PP4AV: A benchmarking Dataset for Privacy-preserving Autonomous Driving

Linh Trinh[✉1] , Phuong Pham[1], Hoang Trinh[1], Nguyen Bach[1],
Dung Nguyen[2], Giang Nguyen[1], and Huy Nguyen[1]

[1]VinFast, Hanoi, Vietnam
v.{linhtk13, phuongpnv, hoangtn, nguyenbh1, giangnd2, vantix01}@vinfast.vn
[2]FPT Software, AI center, Vietnam dungnt20@fsoft.com.vn

Abstract

Massive data collected on public roads for autonomous driving has become more popular in many locations in the world. More collected data leads to more concerns about data privacy, including but not limited to pedestrian faces and surrounding vehicle license plates, which urges for robust solutions for detecting and anonymizing them in realistic road-driving scenarios. Existing public datasets for both face and license plate detection are either not focused on autonomous driving or only in parking lots. In this paper, we introduce a challenging public dataset for face and license plate detection in autonomous driving domain. The dataset is aggregated from visual data that is available in public domain, to cover scenarios from six European cities, including daytime and nighttime, annotated with both faces and license plates. All of the images feature a variety of poses and sizes for both faces and license plates. Our dataset offers not only a benchmark for evaluating data anonymization models but also data to get more insights about privacy-preserving autonomous driving. The experimental results showed that 1) current generic state-of-the-art face and/or license plate detection models do not perform well on a realistic and diverse road-driving dataset like ours, 2) our model trained with autonomous driving data (even with soft-labeling data) outperformed strong but generic models, and 3) the size of faces and license plates is an important factor for evaluating and optimizing the performance of privacy-preserving autonomous driving. The annotation of dataset as well as baseline model and results are available at our github: https://github.com/khaclinh/pp4av.

1. Introduction

Preserving privacy in autonomous driving data is becoming a real and important problem to solve. When machine learning has been used more and more in autonomous driving, companies and research groups have started collecting a large amount of data for developing and validation. Waymo has collected 5 million miles since 2018 [12]. Cruise collected more than 770,000 miles only in 2020 [5]. More collected data comes along with more responsibility for data privacy. For example, data collected in public roads must comply with regulations from European GDPR [6], California CCPA [3], Chinese CSL [4], or Japanese APPI [1]. The regulations require personal identity information of participants in the collected data to be protected, *e.g*. removed when requested. As a response to these regulations, several commercial products have been introduced to de-identify collected data (mostly by blurring camera data). Brighter AI[1], Facebook Mapillary[2], or UAI Anonymizer [11] anonymize the faces and license plates. Celantur[3] goes even further by anonymizing faces, license plates, human body, and whole vehicles.

On the other hand, the scarcity of labeled datasets and baseline models, or the lack thereof, is impeding innovation and progress in solving the anonymization problem in autonomous driving. For the commercial systems mentioned above, their training and evaluation datasets are generally not publicly accessible. Meanwhile, the academic research community has not paid enough attention to this problem. To the best of our knowledge, there is no public dataset for both face and license plate detection for autonomous driving use cases. For example, there are many public face detection datasets, such as PASCAL FACE, FDDB, UFDD, MALF, and recently, WIDER FACE [22, 26, 30, 36, 37], none of which include data from road-driving scenarios. Similarly for license plate detection, there are several public datasets, such as SSIG-SigPlate, UFPR-ALPR, CCPD [24, 27, 35]. Most of the public datasets use normal-lense cam-

[1]https://brighter.ai/video-redaction-in-automotive/
[2]https://www.mapillary.com/geospatial
[3]https://www.celantur.com/

eras, while in autonomous driving, wide-angle lense cameras (*e.g*. fisheye) are usually used to provide a panoramic view around the car. In addition, only a few open source models are available for data anonymization in autonomous driving, *e.g*. Understand AI community edition [11].

In this paper, we introduce a new dataset, named Privacy-Preserving for Autonomous Driving (PP4AV), for face and license plate detection in autonomous driving. PP4AV contains both front camera images and fisheye camera images. We collect front camera images from front-mounted-camera videos of driving scenarios from six European cities. To make the dataset more challenging, we choose data from urban driving where different poses and sizes of target faces and license plates could be available. The fisheye camera images are from WoodScape dataset [38], a public dataset for autonomous driving with fisheye camera. Different from previous datasets where faces license plates were not annotated together and not specific for autonomous driving, PP4AV provides 3,447 annotated driving images for both faces and license plates. The dataset can be used as a benchmark suite (evaluating dataset) for data anonymization models in autonomous driving. The nature of data in PP4AV also enables us finding out that bounding box size is an interesting factor for data anonymization in autonomous driving. Besides PP4AV, we also provide an anonymization model trained with autonomous driving scenarios as a baseline. Due to the lack of labeled data, we use the knowledge distillation approach to train the baseline model from two *teachers*, i.e. YOLO5Face [31] and Understand AI community edition.

