ADGAN: Protect Your Location Privacy in Camera Data of Auto-Driving Vehicles

Zuobin Xiong, Zhipeng Cai, *Senior Member, IEEE,* Qilong Han, Arwa Alrawais, *Member, IEEE,* and Wei Li, *Member, IEEE*

***Abstract*—Computer vision and deep neural networks have been significantly promoting the development of visual perception in these years. Particularly, for autonomous vehicles, real-time image/video data is captured by onboard cameras and analyzed by computer vision techniques in many real applications. In the captured camera data, some contents can be used as auxiliary information to infer individuals' locations and trajectories, which leads to severe privacy leakage but has been rarely studied. Thus, the goal of this paper is to protect individuals' location privacy by hiding side-channel information in the captured data while preserving the the data utility for downstream applications. To this end, the technology of Generative Adversarial Networks (GAN) is utilized to design two novel models, named ADGAN-I and ADGAN-II, both of which can take the original camera data as inputs and generate privacy-preserving outputs according to predefined sensitive object class. Thus, the processed camera data can defend location inference attack from adversaries in off-line applications. Moreover, in ADGAN-I and ADGAN-II, the tradeoff between location privacy and data utility can be effectively balanced. Finally, the results of extensive real-data experiments validate the superiority of our proposed models over the state-of-the-arts in utility preservation and privacy protection for autonomous vehicles' images and videos.**

***Index Terms*—Computer Vision, Autonomous Vehicles, Generative Adversarial Networks, Location Privacy.**

I. Introduction

THANKS to the technical innovation of visual perception and deep neural networks, the techniques of autonomous vehicles are becoming increasingly mature, greatly accelerating the development of automobile industry. More and more autonomous vehicles come into being, including Tesla, BMW, and Ford, *etc*. [1]. Nowadays, the autonomous vehicles have successfully driven millions of miles without human control, which is mainly benefited from their high-performance perception and decision-making systems guided by a huge amount of perceptual data. The perceptual data comes from various onboard sensors; for instances, GPS sensors collect real-time location data with exact coordinates, radar sensors detect surrounding objects and their distance to vehicles, behavior-relevant sensors monitor environment inside the car and record operation of passengers, and camera sensors work as the eyes of vehicles to perceive visual view and instruct driving behaviors. These sensory data not only can facilitate autonomous vehicles but also can be utilized as precious source for smart city, smart transportation and many other applications.

Meanwhile, carrying such a valuable data source makes the autonomous vehicles vulnerable to malicious inference attacks, in which attackers can obtain private information of the vehicles/drivers/passengers through data mining. For instances, with behavior-relevant data, attackers can infer passengers sensitive information, such as sex, age, and hobby [2], [3]; and with GPS data, attackers can figure out passengers' locations, trajectories, home/work addresses, health conditions and move patterns. In this paper, our privacy concern mainly focuses on the use of camera data in off-line (or non real-time) applications, such as 3D street view model construction. The camera is deployed as a core component of perception systems in the autonomous vehicles to monitor and record real-time road conditions. Additionally, the camera can collect lots of "over-captured" information more than it needs, *e.g.*, street view background, pedestrians' faces on the streets, plate numbers and models of surrounding vehicles. This "over-captured" information contains various private information about vehicles and individuals, which can be used to infer the locations and trajectories of drivers and passengers, leading to severe risk of location privacy leakage [4]. An example is demonstrated in Fig. 1. If an attacker gets a camera image from a victim's vehicle as shown in Fig. 1(a), she is able to figure out that the location of victim's vehicle is nearby "Washington Monument", in which location privacy is totally leaked via image background without any GPS sensor data. Again, if the attacker can get another camera image from the same victim's vehicle, she can identify the location (*i.e.*, "Washington National Cathedral" in Fig. 1(b)) and even estimate the victim's trajectory and speed by learning the driving time between two locations as shown in Fig. 1(c). Thus, for autonomous vehicles, the camera data is a type of critical side-channel information for location inference. More importantly, the power of location inference attack has been greatly enhanced with recent progresses of visual perception, deep neural networks, and other detection techniques. In [5], [6], the authors have already proposed vocabulary tree-based and feature-based matching methods to detect location in the image data with a high recognition accuracy more than 74%. This fact further confirms that the attackers have powerful abilities to identify a victim's location and trajectory through

Zuobin Xiong, Zhipeng Cai and Wei Li are with the Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA (email: zxiong2@student.gsu.edu; zcai@gsu.edu; wli28@gsu.edu)

Qilong Han is with the College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang 150001, China (email: hanqilong@hrbeu.edu.cn)

Arwa Alrawais is with the College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia (email: a.alrawais@psau.edu.sa)

Corresponding Authors: Wei Li and Arwa Alrawais

(a) Camera data from victim near Washington Monument (b) Camera data from victim near Washington National Cathedral (c) Leakage of location and trajectory

Fig. 1. An example to illustrate how victim's location and trajectory privacy is inferred by location inference attack (pictures source from Google Map).

the perceived camera data. Therefore, **the camera data from autonomous vehicles is in desired need of effective solutions to defend location inference attack**.

To the best of our knowledge, [4] is the only existing work studying location privacy through side-channel information in images. However, the solution of [4] is limited to processing low resolution static images and is not able to handle satisfied-quality image and video data. Thus, in this paper, **our objective is to protect location privacy while maintaining data utility so that the processed camera data (*e.g.*, images and videos) are still suitable for off-line applications without privacy leakage**, which takes into account the following critical issues. (i) Since this problem has not been studied well, how to mathematically model the problem and design powerful defense strategies is an open question. (ii) The camera data is supposed to retain utility after being processed by protection mechanisms so that it can used in off-line applications. Hence, it is necessary to consider how to effectively balance the tradeoff between location privacy and data utility. (iii) The existing privacy-preserving method only treated background buildings as side-channel information for privacy leakage [4]. In fact, except for background buildings, some unique or obvious objects (*e.g.* trees and mountains), can also be the identifiers of specific locations. Thus, customized approaches are needed to protect user's defineded private objects.

In order to tackle the aforementioned challenges, we develop two Generative Adversarial Networks (GAN)-based models: **ADGAN-I** and **ADGAN-II**. The protection strategy is to modify the appearance of original captured data slightly according to model feedback such that the generated results can avoid feature extraction and location inference attack while keeping recognition utility for real-world applications (*e.g.*, traffic analysis and 3D street view model construction). It is worth mentioning that, besides a generator and a discriminator in traditional GAN framework, an additional helpful component called "target model" is proposed for performance improvement. The tradeoff between privacy protection and utility preservation is managed by well designed loss functions. Furthermore, to maintain the context structure of the original data, multiple discriminators are deployed to enhance the ability of reconstructing the synthetic data. In conclusion, our contributions are addressed below.

- We design two novel GAN-based approaches, including ADGAN-I and ADGAN-II, for data generation, which can prevent autonomous vehicles' locations being inferred through their vehicular camera data.
- To improve the flexibility of our methods for various applications, in ADGAN-I and ADGAN-II, users are able to customize their private objects in camera data.
- We make full use of unlabeled complex street view data with the help of semi-supervised training idea, which enables our model to adapt more diverse data distribution and improve performance.
- We implement comprehensive experiments on real image and video datasets to illustrate that our models can resist location inference attack while preserving utility.

