *Major Cities are Introducing Noise Radars that Automatically Issue Fines to Loud Vehicles to Combat Noise Pollution*. Business Insider. https://www.businessinsider.com/major-cities-introducing-noise-radars-to-fine-loud-vehicles-2019-9nr_email_referer=1&utm_source=Sailthru&utm_medium=email&utm_content=Tech_select

National Center for Statistics and Analysis. (2023). *Early Estimate of Motor Vehicle Traffic Fatalities for the first half of 2023* Report No. DOT HS 813 514). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813561

Stewart, T. (2023). *Overview of Motor Vehicle Traffic Crashes in 2021* (Report No. DOT HS 813 435). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813435

New York City. (2023). New York City Automated Speed Enforcement Program, 2022 Report. https://www.nyc.gov/html/dot/downloads/pdf/speed-camera-report.pdf

New York City Independent Budget Office. (2016, March). Transportation Funds Added for Vision Zero, Traffic Enforcement Cameras. https://www.ibo.nyc.ny.us/iboreports/transportation-funds-added-for-vision-zero-traffic-enforcement-cameras.pdf

NYC.gov. (2022, February). Roadside Sound Meter and Camera that is Activated by Loud Mufflers Now Sending Notices. https://www.nyc.gov/site/dep/news/22-005/roadside-sound-meter-camera-is-activated-loud-mufflers-now-sending-notices-vehicle

Özlü. A. (2018). Color Recognition. https://github.com/ahmetozlu/color_recognition

Parekh, N., & Campau, T. (2022, March). Average Age of Vehicles in the US Increases to 12.2 years, According to S&P Global Mobility. *S&P Global Mobility*. https://www.spglobal.com/mobility/en/research-analysis/average-age-of-vehicles-in-the-us-increases-to-122-years.html

PlateRecognizer. (Accessed March 2024). Plate Recognizer, https://platerecognizer.com/

Pytesseract. (Accessed March 2024). PyPI. https://pypi.org/project/pytesseract/

Rakesh Patel. (February 2024). What is Driver Tracking & How Does it Help Your Business? Upper. https://www.upperinc.com/blog/driver-tracking/

RapidAPI. (Accessed March 2024). Rapid API, https://rapidapi.com/dominonet-lTpEE6zONeS/api/vehicle-make-and-model-recognition.

Ravish, R., Rangaswamy, S., & Char, K. (2021, October). Intelligent Traffic Violation Detection. In *2021 2nd Global Conference for Advancement in Technology (GCAT)* (pp. 1-7). IEEE. https://doi.org/10.1109/gcat52182.2021.9587520

Rivara, F. P., & Mack, C. D. (2004). Motor Vehicle Crash Deaths Related to Police Pursuits in the United States. *Injury Prevention*, *10*(2), 93-95. https://doi.org/10.1136/ip.2003.004853

Sadeghi-Bazargani, H., & Saadati, M. (2016). Speed Management Strategies: a Systematic Review. *Bulletin of Emergency & Trauma*, *4*(3), 126. https://www.researchgate.net/publication/304396305_Speed_Management_Strategies_A_Systematic_Review

Silva, S. M., & Jung, C. R. (2020). Real-time License Plate Detection and Recognition Using Deep Convolutional Neural Networks. *Journal of Visual Communication and Image Representation*, *71*, 102773. https://doi.org/10.1016/j.jvcir.2020.102773

Statista Research Department. (2023). *European Countries by Number of Permanently Installed Traffic Speed Cameras 2023*. https://www.statista.com/statistics/1201490/speed-cameras-permanent-traffic-countries-europe

Tafazzoli, F., Frigui, H., & Nishiyama, K. (2017). A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1-8). https://doi.org/10.1109/CVPRW.2017.121

Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future directions. *Computation*, *11*(3), 52. https://doi.org/10.3390/computation11030052

Tilakaratna, D. S., Watchareeruetai, U., Siddhichai, S., & Natcharapinchai, N. (2017, March). Image Analysis Algorithms for Vehicle Color Recognition. In 2017 International Electrical Engineering Congress (iEECON) (pp. 1-4). IEEE. https://doi.org/10.1109/IEECON.2017.8075881