In summary, the main contributions of this work are in 3 folds:

- We introduce PP4AV, a dataset for privacy-preservation autonomous driving. To the best of our knowledge, PP4AV is the first public dataset with faces and license plates annotated with driving scenarios. PP4AV can be used as an evaluating suite for privacy-preservation autonomous driving.
- We propose a baseline anonymization model for autonomous driving. While our baseline is trained without actual annotated dataset, the experimental results showed that the baseline outperformed other strong but *generic* models from Amazon or Google.
- In depth-analysis, we found that the bounding box size of face and license plate plays an important role in anonymization models' performance. Interestingly, many models did not perform well with large face or license plate, which is a red-flag for privacy-preserving as the faces or license plates could be easily identified.

The remainder of this paper is organized as follow: we present a survey of related works in the section 2. Then we introduce PP4AV in the section 3. In section 4, we propose a distilled knowledge model as a baseline without annotated dataset. We discuss about the experiments, results, and the deep analysis of failure in the section 5. Finally, section 6 aim to conclude our paper.

2. Related Work

2.1. Data Anonymization Datasets

Existing public datasets annotate faces and license plates separately. Moreover, most face detection datasets do not focus on, or even include driving scenarios, while most of license plate detection datasets only focus on parking lot areas instead of highway or urban scenarios.

Face detection datasets. Public face detection datasets are mostly collected from the Internet or natural scenes. Dataset size ranges from a few hundreds images (*e.g*. AFW [40] with 205 images or PASCAL FACE dataset [22] with 851 images) to several thousands images (*e.g*. WIDER FACE [37] with 32,203 images or MALF [36] with 5,250 images). Images were annotated with bounding boxes surrounding faces and other facial attributes (*e.g*. yaw, pitch and roll). While providing a wide range of conditions (celebrities in news [22], weather-based degradations and motion blur [30]), no existing public face detection datasets are dedicated to driving scenarios. The human faces in the street scenes are usually smaller due to the distance between the ego vehicle and the traffic participants, and can be seen from different viewing angles. These differences pose some unique challenges for data anonymization models in autonomous driving. Our experiments show that models trained with specific autonomous driving data, can outperform strong but *generic* models in this task.

License plate detection datasets. Most datasets for license plate detection were collected from images from traffic monitoring systems, highway toll stations, or parking lots. Zemris [13] provided a dataset that contains fewer than 700 images with only one vehicle in each image. SSIG-SegPlate [24] and UFPR-APLR [27] captured images by cameras on the road. These images were collected on a sunny day. The CCPD dataset [35] is collected from China with more than 200k images, and it has become the largest dataset for license plate recognition. Our dataset PP4AV is different from other public license plate datasets in that it has images from real-world driving situations, which poses new but practical challenges to the state-of-the-art models which were not trained using such data.

Fisheye camera datasets. While data from fisheye cameras are not as popular as normal cameras, they are usually used in autonomous driving systems, *e.g*., 360-degree surrounding view features or smart parking features. There are a few face or license plate detection datasets with fisheye cameras. FDDB-360 [16] and Wider-360 [15] were syn-

thesized (generated) by transforming the original images and annotations of FDDB and WIDER FACE separately to obtain the fisheye-like images and fisheye-like annotations. To our best knowledge, there is no public license plate detection dataset by fisheye cameras in driving scenarios. PP4AV provides license plate annotations from the WoodScape dataset [38], a multi-camera fisheye dataset for autonomous driving.

Last but not least, existing face detection and license plate detection datasets are mutually exclusive. Having both face and license plate detection labels on the same image will give a complete evaluation result for the data anonymization task in autonomous driving. To our best knowledge, PP4AV is the first public dataset for privacy-preserving autonomous driving that offers annotations for both faces and license plates on the same image and include data collected from both normal and fisheye cameras.

2.2. Data Anonymization Models

Commercial productions. In response to the demand for data privacy compliance, commercial products for data anonymization in autonomous driving have emerged. Brighter AI has recently announced their product dedicated to data anonymization in autonomous driving. The deep learning model has been named Deep Natural Anonymization (DNAT) based on R-CNN to do face, license plate, and human body anonymization. Brighter AI provides data anonymization for images of both normal and fisheye cameras. The model has reached 99% accuracy on their private dataset. Celantur launched their product to anonymize people, vehicles, faces, and license plates. Their technique is based on instance segmentation (Mask-RCNN) and keypoint detection. Facebook Mapillary, acquired by Facebook in 2020, also provides data anonymization in autonomous driving by anonymizing faces and license plates. They shared a comparative analysis demonstrating the superiority of their product against other public APIs from Amazon, Google, and Microsoft [8]. dSpace Understand AI has released the beta version of Anonymizer [11]. The model was trained on millions of street view samples, achieving more than 99% detection rate. NavInfo provides the Anonymization product[4] to comply with GDPR [6], CCPA [3], and CSL [4] by detecting and blurring faces and license plates from image data for ADAS validation. Their reported performance for license plates on CCPD is 98.5%, 99.42%, and 98.96% (for average precision, average recall, and F1-score, respectively). For face detection, these numbers are 95.59%, 98.05%, and 96.80%, on the IJB-C dataset [29]. All of these aforementioned products do not publish their models and private datasets for face and license plate anonymization, with the exception of NavInfo, who reported their model performance on public datasets. However, as we pointed out in section 2.1, these datasets are generally not suitable for evaluating data anonymization in autonomous driving. To add to that, the CCPD dataset were only collected in China, therefore not suitable for benchmarking GDPR in the EU or US. Additionally, except for Brighter AI, all the solutions listed above are not applicable to fisheye image data.