The rest of this paper is organized as follows. The related works are introduced in Section II. Then the methodology of proposed models is presented in Section III. Intensive experiments are conducted in Section IV. Finally, in Section V, this paper is concluded, and our future work is discussed.

II. Related Works

The most related works are summarized from three aspects: *privacy protection with GAN*, *data privacy in autonomous vehicles*, and *visual privacy against machine learning*.

A. Privacy Protection with GAN

To protect data privacy, the generator G and the discriminator D of GAN are respectively modeled as a defender and an attacker such that the attacker cannot infer private information after the training process of GAN is completed. In [7], [8], the authors proposed generative methods of full body and face de-identification to prevent human ID from being recognized by attackers as well as to preserve the utility of generated data. For images, GAN-based visual protection methods are also introduced by [9], [10], in which GAN was utilized as an obfuscation mechanism to decrease the detection accuracy of specific pixels in images. Then Chen *et al.* [11] designed a VGAN-based image representation learning scheme for privacy-preserving facial expression recognition, which can protect human ID privacy and maintain expression recognition accuracy. In the field of Natural Language Processing (NLP), Li *et al.* proposed a GAN-based method that can prevent attackers from inferring the age and sex of text writers while maintaining the utility of emotion analysis for NLP. In [12], the loss function of GAN was used as a regularization term to train a robust machine learning model to resist the membership inference attack. Besides, GAN-based schemes are also used for secure wireless communications [13], private data publishing [14], and context-aware medical image synthesis [15], *etc*.

However, when it comes to our problem considered in this paper, most of the existing GAN-based methods can not be used directly. Because the street view data has very complex structure and many different object classes, the existing methods that mainly focus on small objects (such as individuals' faces and vehicle plates) may not be able to perform complicated object classification. Moreover, it is easy to modify the small objects without losing structure information of entire images; on the contrary, for the large objects (*e.g.*, background constructions), it becomes harder to process imperceptible modification for privacy preservation while retaining the recognition utility of data.

B. Data Privacy in Autonomous Vehicles

With numerous data collected by embedded sensors, autonomous vehicles are easily targeted by malicious attackers who intend to mine individuals' private information for various purposes. Particularly, location related data, *e.g.*, accurate GPS coordinates, is demanded by many location based services (LBS). Moreover, such location data can be easily linked to a variety of sensitive information that individuals usually are not willing to publish, such as home and workplace addresses, sexual preferences, and political views [16]–[18].

Protecting location privacy for the vehicles in traditional vehicular networks has attracted lots of research interest, but few of them are for autonomous vehicles. In [19], the authors proposed a Social-based PRivacy-preserving packet forwarding (SPRING) method based on a cryptography framework to guarantee secure and private data transmission. To protect location privacy, Lu *et al.,* designed a Social spot-based Packet Forwarding (SPF) protocol to obfuscate real time location of vehicles [20]. In [16], a differentially private approach was developed to protect location and trajectory privacy by using Hilbert spatial division. In [21], the authors presented a novel protocol, named Social-Tier-Assisted Packet (STAP), for Vehicular Ad Hoc Networks, which works well in disseminating packets and preserving receivers' locations. Nevertheless, the location privacy in the above works is limited on LBS.

So far, [4] is the only existing work that considers location privacy leakage through side-channel information in images. But the image quality and flexibility of the mechanism in [4] may not be adequate for autonomous vehicles, especially in high resolution. In this paper, we focus on the resistance of location inference attack for camera data in auto-driving vehicles by developing two novel GAN-based defense strategies.

C. Visual Privacy against Machine Learning

Visual data, such images and videos, is vulnerable to privacy leakage, because it usually contains abundant (even human-imperceptible) graphical and semantic information that can be easily learnt by machine learning-based classification/detection models. To protect the privacy of visual data, adversarial idea has been applied in many studies to defend model prediction, which is normally accomplished by adding imperceptible crafted-noise to obfuscate original visual data [22]–[25]. In [26], a encryption-based method was used to keep visual privacy during transferring by pseudorandomly flipping private information so that only the authorized receiver who has the secret key can decode and recover the private data while unauthorized attackers get noisy results. The goal of [27] was to preserve human face privacy in visual data, where a private face was morphed by adding another face on it pixel by pixel such that the private face is transferred to an interpolated unknown face. However, the above methods work only on small objects (*e.g.* human faces and license plates) through private object detection and noisy blurring, and they also require additional information for implementation, such as secret key, secure channel, and additional faces images. Advanced methods were developed to process other variable objects with the help of deeper neural networks [28], [29]. In [28], [29], multi-classification CNN was utilized to classify all objects in social network images into 68 predefined attributes, and the private information in published images was noised or removed automatically based on user's privacy preference. This method is efficient but often reduce image utility because of brute removal. Liu *et al.* [30] proposed a mapping distortion-based method to construct a modified dataset for protecting privacy of entire dataset as a whole. The authors in [31] used a concept termed *sensitivity map* to learn the sensitivity of each pixel in an image regarding to a detection/classification model and then treated the *sensitivity map* as prior knowledge to add context-aware noise to original data. Although the methods of [30], [31] perform well on image privacy protection, they also get stuck in image field and can not resist location inference attack for visual data.

Notably, in this paper, the objective, the challenges, and the proposed methods are significantly different from the state-of-the-arts. Besides protecting visual privacy of objects in images, we also use visual data as an intermediate-bridge to defend location inference attack. It is worth mentioning that in visual data, the objects that are side-channel information for location inference are varied and not easy to be identified. Our proposed models, including ADGAN-I and ADGAN-II, not only can generate high-quality privacy-preserving results for complex visual data smoothly and efficiently, but also can be utilized on video data for real applications.

III. Design of Auto-Driving GAN Models

To preserve location privacy in image and video data for autonomous vehicles, we propose two Auto-Driving GAN (ADGAN) models, including ADGAN-I and ADGAN-II. Since a video can be split into a sequence of frames for processing, this section focuses on the details of generating synthetic images in ADGAN-I and ADGAN-II for presentation simplicity. In Section IV, extensive experiments are analyzed to show the applicability of ADGAN-I and ADGAN-II to both images and videos.

A. Methodology Overview

The frameworks of ADGAN-I and ADGAN-II are presented in Fig. 2 and Fig. 3, respectively, where there is a generator module, a discriminator module, and a target module that is used as attack feedback of the generator. To prevent privacy

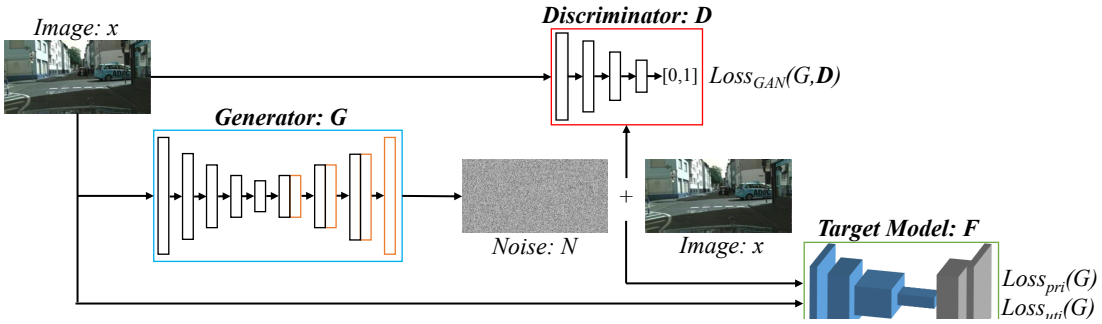


Fig. 2. 'een Noise .