UK Department for Transport. (2007). Using Speed and Red-light Cameras for Traffic Enforcement: Deployment, Visibility and Signing. https://assets.publishing.service.gov.uk/media/5a819278e5274a2e87dbe588/dft-circular-0107.pdf

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464-7475). https://doi.org/10.48550/arxiv.2207.02696

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced Super-resolution Generative Adversarial Networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0). https://doi.org/10.1007/978-3-030-11021-5_5

WHO (2023). *Road Traffic Injuries*. World Health Organization. Available at https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple Online and Realtime Tracking with A Deep Association Metric. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3645-3649). IEEE. https://doi.org/10.1109/icip.2017.8296962

Yang, L., Luo, P., Change Loy, C., & Tang, X. (2015). A Large-scale Car Dataset for Fine-grained Categorization and Verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3973-3981). https://doi.org/10.1109/CVPR.2015.7299023

Zhai, M., Workman, S., & Jacobs, N. (2016). Detecting Vanishing Points Using Global Image Context in A Non-manhattan World. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5657-5665). https://doi.org/10.48550/arXiv.1608.05684

Zhang, C., Wang, Q., & Li, X. (2021). V-LPDR: Towards A Unified Framework for License Plate Detection, Tracking, and Recognition in Real-world Traffic Videos. *Neurocomputing*, *449*, 189-206. https://doi.org/10.1016/j.neucom.2021.03.103

HIDDEN EXTRA CONTENT HERE

Average speeds on roadways increased during COVID, as drivers appeared to drive more recklessly because of lighter traffic and higher speed limit allowances in many places (Tucker and Marsh, 2021; Aashto 2022, Speed Camera Program 2023).

To get around this issue, other systems being used in places such as the UK or Australia are designed with multiple cameras to operate over a larger stretch of road, calculating the average speed of the driver (Delaney et al. 2005; Owen et al. 2016; Bates et al. 2016). However, such systems run into issues where the exact or momentary speed of a vehicle is unknown.

Another technology employed to restrict speed is speed limiters or governors, which are widely used to impose maximum speed limitations on trucks in various countries. For example, Canada's speed governors limit trucks to 65 mph, while in Australia, trucks are restricted to 65 mph or less depending on the region (Canada 2013; Buchs 2022). The USA is also planning to require speed governors for trucks, with a proposed limit of 70 or 75 mph (FreightWaves, n.d.).

For instance, in a Texas city, a single police officer scanned 48,101 plates using such a system, resulting in 255 traffic citations, 26 suspended licenses, and other violations in the course of 96 hours of use over 27 days (Wood, 2021).

In addition, various tools can be used to identify high-emission vehicles, including remote sensing devices that detect pollutants in the exhaust using infrared and ultraviolet technology, on-board diagnostics that monitor engine performance and emissions control systems, and video analytics that analyze factors like vehicle speed (Huang et al. 2018; Oluwaseyi and Sunday 2020; Valido et al. 2022).

In the safety realm, many nations now subscribe to a Vision Zero plan, whose goal is zero road deaths (and near-zero debilitating injuries) (Tingvall et al. 1999; Marusin et al. 2018). Avoiding excessive speed and other illegal driving maneuvers (like left turns on red, wrong-way travel and rapid lane changes in congested traffic) are one of the best ways to lower severe-crash counts, and could save trillions of dollars a year around the globe (with US crashes alone currently costing nearly $1 trillion annually, or roughly $3,000 per capita per year (Liu and Subramanian 2009). In Asia, at least 8 countries rely heavily on automated enforcement predominantly to enforce speed limits, with another 10 Asian nations using a mix of automated and manual enforcement (UN.ESCAP, 2020). In Hong Kong, speed cameras are used in conjunction with manual policing.

As of February 2023, 22 U.S. states, plus Washington D.C. and at least 35 Texas communities with preexisting contracts (The Texas Tribune 2019, KVUE 2022) (Figure 1a) use red-light cameras.