Open source projects. There are a few open source projects available for data anonymization tasks. [11] is the community version of dSpace Understand AI. In this paper, we use the community version of Understand AI as one of our baselines. In response to the lack of baseline models for data anonymization in autonomous driving, we have developed a baseline model and made it available for the community. Due to the limited amount of annotations for driving scenarios, we chose a knowledge-distilled approach to train our baseline model. Using current state-of-the-art models as *teacher models*, we can get a lot of soft labels and train our baseline to achieve a reasonable performance for this task.

3. The PP4AV Dataset

3.1. Data collection

Our objective is to build a benchmark dataset that can be used to evaluate face and license plate detection models for autonomous driving. For normal camera data, we sampled images from the existing videos in which cameras were mounted in moving vehicles, running around European cities. We focus on sampling data in urban areas rather than highways in order to provide sufficient samples of license plates and pedestrians. The images in PP4AV were sampled from 6 European cities at various times of day, including nighttime. Given our objectives, we ensure that all images contain at least one object such as a license plate or human face. We use the fisheye images from the WoodScape dataset to select 244 images from the front, rear, left, and right cameras for fisheye camera data. In total, 3,447 images were selected and annotated in PP4AV. The summary of data collection results is summarized in Table 1.

Camera	Cities	Conditions	Road types	Resolution	Images	Face	Plate
Normal	Netherlands	Daytime	Urban	1,920×1,080	388	753	498
Normal	Netherlands	Nighttime	Urban, Highway	1,280×720	824	0	884
Normal	Paris	Daytime	Urban	1,280×720	1,450	2,301	5,571
Normal	Strasbourg	Daytime	Urban	2,048×1,024	50	207	82
Normal	Stuttgart	Daytime	Urban	2,048×1,024	69	132	185
Normal	Switzerland	Daytime	Urban	1,280×720	372	52	449
Normal	Zurich	Daytime	Urban	2,048×1,024	50	154	118
Fisheye	Europe	Daytime	Urban	1280×966	244	296	241
				Total	3,447	3,895	8,028

Table 1: The summary of data collection and annotation PP4AV dataset.

[4]https://www.navinfo.eu/services/ai-business-solution/anonymization/

3.2. Data annotation

Annotating policies. We annotate facial and license plate objects in images. For facial objects, we define the bounding boxes of all detectable human faces from the forehead to the chin to the ears. We label faces with diverse sizes, skin tones, and faces partially obscured by a transparent material, such as a car windshield. A benchmark dataset with a predominance of front faces would enhance the accuracy and efficacy of evaluating data anonymization techniques. For license plate objects, we detect the bounding boxes of all recognizable license plates with high variability, such as different sizes, countries, vehicle types (motorcycle, automobile, bus, truck), and occlusions by other vehicles. In addition, we annotate the license plates of vehicles involved in moving traffic. To ensure the quality of our annotation, we apply a two-step process. In the first phase, two teams of annotators will independently annotate identical image sets. After their annotation output is complete, a merging method based on the IoU scores between the two bounding boxes of the two annotations will be applied. Pairs of annotations with IoU scores above a threshold will be merged and saved as a single annotation. Annotated pairs with IoU scores below a threshold will be considered conflicting. In the second phase, two teams of reviewers will inspect the conflicting pairs of annotations for revision before a second merging method similar to the first is applied. The results of these two phases will be combined to form the final annotation. All work is conducted on the CVAT tool[5].

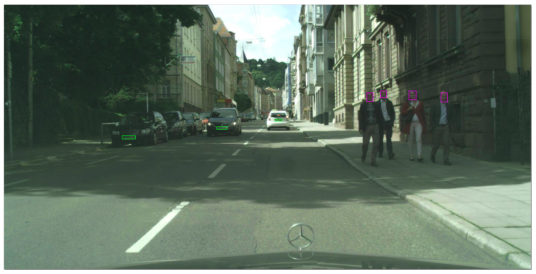
Identifiable objects. The human eye can only recognize objects above a certain size in an image. This means for privacy purpose, we may not need to blur faces or license plates smaller than some threshold. To estimate this threshold, we performed a visual user experience survey with seven participants (aged 22–37, with normal vision capacity) who examined the annotated objects in the image. We created seven different sets for the survey, each containing 35 faces in 7 size groups and 30 license plates in 6 size groups. We randomly assigned each participant to a set and asked them to rate each object as recognizable, hard to recognize, or unrecognizable.

Plate height (pixels)	<5	6-7	8-9	10-14	15-19	>20
Unrecognizable	35	29	19	11	2	0
Hard to recognize	0	6	16	15	0	1
Recognizable	0	0	0	9	33	34

Table 2: Survey of license plate recognizability at different plate heights.

We summarized the object size (width for face and height for license plate) and the rating frequency in Tables 3 and 2. The survey results indicated that bounding boxes need to have a minimum edge of at least 10 pixels for the face and 8 pixels for the license plate in order to be identifiable (by humans). Therefore, we filtered the base annotation by face width and license plate height to generate a new annotation for testing face and license plate detection methods.