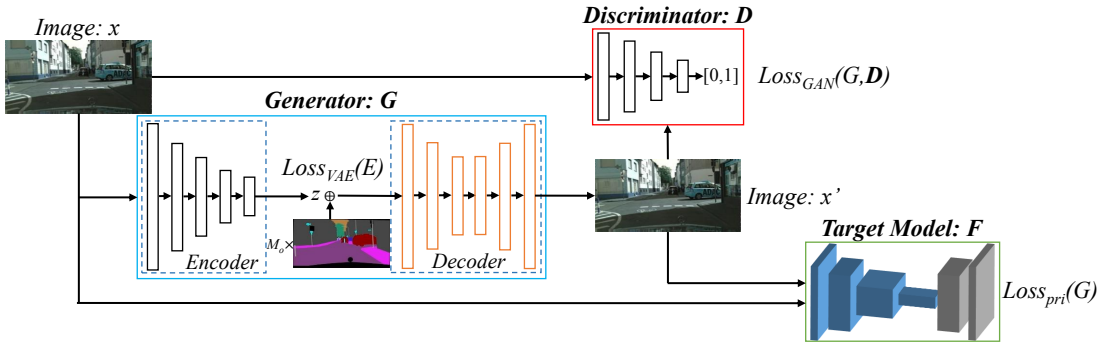


Fig. 3. System framework of ADGAN-II with a VAE-based generator G, a multi-discriminator structure $\mathbf{D}$, and a target model F. The notation $\oplus$ between z and label map $M_o \times Y$ represents concatenation operation. $\mathcal{L}_{GAN}(G,D)$, $\mathcal{L}_{VAE}(G)$ and $\mathcal{L}_{pri}(G)$ are illustrated in section III-C.

leakage from side-channel information (*e.g.*, background constructions) of the original image set I, the generator G of ADGAN-I and ADGAN-II is trained to generate a synthetic image x' from $G(x)$, where $x \in I$ is an original image and x' is a sample of the synthetic image set I'. That is, every image x captured by the onboard cameras is input into G to produce a synthetic image x'. To achieve our expected image synthesis, the generator G needs to accomplish three objectives simultaneously: (i) the generated images should be as real as possible; (ii) the recognition accuracy of non-sensitive objects should be maintained; (iii) the accuracy of location inference should be decreased. On the other hand, we hire multiple discriminators in our discriminator module $\mathbf{D}$, for which the benefits are explained in Section III-B.

Corresponding to the aforementioned three objectives, we integrate the generator G, multi-discriminator $\mathbf{D}$, and target module F to establish a general loss function for both ADGAN-I and ADGAN-II as follows:

$$\mathcal{L}_{ADGAN}(G,\mathbf{D}) = \mathcal{L}_{GAN}(G,\mathbf{D}) + \lambda_1 \mathcal{L}_{uti}(G) + \lambda_2 \mathcal{L}_{pri}(G), \quad (1)$$

where $\mathcal{L}_{GAN}$ is the loss function of GAN for objective (i), $\mathcal{L}_{uti}(G)$ is used to maintain similarity between the original image x and the synthetic image x' for objective (ii), $\mathcal{L}_{pri}(G)$ quantifies the loss of private information to control the privacy protection of targeted objects (*e.g.* background constructions) for objective (iii), and λ_1 and λ_2 are hyperparameters.

After the training process is completed, the synthetic privacy-preserving images can be generated from G^* by

$$G^* = \arg\min_G \max_D [\mathcal{L}_{GAN}(G,\mathbf{D}) + \lambda_1 \mathcal{L}_{uti}(G) + \lambda_2 \mathcal{L}_{pri}(G)], \quad (2)$$

where "min-max" implies that G can generate privacy-preserving images even $\mathbf{D}$ is optimized as the most powerful discriminator. Given G^*, the privacy-preserving data can be produced quickly by the generator modules for autonomous vehicles without location privacy leakage.

B. Details of ADGAN-I Model

In the following, we introduce how ADGAN-I model works.

1) **Generator Module:** The generator aims to produce privacy-preserving data by obfuscating the original data. In the original images, the private objects (*e.g.,* background constructions) should be modified to be different, while the other non-private objects should be kept as similar as possible. In other words, it is requested that the original images and the synthetic images are different in private object pixels but have similar whole image structure. To this end, we utilize "U-Net" structure [32] to create specific noise for each image. The encoder of "U-Net" compresses the original images into latent representation, which can preserve the most essential information of original images. Based on this image-dependent information, the decoder of "U-Net" can produce appropriate noise that achieves privacy protection and utility preservation.

The network structure of G is shown in TABLE I, where the encoder and decoder are built according to the structure of [33]. Specifically speaking, given any image x, the output of the generator G is an image-dependent noise denoted by N, *i.e.*, $N = G(x)$ and the final output of the generator module is the corresponding synthetic image $x' = x + N = x + G(x)$ that is sent to the discriminator $\mathbf{D}$ and the target mode F as

TABLE I
STRUCTURE OF U-NET IN ADGAN-I

L	Encoder	Decoder
1	$(5,5) \times 64$, Leaky ReLU	$(5,5) \times 512$ deconv, BN, ReLU
2	$(5,5) \times 128$, BN, Leaky ReLU	$(5,5) \times 512$ deconv, BN, ReLU
3	$(5,5) \times 256$, BN, Leaky ReLU	$(5,5) \times 512$ deconv, BN, ReLU
4	$(5,5) \times 512$, BN, Leaky ReLU	$(5,5) \times 512$ deconv, BN, ReLU
5	$(5,5) \times 512$, BN, Leaky ReLU	$(5,5) \times 256$ deconv, BN, ReLU
6	$(5,5) \times 512$, BN, Leaky ReLU	$(5,5) \times 128$ deconv, BN, ReLU
7	$(5,5) \times 512$, BN, Leaky ReLU	$(5,5) \times 64$ deconv, BN, ReLU
8	$(5,5) \times 512$, BN, Leaky ReLU	$(5,5) \times 3$ deconv, tanh

shown in Fig. 2. For the simplicity of presentation, we use $x' = G(x)$ to represent the image generation process.

Next, to achieve the objective (i), the loss function of the generator module can be formulated in Eq. (3).

$$\mathcal{L}_{GAN} = \mathbb{E}_{x\sim I}[\log \mathbf{D}(x)] + \mathbb{E}_{x\sim I}[\log(1 - \mathbf{D}(G(x))]. \quad (3)$$

Additionally, "utility loss" for the objective (ii) and "privacy loss" for the objective (iii) are analyzed in Section III-B3 and Section III-B4, respectively.