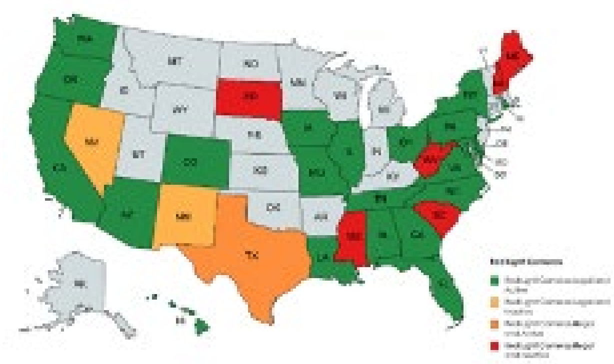

Figure 1a: Red Light Camera Use in US States

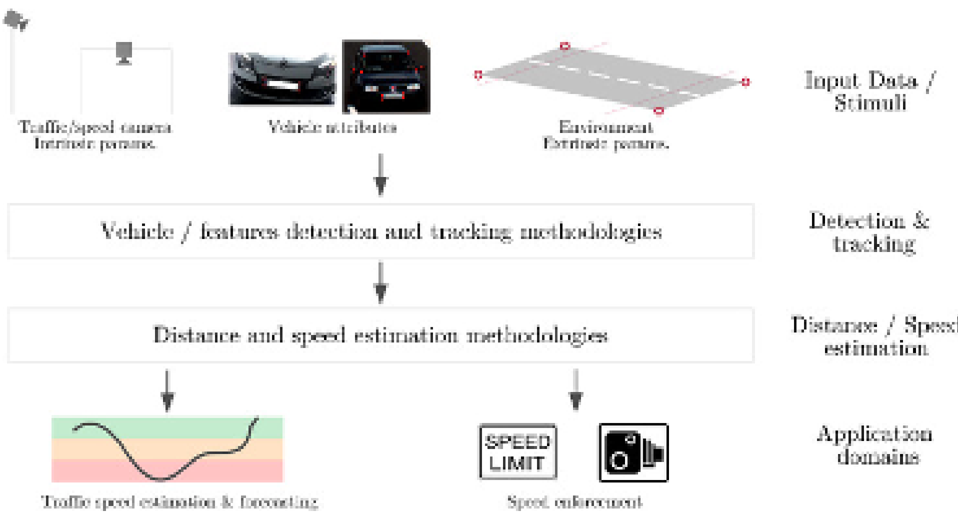

Figure 3. High-level components of vision-based speed estimation (Source: Llorca et al. 2021)

Vehicle or object-tracking algorithms use deep learning (a series of neural networks) to predict object positions across video frames using spatial and temporal features. Tracking tools generate bounding boxes to improve object detection and identification.

Hu et al. (2023) proposed a Smooth Modulation Neural Network with Multi-Scale Feature Fusion (SMNN-MSFF) model with multiscale feature fusion and smooth modulation. Their experimental results demonstrate that the mean average precision of the SMNN-MSFF model is 94.96% in recognizing 24 colors and 97.25% in recognizing 8 colors.

Aarathi and Abraham (2017) adopted a preprocessing step to remove haze from images and then used CNN to extract features from vehicles, passing them to a SVM for color classification. Wang et al. (2023) presented a vehicle color recognition method using GoogLeNet and Inception v1, which can recognize eight vehicle colors with 90%-95% accuracy.

This paper adopts two different modes to obtain the VPs. The first relies on the users to provide the estimated locations of VPs. The users are instructed to infer the approximate locations of VPs by themselves. Generally, the first VP can be located by finding the intersection of the left-side and the right-side of the road in the video, and the second VP can be approximately located by extending the edges of the vehicles.

Krause et al. (2013) introduced a dataset of 207 fine-grained categories, consisting of a small, ultra-fine-grained set of 10 BMW models (512 images) and a larger set of 197 car types (16,185 images). The images in the larger set have a resolution of 360 x 240 pixels and are divided into a 50-50 train/test split, with 8,144 train images and 8,041 test images.

Ali et al., 2022 presented 3,847 images of 48 U.S. models from high resolution of 1920 x 1080 pixels videos collected from camera units installed on a highway in Karachi, Pakistan between 10 am and 7 pm. The videos involve different viewpoints and lighting conditions with variable frame rates, which are able to reflect real lighting conditions for moving vehicles.