[5]https://github.com/openvinotoolkit/cvat



(a) Example of annotation face and license plate on normal camera image

(b) Example of annotation face and license plate on fisheye image

Figure 1: Examples of annotation in PP4AV dataset. (Pink: Face, Green: License Plate).

Face Width (pixels)	<7	8-11	12-14	15-19	20-24	25-29	>30
Unrecognizable	35	28	17	13	2	5	3
Hard to recognize	0	7	15	10	19	10	10
Recognizable	0	0	3	12	14	20	22

Table 3: Survey of license plate recognizability at different face widths.

Compare with other benchmarking datasets. Table 4 shows a detailed comparison between our dataset and other benchmarking datasets. To our best knowledge, there is no specific public dataset for face identification in traffic scenes. The facial recognition benchmarking dataset was obtained from various online sources, while the traffic road scenes were used to generate the open dataset for license plate detection. This dataset is more concentrated on scenes

with license plates, with less human presence. Therefore, we create our own benchmark dataset, since adding extra face annotation to existing datasets is not feasible. We manually select images with abundant people and cars on the road when collecting data for our dataset. Another advantage is that our dataset is the first to provide data for license plate and face detection in diverse European cities. Data obtained from the Internet have different resolutions due to the involvement of multiple sources. Our data selection has a broader range of image resolutions than the open dataset for license plates, which has a fixed resolution. We do not split our dataset into train, val, and test, as our aim is to offer a benchmarking dataset to evaluate the performance of models trained on other datasets.

Last but not least, PP4AV is the first to annotate the faces and license plates of common vehicles, such as cars, buses, trucks, vans, trailers, and motorcycles in both normal camera and fisheye images.

4. Baseline Model

4.1. Model

Training loss. Our approach is based on the Knowledge Distillation technique that was introduced by Hinton *et al*. [25]. The class probabilities of teacher models are distilled into the student model. Kullback-Leibler (KL) divergence is used to measure the distance between student probabilities $p^s_{i,c}$ and teacher probabilities $p^t_{i,c}$ where $c \in 1, 2, .., C$ is the class number. Then a KL-loss is formulated to minimize the KL divergence:

$$loss_{KL} = -\sum_{c=1}^{C} p^t_{i,c'} log(\frac{p^s_{i,c'}}{p^t_{i,c'}}) \quad (1)$$

where

$$p^s_{i,c'} = \frac{exp(p^s_{i,c}/T)}{\sum_{j=1}^{C} exp(p^s_{i,j}/T)} \quad (2)$$

$$p^t_{i,c'} = \frac{exp(p^t_{i,c}/T)}{\sum_{j=1}^{C} exp(p^t_{i,j}/T)} \quad (3)$$

with T being the temperature. To handle the class imbalance problem due to the high scale of resolution and small objects in the image, we replace cross-entropy loss with focal loss [28]. The new total loss function is as follows:

$$loss = \lambda \cdot loss_{iou} + loss^{fl}_{cls} + loss^{fl}_{obj} + \gamma \cdot loss_{KL} \quad (4)$$

where λ is the regression weight for IoU loss $loss_{iou}$, γ is the weight factor for KL divergence loss $loss_{KL}$, and $loss^{fl}_{cls}$, $loss^{fl}_{obj}$ are focal loss for classification and regression respectively.

Teacher models. We consider selecting the candidate of teacher models to be the one with good performance on face or license plate detection tasks. We also evaluate the model on our PP4AV to check the performance of face and license plate detection on street scenes. To the best of our knowledge, UAI Anonymizer is the only public model for data anonymization in autonomous driving. In UAI Anonymizer, there are face and license plate models built separately. For face detection, as presented by Yang *et al*. [37], most of the state-of-the-art methods use WIDER FACE as a training dataset. Based on this argument, we select the algorithms achieving among the top performance on WIDER FACE. The first teacher is the yolo-l version of YOLO5Face [31]. Next, RetinaFace [20] is the face detector of Meta's DeepFace[6] (now this software was stopped due to data privacy). For license plate, we use only license plate detection model from UAI Anonymizer as the teacher model because the training data was designed for EU.

Student Model. YOLOX [23] demonstrated the SOTA as the top ranking method in COCO object detection. So we use the YOLOX as our baseline and optimize it for face and license plate detection. We apply the idea of modifications from YOLO5Face [31] to YOLOX for the detection of small and large objects. We apply 3 modifications to the YOLOX network architecture: (1) replacing the Focus layer with a stem block structure; (2) changing the SSP block to use a smaller kernel; and (3) adding a P6 output block with a stride of 64. Due to faces and plates from far away, the traffic scene was very small, so we disabled Mixups and closed the Mosaic scale in Data Augmentation. We use shear, HSV, and rotation in the data augmentation.