2) **Discriminator Module:** The discriminator module $\mathbf{D}$ is deployed to distinguish whether its input is from the real image set I or the synthetic image set I'. Considering the properties of street view images, we employ multiple discriminators to configure the discriminator module $\mathbf{D}$ for better generation performance, *i.e.*, $\mathbf{D} = \{D_1, D_2, \ldots, D_k\}$ where k is a positive integer larger than 1. It has been illustrated that a single discriminator does not have adequate ability to distinguish the complex images because of high-resolution and multiple classes [34]. What's more, only one discriminator with fixed receptive field can only perceive certain pattern in the complex street view image data, thus it is very easy to be fooled by a powerful generator. While, multiple discriminators with different filters can augment the discrimination capability, improving the performance of data utility preservation and privacy protection even in the presence of a powerful adversary.

In our setting, each discriminator, D_i ($1 \leq i \leq k$), is a convolutional neural networks (CNN) built for real/fake binary classification. For the purpose of data generation, we first need to ensure the distribution of the generated data is the same as that of the original data from a global aspect, and then aim to perfect the details of generated data as realistic as possible. More concretely, D_i with smaller i is designed as a CNN with a smaller receptive field to improve the details of generated images, and D_i with larger i is designed as a CNN with a quite large receptive field to scan the entire input image. The output of $\mathbf{D} = \{D_1, D_2, \ldots, D_k\}$ is a scalar that indicates the probability of real data, which is the summation of all D_i on the original and generated data. Thus, in our multi-discriminator setting, the loss function Eq. (3) can be formally rewritten by Eq. (4).

$$\mathcal{L}_{GAN} = \sum_{i=1}^{k}[\mathbb{E}_{x\sim I}[\log D_i(x)] + \mathbb{E}_{x\sim I}[\log(1 - D_i(G(x)))]]. \quad (4)$$

3) **Utility Preservation:** As well known, GAN-based models are vulnerable to the mode collapse [35]. Therefore, we propose two strategies to mitigate the impact of mode collapse and improve the data utility. (i) Generating small magnitude noisy masks. In Section III-B1, the generator G does not produce image x directly. Instead, the final result is produced by $x' = x + N$, where even the mode collapse happens, the degradation of N will not impact x' too much because its magnitude is less than that of real image x. (ii) Regularization with additional loss function. The loss function of ADGAN-I is not just a min-max game as the original GANs. We define the utility loss $\mathcal{L}_{uti}(G)$ and the private loss $\mathcal{L}_{pri}(G)$ and introduce a target model F, such that the output of F can have more features for training the entire model. This design is inspired by mini-batch features, which is effective to defend mode collapse as illustrated by previous researches [36], [37].

The target model F is a semantic segmentation model named FCN8s [38] taking an image x as input and outputting a label map Y where $Y = \{Y(i,j)\}$ is a matrix with the same size as x, and each element $Y(i,j)$ is the label of corresponding pixel in x. Let y_t denote the label of private objects in x. The goal of our image synthesis is to maintain the non-private objects. Thus the classification result maps of the non-private objects in both the original image x and the generated image x' should be as similar as possible. We divide each image into a private part and a non-private part via masks M_t and M_o, so that the optimization of privacy protection on the private part and the optimization of utility preservation on the non-private part will not have much impacts on each other. Particularly, $M_t = \{M_t(i,j)\}$ is a 0-1 binary matrix where $M_t(i,j) = 1$ iff $Y(i,j) = y_t$, and $M_o = \mathbf{1} - M_t$ where $\mathbf{1}$ is an all 1 matrix with the same size as M_t.

With the help of F, we formulate the utility loss function to measure the utility loss of generated images x' as follows.

$$\mathcal{L}_{uti}(G) = \log(H(M_o \cdot F(G(x)), M_o \cdot Y)), \quad (5)$$

where $H(\cdot,\cdot)$ is the cross-entropy measurement. Minimizing $\mathcal{L}_{uti}(G)$ is to push generator produce precise noise N such that the non-private objects will not be affected by disturbation.

4) **Privacy Protection:** The generated noise N is added on image x to protect privacy. We define $\mathcal{L}_{pri}(G)$ as the measurement of "privacy loss" that indicates the performance of privacy protection. To reduce recognition accuracy of the private objects, we intend to turn those private objects to their least likely classes, so that they can not be used to perform location inference attack.

Formally, let $Y_l = \arg\min(F(x))$, in which Y_l is a matrix and each element of Y_l is the least likely class of the corresponding pixel in x. That is, Y_l indicates the most impossible classification result map of the original image x. Then, $\mathcal{L}_{pri}(G)$ can be calculated by:

$$\mathcal{L}_{pri}(G) = \log(H(M_t \cdot F(G(x)), M_t \cdot Y_l)). \quad (6)$$

Finally, the loss function of ADGAN-I model can be formulated as a combination of Eq. (4), Eq. (5), and Eq. (6), *i.e.*,

$$\begin{aligned}\mathcal{L}_{ADGAN-I} = &\sum_{i=1}^{k}[\mathbb{E}_{x\sim I}[\log D_i(x)] + \mathbb{E}_{x\sim I}[\log(1 - D_i(G(x)))]] \\ &+ \lambda_1 \mathbb{E}_{x\sim I} \log(H(M_o \cdot F(G(x)), M_o \cdot Y)) \\ &+ \lambda_2 \mathbb{E}_{x\sim I} \log(H(M_t \cdot F(G(x)), M_t \cdot Y_l)),\end{aligned} \quad (7)$$

where λ_1 and λ_2 are hyper-parameters that are used to adjust the scale of utility loss and privacy loss.

C. Details of ADGAN-II Model

Different from ADGAN-I that uses a "U-Net"-based network as the generator, in ADGAN-II we exploit the variational auto-encoder (VAE) to design a new generator, which is motivated by the following considerations. First of all, in the data generation branch of deep learning, GAN and VAE are the most powerful and useful basic generative structures. Integrating the two structures as a whole one is promising to further enhance the performance of data generation, which is the goal of ADGAN-II. Besides, with the design of ADGAN-I and ADGAN-II, the performance of U-Net based generator and VAE based generator in street view obfuscation can be clearly compared. Moreover, another advantage of ADGAN-II is its flexibility and extensibility. In ADGAN-I, the privacy-preserving result $x' = x + N$ is obtained via the original data x and the generated noise N. That is, the original data x is necessary to get the generated data in ADGAN-I, which may restrict the extensibility of ADGAN-I in practice. While, technically speaking, with the help of VAE's decoder in ADGAN-II, the results of generator can be produced by a latent vector z without any original data x, which makes ADGAN-II more flexible and extendable in real applications. Although such a functionality has not been fully implemented in the existing works, it is a very interesting topic in our future research to investigate the generation of complex street view data from specific low dimension vectors.

1) **Generator Module:** A generative VAE model [39], [40] takes an input x and outputs its generated x' by optimizing maximum likelihood expressed as:

$$\mathcal{L}(\phi, \theta, x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - KL(q_\phi(z|x)\|p(z)), \quad (8)$$

where ϕ and θ are respectively the encoder and the decoder of VAE, z is a predefined low-dimensional vector, and $KL(\cdot||\cdot)$ is the Kullback-Leibler divergence between $q(z|x)$ and $p(z)$.