Speed estimation on public datasets

This paper's speed estimation algorithm was first tested on a public dataset called VS13 (Djukanović et al. 2022). It contains video recordings of 13 different car models (i.e., the Citroen C4 Picasso, Kia Sportage, Mazda 3 Skyactive, Mercedes AMG 550, Mercedes GLA 200D, Nissan Qashqai, Opel Insignia, Peugeot 208, Peugeot 3008, Peugeot 307, Renault Captur, Renault Scenic, and VW Passat B7) at different speeds. Each video is captured in full HD, 10 seconds long, and 24 frames per second. The ground-truth speeds of vehicles are provided in this dataset, which ranges from 30 to 105 kilometers per hour. Since these videos were captured from similar perspectives, the Mazda was arbitrarily selected to assess the performance of the speed estimation algorithm. The 13 vehicles' speeds were estimated between the 100th to 160th video frames for user-estimated VP and 100th to 130th frames for algorithm-estimated VP. The median speed estimates are used for error calculations, as shown in Table 3.

Table 3. CV- and user-estimated speeds versus ground-truth speeds, with associated errors.

(1) CV-estimated Speed (km/h)	**(2) User-estimated Speed** (user-input VPs) (km/h)	**Actual (Ground-truth) Speed** (km/h)	**Errors in Estimated Speeds** (Estimated minus Actual) (km/h)	**% Relative Error** (Automatically Estimated VPs/User-input VPs)
29.30	28.04	30	-0.70 & -1.96	-2.3 & -6.5
39.28	42.30	40	-0.72 & 2.30	-1.8 & 5.8
50.11	56.16	50	0.11 & 6.16	0.22 & 12.3
73.87	81.47	60	13.87 & 21.47	23.1 & 35.8
22.07	77.97	70	-47.93 & 7.97	-68.5 & 11.4
81.47	85.32	81	-0.47 & 4.32	-0.58 & 5.3
86.33	100.30	90	-3.67 & -10.30	-4.1 & -11.4

Table 3 showcases the estimated speed of vehicles using both automatically estimated VPs and user-input VPs. It can be observed that automatically estimated VPs give better predictions on speed except for one case. The absolute errors are smaller than 5 kilometers per hour on five out of seven cases that have been tested. The user-input VPs can also provide an accurate estimation of speeds, despite the fact that the errors are slightly larger compared to using the automatically estimated VPs on several cases. However, the VPs specified by users would still be valuable as they can serve as the backup solutions for speed inference once the VP estimation algorithm cannot provide reliable results.

The license plate and vehicle make/model algorithms were tested on 2-3 videos collected using each smartphone type from Table 2. Analyzing algorithm outputs for speeds under 40 mph, the ALPR algorithm excelled when vehicles in the nearest lane were observed, with performance decreasing as distance increased. In Figure 6, the ALPR predicts the license plate of the closest lane vehicle at 30-40 mph, with a detection confidence of 0.96. The text ‘TEXAS’ is predicted with a confidence score of 0.61, while the number is predicted as ‘RL* 23**’ with a confidence score of 0.28. The vehicle is identified as a Mini Cooper 2009.

Table 2. Cell Phone Resolution Examples (Source: GSMArena.com)

Phone Type	Camera Resolution	Video Resolution
iPhone 14 Pro	48 megapixels (MP)	1080 pixels (p) width at 30 frames per second (fps) Full HD (high-density)
Samsung M33	50 MP	2160 p @ 30 fps Ultra HD
Oneplus Nord2	50 MP	1080p@30fps Full HD

The ALPR model did not provide estimates for speeds between 40 and 60 mph due to low-quality plate images (Figure 7). Although the model detected a vehicle's license plate, no text prediction was possible. The predicted make and model, the GMC Terrain SUV, differed from the actual Toyota Sienna. The vehicle make and model algorithm accurately identified makes but struggled with model types. Make estimates were occasionally inconsistent, as shown in Figures 6 and 7.

Figure 6. Processed output: Vehicle at 30-40 mph on I35 frontage road in Austin

Figure 7. Processed output: Vehicle at 50-60 mph on IH-35 in Austin, Texas

Warnings or tickets would be issued after formal review by deputized officers only when there is clear and compelling evidence. Data storage security is also key, similar to existing expectations of law enforcement agencies. Those who receive tickets have the right to challenge the process, mainly if the images appear fuzzy or unclear, ensuring the protection of due process. Due process is a fundamental concept in American law that refers to the idea that individuals are entitled to fair treatment and legal protections when facing government actions that may adversely affect their life, liberty, or property.