4.2. Training data preparation

We construct our model training dataset from existing open datasets for autonomous driving. These datasets cover a wide range of environments, but lack face and license plate annotations, which are drawbacks for our task. Since we focus on public datasets for self-driving vehicles, we disregard all general-purpose public datasets because they are unrelated to the driving scenario. Another issue is that there are no face and plate annotations in any of the public datasets for self-driving cars. In our approach, we attempt to use the pretrained model (we then utilize this model as a teacher model to train our model), as we have already researched, to teach our model via its prediction rather than annotating and feeding the prediction of these models into our model. Table 5 summarizes the training and validation set in this experiment. We collect the public dataset for autonomous driving. Although the test set was collected only in European cities, the training set contains data not only in Europe. The datasets BDD100K [39], Comma2K19 [32], Bosch [18], India Driving [17], and LeddarPixset [21] were collected outside of Europe. We also leverage CrowdHuman [33] to enrich the facial objects in the street scene. The

[6] https://en.wikipedia.org/wiki/DeepFace

Dataset	Data collection					Number of samples			Annotation objects
	Collected source	Location	Camera	Resolution	Distance	Train/val	Test	Total	
WIDER FACE [37]	www	-	Normal	Varied	-	16k	16k	32,203	faces
FBDD [26]	Yahoo	-	Normal	Varied	-	-	2,845	2,845	faces
IJB-C [29]	www	-	Normal	Varied	-	-	8.3k	130k	faces
MALF [36]	Flickr, www	-	Normal	Varied	-	-	5,250	5,250	faces
AFW [40]	Flickr	-	Normal	Varied	-	-	250	250	faces
UFDD [30]	www	-	Normal	Varied	-	-	6,424	6,424	faces
UFPR-ALPR [27]	Parking slot	Brazil	Normal	1,920×1,080	Close	1,800/9,000	1,800	4,500	plates of cars, motorcycles
Lucian [10]	Traffic road	Romania	Normal	1,280×720	Close, Far	427	107	534	plates of cars
CCPD [35]	Parking slot	Chinese	Normal	1,160×720	Close	100k/100k	-	200k	plates of cars
SSIG-SegPlate [24]	Traffic road	Brazil	Normal	1,920×1080	Close, Far	800/400	800	2,000	plates of cars
Our	On road	4 EU countries	Normal, Fisheye	Varied	Close, Far	-	3,447	3,447	faces, plates of vehicles

Table 4: The comparison of our dataset with other benchmarking dataset ('-': not contained or mentioned in the paper).

diversity of resolution across the dataset will aid the model's multi-scale adaptation over training on only one resolution. All of the collected public datasets guarantee diverse conditions such as weather conditions, various time ranges in a day, road types, and location. We do not use any of these images in the creation of PP4AV because fisheye datasets for autonomous driving are scarce, and we have found no other datasets for fisheye cameras besides WoodScape.

Dataset	Location	Resolution	Train	Val
Cityscape [19]	50 cities Euro	$2{,}048\times1{,}024$	2,921	488
BDD100K [39]	US	1,280×720	41,568	7,370
Comma2K19 [32]	California	1,164×874	6,358	1,414
Bosch [18]	US	$2{,}464\times2{,}056$	3,500	750
India Driving [17]	India	$1{,}920\times1{,}080$	5,332	819
LeddarPixSet [21]	Canada	$1{,}440\times1{,}080$	1,062	228
Kitti [14]	5 cities Euro	1,240×376	7,518	0
CrowdHuman [33]	Varied	Varied	5,332	819
Lucian [10]	Romania	1,280×720	427	107
		Total	74,018	11,795

Table 5: The overview and number of images in the training and validation set of the baseline model.

Algorithm 1: Process pseudo label for training set

Input: Training images T
Teacher models $\theta = \{\theta_1, \theta_2, ..., \theta_n\}$
IoU thresholds V_t
Output: Pseudo label $\mathcal{L}$

$\mathcal{L} = \{\}$
$B = \{B_i | B_i \leftarrow \theta_i(T)\}$
for $s \in T$ **do**
  while $B(s) \neq \varnothing$ **do**
    $(i, p^*) \leftarrow argmax_{p \in B(s)} CF(s)$
    $\hat{B} = B(s) - \{p\}$ where $p \in B_i(s); p \neq p^*$
    for $\hat{p} \in \hat{B}$ **do**
      if $IoU(p^*, \hat{p}) \geq V_t$ **then**
        $B(s) = B(s) - \{\hat{p}\}$
    $\mathcal{L} = \mathcal{L} + \{p^*\}$
    $B(s) = B(s) - \{p^*\}$

4.3. Data preprocessing

In order to prepare data for a training model via distillation, we propose a framework for ensembling multiple teacher models. We propose an algorithm for generating pseudo-labels for training sets. In our algorithm, after we collect a training image T, we use a n-teacher model $\theta_1, \theta_2, ..., \theta_n$ to create the pseudo annotation. With each model M_i, we generate a pseudo label B_i, which contains bounding boxes, class of object, and confidence score. In this step, the output of each teacher model will be processed to make its predictions more robust. We eliminate the bounding boxes with low confidence scores or those with such small bounding box sizes. In the next step, the key idea to enhance the pseudo model is that the candidate of the bounding box will be selected among teacher models by the highest confidence score. In Algorithm 1, with each class of target objects, we process each image s in the training set T. Denote $B(s)$ is the set of predictions of all teacher models on image s, and p is a prediction in $B(s)$ that contains the bounding box, class, and confidence score. We find the best prediction p^* by ranking the confidence score in $CF(s)$ of image s and its corresponding index i of model θ_i. We compare this p^* with all the predictions $\hat{p}$ of model $\theta_j, j \neq i$ by checking the IoU score. If the IoU score of a pair p^* and $\hat{p}$ is greater than a threshold V_t, we consider these two predictions are for the same object, then we eliminate the bounding box with a lower confidence score. We repeat this process on each image until all the predictions have been processed. The algorithm's output, pseudo label $\mathcal{L}$, contains information about each object's bounding box, class, and confidence score. We keep the confidence score in the pseudo label for the distillation task later. We also consider processing the data without the back head of a face to train a model that focuses on the front face and license

plate detection. In this case, we keep the teacher model that only detects the front face and then performs Algorithm 1.