According to our objectives (ii) and (iii), we only need to reconstruct the non-private objects for utility preservation while remaining the private part un-recognizable for privacy protection. However, the traditional VAE is not capable of reconstructing such complex street view images. To provide more information for the reconstruction process in the decoder, we keep the encoder process unchanged but inject more knowledge about Y as conditional information into the decoder.

By defining $Y' = M_o \times Y$, the conditional information Y' has the same size as Y but only contains the information about non-private objects. Thus, during learning process, the conditional VAE can recover most details for image utility without privacy leakage. The loss function for our conditional VAE-based generator module can be formulated as:

$$\mathcal{L}_{cVAE} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z, Y')] - KL(q_\phi(z|x)\|p(z)). \quad (9)$$

From our general representation in Eq. (1), the loss function of generator in ADGAN-II can be expressed as the utility loss function $\mathcal{L}_{uti}(G) = -\mathcal{L}_{cVAE}$.

2) **Discriminator Module:** The model configuration and loss function of the discriminator module in both ADGAN-I and ADGAN-II are the same, which is illustrated in Section III-B2.

3) **Privacy Protection:** Recall that in the conditional VAE generator, the private information has been removed from auxiliary information Y'. To preserve more privacy, privacy loss function, $\mathcal{L}_{pri}(G)$, is added to differ those private objects from the original images, which is given by Eq. (10).

$$\mathcal{L}_{pri}(G) = -\log(H(M_t \cdot F(G(x)), M_t \cdot Y)). \quad (10)$$

To sum up, the loss function of ADGAN-II can be expressed by Eq. (4), Eq. (9), and Eq. (10) in the following equation:

$$\begin{aligned}\mathcal{L}_{ADGAN-II} = \sum_{i=1}^{k}[\mathbb{E}_{x\sim I}[\log D_i(x)] + \mathbb{E}_{x\sim I}[\log(1 - D_i(G(x)))]] \\ + \gamma_1 \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z, Y')] - KL(q_\phi(z|x)\|p(z)) \\ + \gamma_2 \mathbb{E}_{x\sim I} \log(H(M_t \cdot F(G(x)), M_t \cdot Y)),\end{aligned} \quad (11)$$

where γ_1 and γ_2 are scale parameters to adjust loss function.

Remarks: Both ADGAN-I and ADGAN-II can generate privacy-preserving images and videos. Compared with images, additional pre-process and post-process are needed to deal with videos, which is demonstrated in Section IV-D. Moreover, in ADGAN-I and ADGAN-II, the private object class can be customized according to different application requirements. The masks M_t and M_o can be set flexibly to satisfy various needs and then used to retrain the models for data generation.

D. Differences between ADGAN-I and ADGAN-II

The main differences between ADGAN-I and ADGAN-II lie in two aspects.

(i) Model Structure. ADGAN-I uses a "U-Net"-based structure, while ADGAN-II uses a conditional VAE based-generator. As a result, ADGAN-II needs more additional condition to generate images, which causes the different utility loss function. Concretely, ADGAN-I can directly use cross-entropy loss and masks, M_t and M_o, to define its utility, and ADGAN-II needs to use loss function of the conditional VAE to define the loss function.

(ii) Data Generation. In ADGAN-I, the generator G only generates a noise $N = G(x)$. The final synthetic image x' is actually produced by $x' = x + G(x)$. Adding such an image-dependent noise into the original image, x, may not change the quality of x' too much, because the original image x contains much more information. While in ADGAN-II, the synthetic images are generated directly by a transformation $x' = G(x)$. The original image, which is input to the cVAE structure, passes through the encoding and decoding process, which may lead to loss of original image information as well as low image quality. Therefore, the image quality of ADGAN-I would be better than that of ADGAN-II.

IV. Experiments

In this section, extensive experiments are conducted to validate the effectiveness of our two models.

A. Experiment Setup

1) **Datasets:** To demonstrate the performance of utility preservation and privacy protection of our ADGAN-I and

ADGAN-II models, two different vehicular camera datasets are used in our experiments.

- *Cityscapes Dataset* [41], which has pairs of images and extra video data for training and evaluating.
- *Google Street View Dataset* [42], which contains street view images covering 10,343 related place-mark in USA. Every 6 of the images belong to a place-mark so that they can be used to perform location inference detection.

Due to the constraints of hardware, we downscale the resolution of above two datasets into 512×512. After the results are generated, we recover them back to higher resolution 1024×512 for better visualization.

2) **Location Inference Attack:** In location inference attack, the adversaries can obtain street view data as prior knowledge from Google Map API or other resources and thus can identify the real location of a given image or video. First, the adversaries collect enough camera data and extract features of the data. Next, at the attack stage, the adversaries acquire camera data from victim vehicles' storage and/or vehicles' remote servers and perform feature extraction. Finally, learning or matching methods can be used to detect the location where the victim image/video was taken. More details about location inference attack can be referred to [42]. Therefore, to defend such attack, it is critical to obfuscate the feature extraction process of private objects in the original data. In our two models, our goal is to conceal the location-related private objects and preserve other useful objects in the generated data.

3) **Baseline Model:** To our best knowledge, [4] is the only existing work that studies location inference attack towards images in autonomous vehicles. Hence, its proposed mechanism PPAD is adopted as baseline for performance comparison.

PPAD aims to protect location information for privacy preservation. Briefly speaking, the significant differences between PPAD and our two models are: (i) In PPAD, Structure Similarity Index Measurement (SSIM) and L_1 distance are used as utility and privacy metrics, respectively, which partially hinder their performance. (ii) Different from PPAD that only has generator and discriminator, our two models also have a target model. The target model F provides interactive feedback to training process, which is the reason why our two models outperform PPAD as shown in Fig. 5 and Fig. 6.

B. Analysis of Utility and Privacy

To quantitatively evaluate utility and privacy of our proposed methods, the pixel accuracy (PA) and interaction over union (IoU) are selected from FCN-scores [34] to measure the object detection accuracy on images. In the experiments, the segmentation model is run to calculate the values of PA and IoU for all images generated by PPAD and our ADGAN-I and ADGAN-II. Based on PA and IoU, three types of metric are defined for performance comparison, including: (i) *quality*, which is the average PA and the average IoU over the entire images; (ii) *privacy*, which is the average PA and average IoU over the private objects (*i.e.,* background constructions in our experiments); (iii) *utility* that is the average PA and IoU over the non-private objects. For quality and utility, a higher value means a better performance; while for privacy, the lower value the better performance.