Michael Wilson. (2022, March). *$87.50 for 3 Minutes: Inside the Hot Market for Videos of Idling Trucks*. The New York Times. https://www.nytimes.com/2022/03/19/nyregion/clean-air-idle-car.html

ASURA. (Accessed March, 2024). *Asura Recognition Unit Make and Model Recognition (ARU MMR)*. https://asuratechnologies.com/vehicle-make-and-model-recognition/.

Malagavalli. J., Wang. T., Kockelman, K.M., et al. (2023). Catching speeders via mobile phones and machine learning: An opportunity to improve speed enforcement. https://www.caee.utexas.edu/prof/kockelman/public_html/TRB24CVforSpeedEnforcement.pdf

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 554-561). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6755945

Rachmadi, R. F., & Purnama, I. (2015). Vehicle color recognition using convolutional neural network. *arXiv preprint arXiv:1510.07391*. https://doi.org/10.48550/arXiv.1510.07391

Chen, Z., Pears, N., Freeman, M., & Austin, J. (2014). A Gaussian mixture model and support vector machine approach to vehicle type and colour classification. *IET Intelligent Transport Systems*, *8*(2), 135-144. https://doi.org/10.1049/iet-its.2012.0104

Aarathi, K. S., & Abraham, A. (2017, March). Vehicle color recognition using deep learning for hazy images. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 335-339). IEEE. https://doi.org/10.1109/ICICCT.2017.7975215

Wang, J., Zhang, X., Yang, X., & Zhao, Z. (2023, January). Study on vehicle color detection and recognition based on deep learning. In *Optoelectronic Imaging and Multimedia Technology IX* (Vol. 12317, pp. 166-172). SPIE. https://doi.org/10.1117/12.2642497

Hu, M., Bai, L., Fan, J. *et al.* Vehicle color recognition based on smooth modulation neural network with multi-scale feature fusion. *Front. Comput. Sci.* **17**, 173321 (2023). https://doi.org/10.1007/s11704-022-1389-x

Yicong, J.O. (Accessed March 2024). Google Image Scraper. https://github.com/ohyicong/Google-Image-Scraper.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*. https://doi.org/10.1145/3065386

Jocher, G., Chaurasia, A. and Qiu, J. (Accessed March 2024). YOLO by Ultralytics. https://github.com/ultralytics/ultralytics

Keya wrote: yolov7 is for license plate detection, and yolov8 is for object detection. The newest progress Jahnavi made is to use yolov8 to lower system complexity for object detection.

Histogram technique is one of the color classification approaches, it has an open-source library.

I believe that if we can import enough images (about the vehicle make and model) to train the model, it can be added to the top 100 list soon.

Gainsboro is a color category (which is gray).

Conf means confidence.

HIDE: Additionally, a web-based user-interface system is developed and investigated to form a comprehensive traffic reporting mechanism, alleviating enforcement burdens.

HIDE It also presents the initiation of a comprehensive user-interface traffic violation reporting system by combining U.S. jurisdiction maps and OpenStreetMap speed limits.

HIDE 3.4 Web-based User-Interface for Video Submission

In this paper, a web-based user-interface system is developed to leverage this automated system. It allows users to upload video clips they suspect involve serious speed violations, provide location information, and estimate relevant details such as vehicle make, model, speed, color, or license plate. The clips are transferred from the cloud to the server periodically, undergo the flowchart process in Figure 2, and then are pushed back to the cloud before being displayed on the website. Meanwhile, this system extracts speed limits for the input location based on OpenStreetMap and sends information about speeding vehicles to associated law enforcement agencies. Figure 3 illustrates the web-based user-interface system.