Instead of training a model on all of the data after data preprocessing on the above training dataset, we curate a small, meaningful subset for final training. Based on the processed dataset, we curated three subsets in the order below: (i) firstly, we select the top 20% of images on each dataset by the top number of faces and license plates; (ii) after that, we select the next 20% of the number of images remaining in each dataset by randomly selected; and (iii) finally, our training dataset has carried around 40% of the original selected datasets. The curation of datasets would help the model train faster but keep the performance near the same as training on the whole datasets. The final training dataset was described in Table 5.

5. Evaluation and Analysis

Experiment settings. In our experiment, we compared our baseline methods with unconstrained methods and constrained methods. Unconstrained methods include Google API [7], AWS API [2] for face detection, and UAI Anonymizer for both face and license plate detection. Constrained methods consist of RetinaFace and YOLO5Face, which are trained on the WIDER FACE dataset for face detection, and ALPR [34], NVIDIA LPDnet [9] models for license plate detection. For hyperparameter, λ, γ, and T were all set to 1, and V_t is set to 0.2. All experiments were conducted on an NVIDIA DGX A100 server with eight GPUs.

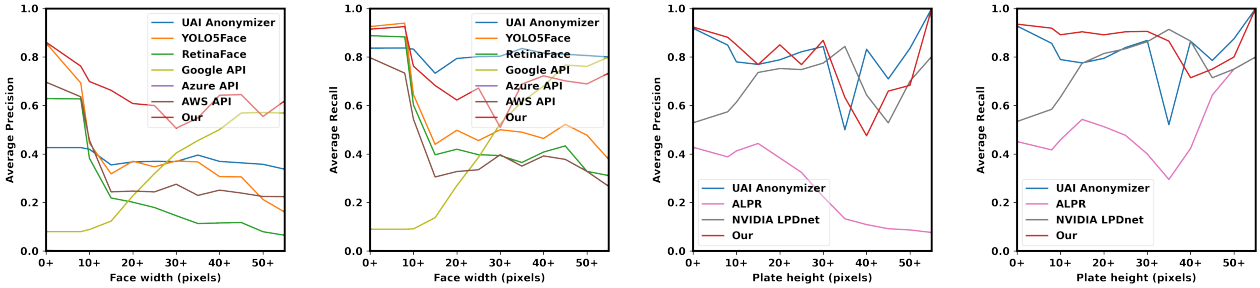
	Methods	Normal images		Fisheye images	
		AP_50	AR_50	AP_50	AR_50
Face	UAI Anonymizer [11]	42.62%	83.7%	43.98%	53.33%
	AWS API [2]	63.69%	73.33%	40.72%	46.67%
	Google API [7]	7.97%	8.99%	7.64%	8.89%
	RetinaFace [20]	62.71%	88.28%	43.82%	62.96%
	YOLO5Face [31]	69.31%	**93.96%**	**69.59%**	**82.96%**
	Our	**76.22%**	92.52%	59.2%	63.92%
Plate	ALPR [34]	38.79%	41.68%	17.26%	31.21%
	NVIDIA LPDnet [9]	57.41%	58.44%	24.9%	26.24%
	UAI Anonymizer [11]	84.89%	85.61%	44.14%	53.9%
	Our	**88.12%**	**91.88%**	**49.53%**	**58.17%**

Table 6: Average Precision (AP) and Average Recall (AR) scores corresponding to different methods on PP4AV dataset with face width and license plate height greater than 8 pixels ('-': model have no detection).

Results. Table 6 presents the comparison of the performance of faces and license plate detection methods on both normal and fisheye images in the PP4AV dataset with the object size filtered by face width and plate height greater than or equal to 8 pixels. Performance from a wide range of state-of-the-art models shows the importance of using same-domain training data in order to achieve high performance. The results demonstrated that normal images perform better than fisheye images for both face and license plate detection. This trend would come from the fact that training (labeled) data for fisheye cameras is not as available as data from normal cameras.