TABLE II
FCN-SCORES COMPARISON OF 3 MODELS ON CITYSCAPES

Model	PPAD	ADGAN-I	ADGAN-II
Quality PA	76.52%	80.93%	71.91%
Privacy PA	64.65%	11.67%	16.21%
Utility PA	81.96%	84.93%	78.97%
Quality IoU	22.36%	28.80%	22.01%
Privacy IoU	11.75%	7.45%	8.71%
Utility IoU	29.93%	35.56%	29.65%

TABLE III
FCN-SCORES COMPARISON OF 3 MODELS ON GOOGLE STREET VIEW

Model	PPAD	ADGAN-I	ADGAN-II
Quality PA	72.31%	79.87%	70.70%
Privacy PA	62.82%	13.54%	11.65%
Utility PA	78.97%	83.05%	77.54%
Quality IoU	18.37%	24.37%	17.40%
Privacy IoU	13.40%	6.27%	4.72%
Utility IoU	21.89%	28.81%	21.05%

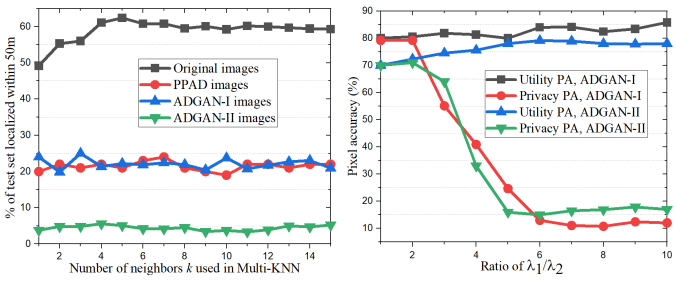
The results of our two models and the baseline model are presented in TABLE II and TABLE III. For Cityscapes dataset, the results of TABLE II are analyzed below.

(i) Our ADGAN-I model achieves the highest quality PA and quality IoU which indicate that the quality of entire images generated by ADGAN-I is the best, and ADGAN-II and PPAD models have comparable quality PA and quality IoU. The imperceptible modification plays an important role in privacy protection and visualization, for which the results in Fig. 5 also demonstrates the superior image quality of ADGAN-I.

(ii) PPAD, ADGAN-I, and ADGAN-II receive lower privacy PA than their corresponding quality PA to protect the private objects in the images. PPAD's PA is decreased from 76.52% to 64.65%, which means there is still more than 60% chance for attackers to identify the private objects. While, the values of PA in ADGAN-I and ADGAN-II are reduced to 11.67% and 16.21%, respectively. Compared with PPAD, privacy PA of ADGAN-I and ADGAN-II is much more lower, which reflects that our two models outperform PPAD in terms of privacy protection. Through observing quality IoU and privacy IoU of the three models, the same conclusions can be drawn.

(iii) For utility preservation, the performance of ADGAN-I is also the best, and the performance of ADGAN-II and PPAD is comparable. In ADGAN-I, utility PA is 84.93%, and utility IoU is 35.56%, which are higher than those of ADGAN-II and PPAD. The higher utility PA (and utility IoU) means more information can be preserved in the generated images.

From the above results, we have more insights into the fundamental difference between our two models and PPAD. In our ADGAN-I and ADGAN-II models, the target model F can give real-time feedbacks on which direction we should move on for performance enhancement; while in PPAD, no component can do that. The highest image quality of ADGAN-I benefits from the specific image-dependent noise that is produced and added without changing the image structure. Differently, ADGAN-II and PPAD compress the original images



(a) Localization accuracy w.r.t k (b) Pixel accuracy w.r.t. λ_1/λ_2

Fig. 4. Performance comparison of ADGAN-I and ADGAN-II.

into low-dimension vectors and then recover them to full-size images, leading to the information loss during transformation.

The results for Google Street View dataset is shown in TABLE III, where there is not segmentation label in the dataset. We utilize the semi-supervised idea with pre-trained FCN8s to process data augmentation: we first fetch a few manually labeled images to train a FCN segmentation model and then use the FCN to produce a number of labeled data as additional dataset. From TABLE III, we can obtain the conclusions similar to those from TABLE II.

Next, we conduct Multi-KNN [42] on the original images and the synthetic images generated by PPAD and our ADGAN-I and ADGAN-II to examine the performance of privacy protection for real locations. Note that Multi-KNN quantifies how many percentages of images can be localized within 50m range of their real locations. The location detection performance is shown in Fig. 4(a), where k is the number of nearest neighbors used in Multi-KNN. For the original images, the percentage of images that can be localized within 50m of their real locations increases with the growth of k at the beginning, and this percentage maintains at around 60% percent when $k \geq 4$, which means the leakage of private location becomes more serious when there are more neighbors and keeps stable when the number of neighbors is sufficient for location detection. For PPAD and ADGAN-I, only around 21% percent of images are localized within 50m no matter the change of k. Since the image quality of ADGAN-I is better than that of PPAD, such 21% accuracy is quite acceptable. Our ADGAN-II achieves the lowest percentage about 4%-5% implying a more effective performance in generating privacy-preserving images to resist location inference attack, for which the reason is that the image perturbation yielded by ADGAN-II is the most significant.

Moreover, to balance the tradeoff between utility preservation and privacy protection, we implement our models with different hyper parameter ratio λ_1/λ_2, where λ_1 and λ_2 are the scaling parameters in Eq. (1) to control "utility loss" $\mathcal{L}_{uti}$ and "privacy loss" $\mathcal{L}_{pri}$, respectively. In Fig. 4(b), we fix λ_1=100 and increase λ_1/λ_2 by reducing λ_2 gradually. For ADGAN-I, as λ_1/λ_2 becomes larger, the utility increases first and then keeps almost stable after $\lambda_1/\lambda_2 = 6$. This is because λ_1 enlarges the proportion of $\mathcal{L}_{uti}$ to preserve more image utility. When $\lambda_1/\lambda_2 > 6$, the improvement of image utility reaches an upper bound because of the existence of $\mathcal{L}_{pri}$. On the contrary, with the increase of λ_1/λ_2, the privacy level first goes down and then gradually stable after λ_1/λ_2=6. The reduction of privacy level before the turning point means that "privacy loss" $\mathcal{L}_{pri}$ is still in charge of whole loss function. When λ_2 is smaller enough, λ_1 can finally influence private pixels, which causes the stability of privacy protection. For ADGAN-II, the change trends of utility PA and privacy PA are similar to those of ADGAN-I. Utility PA increases first and then reaches a stable upper bound because enlarging the weight of $\mathcal{L}_{uti}$ can improve image utility. Meanwhile, privacy PA keeps dropping before $\lambda_1/\lambda_2 = 5$ due to the control of $\mathcal{L}_{pri}$. After the turning point, when $\mathcal{L}_{uti}$ is getting more weight, privacy PA slightly climbs to a stable value. According to the above observations, setting $\lambda_1/\lambda_2 = 6$ for ADGAN-I and $\lambda_1/\lambda_2 = 5$ of ADGAN-II can achieve the most effective tradeoff between utility and privacy.