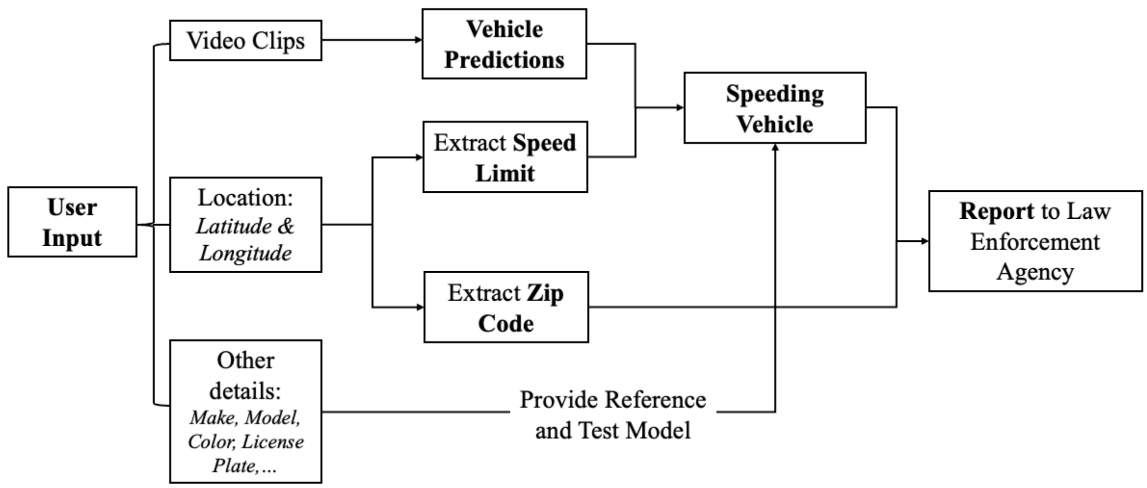

Figure 3: Web-based User-Interface System

4.2 Performance of Make and Model Identification Model

Vehicle make and model identification model is also evaluated on the collected testing dataset in Section 3.2. Among 100 top vehicle models, each has 3 images. Results reveal that the identification model correctly predicts all 3 images of 13 vehicle models, 2 out of 3 images of 30 models, and 1 image of 41 models, with only 16 models (such as Chrysler 300M, Volkswagen Jetta, etc.) not predicted; resulting in an overall accuracy of 46.7%. Nevertheless, for real-world applications, it is not always necessary or feasible to narrow down options to exactly one make and model. Offering a few likely makes and models also narrows down the authorities' search, especially when coupled with human interpretation of the provided video. Thus, Table 5 also shows performance for the top 3 and 5 predictions, with accuracies of 64.0% and 73.0%, respectively.

Table 5. Performance of Vehicle Make and Model Identification Model

# Images Predicted Correctly	Scenario 1: Top 1 Prediction		Scenario 2: Top 3 Predictions[1]		Scenario 3: Top 5 Predictions[2]	
	# Correct Models	Overall Accuracy	# Correct Models	Overall Accuracy	# Correct Models	Overall Accuracy
3	13/100	46.67%	28/100	64.00%	40/100	73.00%
2	30/100		43/100		44/100	
1	41/100		22/100		11/100	
0	16/100		7/100		5/100	

Note: [1]Correct model appears in the top 3 predictions; [2]Correct model appears in the top 5 predictions.

4.4 Law Enforcement Agencies

U.S. states vary greatly in their process of dealing with traffic violations, which adds to the difficulty of developing a comprehensive enforcement map. Different states have their specific jurisdiction agencies for reporting traffic violations, while some states (like Nebraska, Wisconsin, and New York) do not accept individual recordings. Table 7 lists enforcement agency examples for 10 states.

Table 7. Traffic Law Enforcement Agencies at the State Level

State	Agency	State	Agency
Alabama	State Highway Patrol	Arizona	City Courts & State Troopers
Arkansas	State Police	California	Department of Motor Vehicles & State Highway Patrol
Indiana	State Police (commanders)	Kansas	Sheriff's Offices & Police Departments
Kentucky	State Police	Maine	Sheriff's Offices & State Police
Ohio	State Highway Patrol	South Carolina	Department of Public Safety & State Highway Patrol
Why is TX m issing?	...	...	...

Although the predictions remain susceptible to several real-world factors, the web-based user-interface proposed in the paper for individually reporting traffic violations can notably help improve model accuracy. This user-interface platform leverages user inputs, model predictions, OpenStreetMap data, and law enforcement agencies collaboratively and empowers community members to contribute to road safety initiatives. Furthermore, practical implications of implementing this kind of traffic reporting system are discussed in this paper and investigated by taking into account law enforcement agencies at the state level.