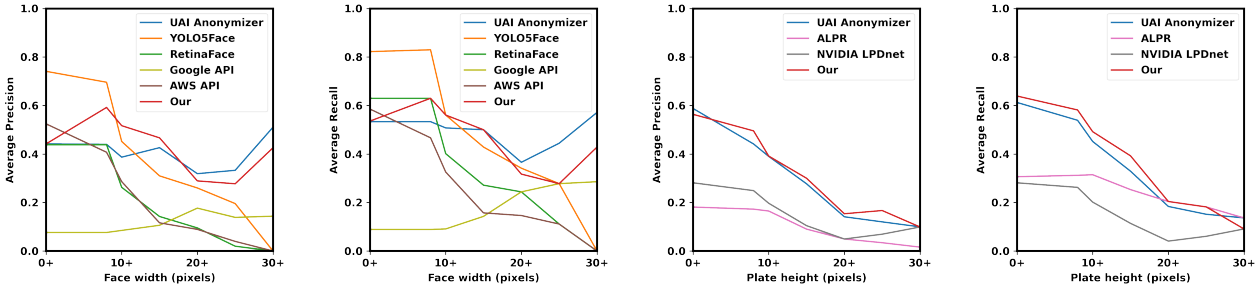
For face detection, it is quite interesting to see some strong but *generic* baseline models (unconstrained models) achieve pretty low performance (in both precision and recall) on PP4AV. On the other hand, models like UAI Anonymizer (trained specifically for anonymization) or RetinaFace and YOLO5Face (which are the best performers in large face detection benchmark suites) showed better performance on our dataset. Our model, learned from RetinaFace and YOLO5Face and trained on unlabeled driving scenarios, achieved the best precision in the normal camera and ranked second in the fisheye camera. It is important to note that we did not use fisheye data to train our baseline. Therefore, our baseline did not outperform the YOLO5Face model in the fisheye camera. The same observation has been made for license plate detection results. UAI Anonymizer outperforms ALPR and NVIDA LPDnet on both normal images and fisheye images. Both of these methods have two stages, with the license plate detection stage coming after the car detection stage, and the low performance due to the car detection stage. Our model, which incorporates hundreds of ground truth samples from Lucian, has outperformed the UAI Anonymizer in terms of performance. Even though some SOTA models achieve pretty good performance on PP4AV, there is still a lot of room for improvement on this problem. This introduces new challenges for new privacy-preserving models. To conduct a more rigorous analysis, we further evaluate the model under the condition that the object size exceeds a certain threshold. We employ face width to filter object size for faces since faces with small or oblique faces will have smaller faces. We apply height as a criterion to filter object size for license plates, as license plates typically have smaller heights than widths. While Figure 3 is for fisheye images, Figure 2 displays the average precision and recall of objects detected by filtering different object sizes in both normal and fisheye images. Except for Google API, other models reduce their AP and AR when the face size increases. This trend would be a big issue for privacy-preserving as the larger the faces, the easier it is to identify the person. When object size increases, even Google API increases AP and AR. UAI Anonymizer maintains the top recall because it has learned to detect human heads. In terms of AP scores, our model continues to perform better than others.

Both the UAI Anonymizer and our model for license plates maintain great performance with various plate heights. When plate size increases, NVIDIA LPDnet improves recall and precision. Our model is 100% accurate



(a) Face Average Precision (b) Face Average Recall (c) Plate Average Precision (d) Plate Average Recall

Figure 2: Average Precision and Average Recall of face (a-b) and license plate (c-d) detection versus object size (face width and plate height) in normal images.



(a) Face Average Precision (b) Face Average Recall (c) Plate Average Precision (d) Plate Average Recall

Figure 3: Average Precision and Average Recall of face (a-b) and license plate (c-d) detection versus object size (face width and plate height) in fisheye images.

and recalls on plates with a height greater than 55 pixels. The AP and AR of face and license plate detection in fisheye images are represented in Figure 2. The graphs 3a and 3b demonstrate that all algorithms rapidly decrease AP and AR for face detection as face width increases. The performance of our model and the UAI Anonymizer has been slowly dropping. When the height of the license plate increases, AP and AR are reduced for all methods. It demonstrates that models are unable to recognize large, distorted plates. It can be difficult and require additional work in the future.

6. Conclusions

We present the first dataset (PP4AV) annotation for faces and license plates in the context of autonomous driving in this paper. PP4AV demonstrated challenges to current state-of-the-art face and license plate detection models in experimental results. We hope that PP4AV will encourage further research into a privacy-preserving model for autonomous driving. Furthermore, by refining the state-of-the-art deep learning method based on YOLOX, we proposed a new baseline for face and license plate detection in autonomous driving. While not requiring any labeled data, our model outperformed some strong but generic SOTA models. We also published a comprehensive failure analysis that investigates the limitations of the existing face and license plate methods in order to provide guidance for the development of future algorithms. Plans for the future include providing a more diverse circumstances dataset for data anonymization. Our proposed datasets and approaches carry the risks associated with large vision models. Our datasets have the potential to spread offensive, social biases, and stereotype images and meta data. To filter out offensive data in real-world applications, we can use rule-based methods or train a specific classifier. This is an area that we intend to investigate further.

7. Acknowledgments

This work would not have been possible without the supports of Vantix Inc and Vinfast LLC, the technology and automotive companies in Vingroup. The technical challenges that we encountered in the projects here are enabling us to do meaningful research.

We would also like to express our gratitude towards colleagues and experts, who we have opportunities to consult with during the course of this research, for their valuable comments and advices.

References

[1] Act on protection of personal information – appi. https://www.ppc.go.jp/files/pdf/APPI_english.pdf.

[2] Amazon rekognition. https://docs.aws.amazon.com/rekognition/latest/dg/faces.html. Accessed: 2022-07-12.

[3] California consumer privacy act. https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf.

[4] China cybersecurity law (csl). http://www.cac.gov.cn/2016-11/07/c_1119867116.htm.