TABLE IV
SSIM MEASUREMENT ON CITYSCAPES AND GOOGLE STREET VIEW

Model	Cityscapes	Google Street View
PPAD	0.6925	0.6610
ADGAN-I	0.8704	0.9013
ADGAN-II	0.6211	0.6305

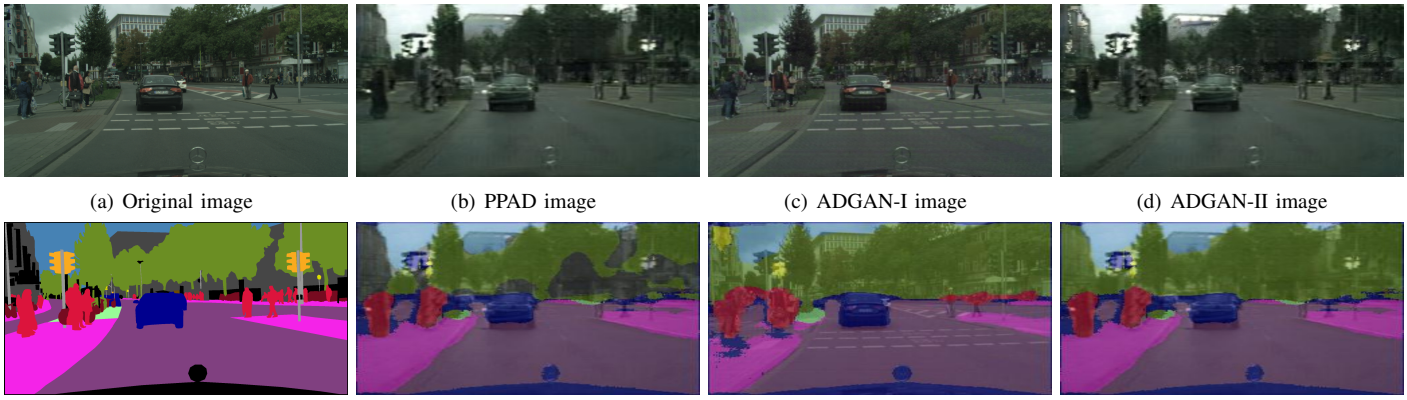
Summary of Analysis. The experiment results show that preserving image utility and protecting location privacy are hard to be achieved simultaneously. On one hand, a higher recognition utility benefits the images for real applications, such as object detection and data mining, but exposes the images to severe privacy threats, *e.g.*, location inference attack. On the other hand, a better privacy protection is realized at the cost of recognition utility, pushing the synthesized images to be unclear and even useless. Nevertheless, our ADGAN-I and ADGAN-II models can provide an effective tradeoff between utility preservation and privacy protection. Furthermore, the experiment results of ADGAN-I and ADGAN-II are consistent with the analysis in Section III-D.

C. Perception Comparison

To visualize the perceptual effectiveness (*i.e.,* image recognition utility) of our two models, we provide SSIM measurement, image quality visualization, and semantic segmentation comparison as well.

1) **SSIM Measurement:** Structure Similarity Index Measurement (SSIM) measures perceptual accuracy, which is very close to human visibility. It can be computed by $SSIM(x, x') = l(x, x')^{\alpha} \cdot c(x, x')^{\beta} \cdot s(x, x')^{\gamma}$ where $l(x, x')$ is luminance similarity, $c(x, x')$ is contrast similarity, and $s(x, x')$ is structural similarity [43]. The output of SSIM is a number in $[0, 1]$ and represents the similarity between x and x', where 0 means totally different and 1 means exactly same. The SSIM results are listed in TABLE IV, we can see that our ADGAN-I model achieves 0.8704 and 0.9013 on Cityscapes and Google Street View, respectively, which is the best performance among the three models with privacy consideration. Besides, the SSIM values of PPAD and our ADGAN-II model are comparable.

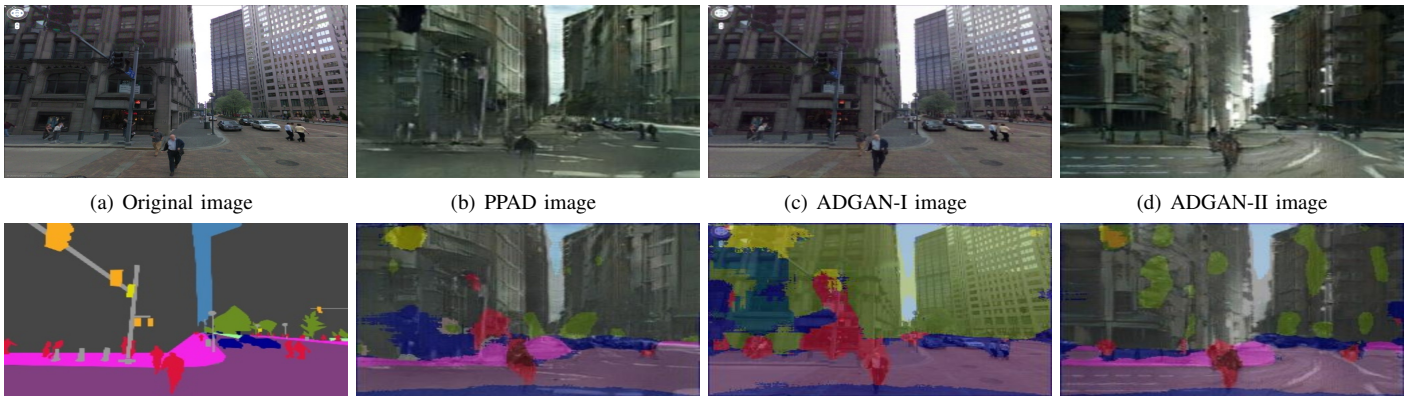
2) **Image Quality Visualization:** In Fig. 5 and Fig. 6, we display some random samples generated from PPAD,



(a) Original image (b) PPAD image (c) ADGAN-I image (d) ADGAN-II image

(e) Ground truth label (f) PPAD result (g) ADGAN-I result (h) ADGAN-II result

Fig. 5. Visual quality and semantic segmentation comparison on Cityscapes dataset. The first row is original image and generated images of PPAD, ADGAN-I, and ADGAN-II. The second row is ground truth label and segmentation results of PPAD, ADGAN-I and ADGAN-II.



(a) Original image (b) PPAD image (c) ADGAN-I image (d) ADGAN-II image

(e) Ground truth label (f) PPAD result (g) ADGAN-I result (h) ADGAN-II result

Fig. 6. Visual quality and semantic segmentation comparison on Google Street View dataset. The first row is original image and generated images of PPAD, ADGAN-I, and ADGAN-II. The second row is ground truth label and segmentation results of PPAD, ADGAN-I, and ADGAN-II.

ADGAN-I and ADGAN-II together with the corresponding original images. We have the following critical findings: (i) compared with PPAD and ADGAN-II, the synthetic images of ADGAN-I are more similar to the original images with better image quality; (ii) compared with ADGAN-I, it is harder to tell backgrounds from the synthetic images of ADGAN-II, which is the reason why ADGAN-II can better protect location privacy; (iii) the synthetic images of PPAD and ADGAN-II look similar, meaning their comparable performance in image generation; These observations confirm that the superiority of our ADGAN-I and ADGAN-II models over PPAD for utility preservation and privacy protection, which is consistent with the analysis in Section IV-B.

3) **Semantic Segmentation Comparison:** For computer vision of autonomous vehicles, semantic segmentation is the most important application. To investigate whether effective semantic segmentation can be achieved under the requirement of privacy protection, we run semantic segmentation on the synthetic images generated from the three models. The segmentation results are compared in Fig. 5 and Fig. 6, from which we can see that our ADGAN-I and ADGAN-II models outperform PPAD in terms of data utility. Meanwhile, the background constructions are hard to be segmented correctly by computer (see the results of ADGAN-I) or human (see the results of ADGAN-II), which corresponds to our conclusions from TABLE II and TABLE III.

More examples of images are presented at: https://www.dropbox.com/sh/zsyaoch6exzvu42/AABQOyu2uPb8uwYTt1mG4u5Na?dl=0.