[5] Cruise sets 2020 mileage record for av testing in california. https://www.govtech.com/fs/cruise-sets-2020-mileage-record-for-av-testing-in-california.html. Accessed: 2022-07-12.

[6] General data protection regulation (gdpr). https://gdpr-info.eu/.

[7] Google cloud vision face detection. https://cloud.google.com/vision/docs/detecting-faces. Accessed: 2022-07-12.

[8] Mapillary faces and license plates detection performance. https://blog.mapillary.com/update/2019/09/12/protecting-privacy-better-maps.html. Accessed: 2022-07-12.

[9] Nvidia license plate detection (lpdnet). https://catalog.ngc.nvidia.com/orgs/nvidia/models/tlt_lpdnet. Accessed: 2022-07-12.

[10] Romanian (european union) dataset of license plates. https://github.com/RobertLucian/license-plate-dataset. Accessed: 2022-07-12.

[11] Understand ai anonymizer. https://github.com/understand-ai/anonymizer. Accessed: 2022-07-12.

[12] Waymo reaches 5 million self-driven miles. https://blog.waymo.com/2019/08/waymo-reaches-5-million-self-driven.html. Accessed: 2022-07-12.

[13] Zemris: Zemris license plate dataset. http://www.zemris.fer.hr/projects/LicensePlates/hrvatski/rezultati.shtml.

[14] Vision meets robotics: The kitti dataset. *International Journal of Robotics Research*, 32, 2013.

[15] Datasets for face and object detection in fisheye images. *Data in Brief*, 27, 2019.

[16] Fddb-360: Face detection in 360-degree fisheye images. *Proceedings - 2nd International Conference on Multimedia Information Processing and Retrieval, MIPR 2019*, 2019.

[17] Idd: A dataset for exploring problems of autonomous navigation in unconstrained environments. *Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019*, 2019.

[18] Karsten Behrendt. Boxy vehicle detection in large images. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 2019.

[19] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 2016.

[20] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.

[21] Jean Luc Deziel, Pierre Merriaux, Francis Tremblay, Dave Lessard, Dominique Plourde, Julien Stanguennec, Pierre Goulet, and Pierre Olivier. Pixset : An opportunity for 3d computer vision to go beyond point clouds with a full-waveform lidar dataset. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2021-September, 2021.

[22] Mark Everingham, S. M.Ali Eslami, Luc Van Gool, Christopher K.I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: a retrospective. *International Journal of Computer Vision*, 111, 2015.

[23] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[24] Gabriel Resende Gonçalves, Sirlene Pio Gomes da Silva, David Menotti, and William Robson Schwartz. A benchmark for license plate character segmentation. *Journal of Electronic Imaging*, 25, 2016.

[25] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv*, 2015.

[26] V Jain and Eg Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report*, 2010.

[27] Rayson Laroca, Evair Severo, Luiz A. Zanlorensi, Luiz S. Oliveira, Gabriel Resende Goncalves, William Robson Schwartz, and David Menotti. A robust real-time automatic license plate recognition based on the yolo detector. *Proceedings of the International Joint Conference on Neural Networks*, 2018-July, 2018.

[28] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[29] Brianna Maze, Jocelyn Adams, James A. Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K. Jain, W. Tyler Niggel, Janet Anderson, Jordan Cheney, and Patrick Grother. Iarpa janus benchmark – c: Face dataset and protocol. *Proceedings - 2018 International Conference on Biometrics, ICB 2018*, 2018.

[30] Hajime Nada, Vishwanath A. Sindagi, He Zhang, and Vishal M. Patel. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. *IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS 2018*, 2018.

[31] Delong Qi, Weijun Tan, Qi Yao, and Jingfeng Liu. Yolo5face: Why reinventing a face detector. *ArXiv preprint ArXiv:2105.12931*, 2021.

[32] Harald Schafer, Eder Santana, Andrew Haden, and Riccardo Biasini. A commute in data: The comma2k19 dataset. *arXiv:1812.05752*, 2018.

[33] Shuai Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu, Xiangyu Zhang, and Jian Sun. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018.

[34] S. M. Silva and C. R. Jung. License plate detection and recognition in unconstrained scenarios. pages 580–596, Sep 2018.

[35] Zhenbo Xu, Wei Yang, Ajin Meng, Nanxue Lu, Huan Huang, Changchun Ying, and Liusheng Huang. Towards end-to-end license plate detection and recognition: A large dataset and baseline. *Lecture Notes in Computer Science*, 11217 LNCS, 2018.

[36] Bin Yang, Junjie Yan, Zhen Lei, and Stan Z. Li. Fine-grained evaluation on face detection in the wild. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2015*, 2015.

[37] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 2016.

[38] Senthil Yogamani, Christian Witt, Hazem Rashed, Sanjaya Nayak, Saquib Mansoor, Padraig Varley, Xavier Perrotton, Derek Odea, Patrick Perez, Ciaran Hughes, Jonathan Horgan, Ganesh Sistu, Sumanth Chennupati, Michal Uricar, Stefan Milz, Martin Simon, and Karl Amende. Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. *Proceedings of the IEEE International Conference on Computer Vision*, 2019-October, 2019.

[39] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.

[40] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.