D. Video Evaluation

Furthermore, in order to show the advantages of our proposed models in protecting location privacy in videos, we implement our models on video data from Cityscapes dataset and present the results at https://www.dropbox.com/sh/zsyaoch6exzvu42/AABQOyu2uPb8uwYTt1mG4u5Na?dl=0. The process of applying our ADGAN-I and ADGAN-II on video data is similar to that on static image data except some differences in pre-processing and post-processing stages. First, we split the video data into sequential frames on the most fine-grained level to get as much training data as possible. Those sequential frames are feed into network structures with batch size 4 as their order in original video. In this way, the continuous features of sequential data are grouped into same batch, which can adapt video data. At the evaluation stage, the generated batch results are combined

together to get an averaged single frame. Finally, all generated frames are concatenated to compose a complete video form with fps=24, which is easy for human observation. The results of videos are similar to the images that we show in Section IV-C, which clearly validate the effectiveness of our methods in processing image and video data.

V. Conclusion and Future Work

To prevent location privacy leakage from camera data of autonomous vehicles in off-line applications, in this paper we develop two novel methods, ADGAN-I and ADGAN-II, which employ GAN to generate privacy-preserving images and videos while retaining recognition utility. Two real datasets are utilized to evaluate the performance of ADGAN-I and ADGAN-II, and comprehensive comparisons with the state-of-the-arts are conducted to confirm the advantages of our models. In our future work, research activities will be carried out along two major directions: (i) investigate privacy protection for higher resolution data; (ii) deeply study efficient solutions for real-time applications.

Acknowledgment

This work was partly supported by the National Science Foundation of U.S. (1704287, 1829674, 1912753, and 2011845) and the Dean-ship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No. 2019/01/10411.

References

[1] A. C. MADRIGAL, "Inside waymo's secret world for training self-driving cars," The Atlantic, Tech. Rep., 2017. [Online]. Available: https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/

[2] C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, and G. Wu, "Differential privacy preserving in big data analytics for connected health," *Journal of medical systems*, vol. 40, no. 4, p. 97, 2016.

[3] C. Lin, P. Wang, H. Song, Y. Zhou, Q. Liu, and G. Wu, "A differential privacy protection scheme for sensitive big data in body sensor networks," *Annals of Telecommunications*, vol. 71, no. 9-10, pp. 465–475, 2016.

[4] Z. Xiong, W. Li, Q. Han, and Z. Cai, "Privacy-preserving auto-driving: A gan-based approach to protect vehicular camera data," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 668–677.

[5] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. Citeseer, 2007, pp. 1–7.

[6] A. Feryanto and I. Supriana, "Location recognition using detected objects in an image," in *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. IEEE, 2011, pp. 1–4.

[7] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1319–1328.

[8] Y. Wu, F. Yang, Y. Xu, and H. Ling, "Privacy-protective-gan for privacy preserving face de-identification," *Journal of Computer Science and Technology*, vol. 34, no. 1, pp. 47–60, 2019.

[9] N. Raval, A. Machanavajjhala, and L. P. Cox, "Protecting visual secrets using adversarial nets," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1329–1332.

[10] F. Pittaluga, S. Koppal, and A. Chakrabarti, "Learning privacy preserving encodings through adversarial training," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 791–799.

[11] J. Chen, J. Konrad, and P. Ishwar, "Vgan-based image representation learning for privacy-preserving facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1570–1579.

[12] M. Nasr, R. Shokri, and A. Houmansadr, "Machine learning with membership privacy using adversarial regularization," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2018, pp. 634–646.

[13] M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," *arXiv preprint arXiv:1610.06918*, 2016.

[14] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Generative adversarial privacy," *arXiv preprint arXiv:1807.05306*, 2018.

[15] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 417–425.

[16] Q. Han, Z. Xiong, and K. Zhang, "Research on trajectory data releasing method via differential privacy based on spatial partition," *Security and Communication Networks*, vol. 2018, 2018.

[17] I. Ullah, M. A. Shah, A. Wahid, A. Mehmood, and H. Song, "Esot: a new privacy model for preserving location privacy in internet of things," *Telecommunication Systems*, vol. 67, no. 4, pp. 553–575, 2018.

[18] Q. Miao, W. Jing, and H. Song, "Differential privacy–based location privacy enhancing in edge computing," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 8, p. e4735, 2019.

[19] L. Rongxing, L. Xiaodong, and S. Xuemin, "Spring: A social-based privacy-preserving packet forwarding protocol for vehicular delay tolerant networks," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[20] R. Lu, X. Lin, X. Liang, and X. Shen, "Sacrificing the plum tree for the peach tree: A socialspot tactic for protecting receiver-location privacy in vanet," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*. IEEE, 2010, pp. 1–5.

[21] X. Lin, R. Lu, X. Liang, and X. Shen, "Stap: A social-tier-assisted packet forwarding protocol for achieving receiver-location privacy preservation in vanets," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 2147–2155.

[22] G. Elsayed, S. Shankar, B. Cheung, N. Papernot, A. Kurakin, I. Goodfellow, and J. Sohl-Dickstein, "Adversarial examples that fool both computer vision and time-limited humans," in *Advances in Neural Information Processing Systems*, 2018, pp. 3910–3920.

[23] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236*, 2016.

[24] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1765–1773.

[25] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131–151, 2016.

[26] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1168–1174, 2008.

[27] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2013, pp. 208–213.

[28] T. Orekondy, M. Fritz, and B. Schiele, "Connecting pixels to privacy and utility: Automatic redaction of private information in images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8466–8475.

[29] T. Orekondy, B. Schiele, and M. Fritz, "Towards a visual privacy advisor: Understanding and predicting privacy risks in images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3686–3695.

[30] P. Liu, Y. Li, Y. Jiang, and S.-T. Xia, "Visual privacy protection via mapping distortion," *arXiv preprint arXiv:1911.01769*, 2019.

[31] Z. Shen, S. Fan, Y. Wong, T.-T. Ng, and M. Kankanhalli, "Human-imperceptible privacy protection against machines," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1119–1128.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[33] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[34] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
[35] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.
[36] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in neural information processing systems*, 2016, pp. 2234–2242.
[37] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.
[38] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
[39] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
[40] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in neural information processing systems*, 2015, pp. 3483–3491.
[41] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.
[42] A. R. Zamir and M. Shah, "Image geo-localization based on multiplenearest neighbor feature matching usinggeneralized graphs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 8, pp. 1546–1558, 2014.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

Zuobin Xiong received the B.S. degree from the Department of Mathematics, Northeast Forestry University, Harbin, Heilongjiang, China in 2016, and the M.S. degree from College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang, China in 2019. Currently, he is pursuing the Ph.D. degree at the Department of Computer Science, Georgia State University, Atlanta, GA, USA. His research interests lie in the area of Machine Learning, Data Mining, Internet of Things (IoT) along with the privacy and security issues of above related topics.

Zhipeng Cai (SM'06) is currently an Associate Professor at Department of Computer Science, Georgia State University, USA. He received his PhD and M.S. degrees in the Department of Computing Science at University of Alberta, and B.S. degree from Beijing Institute of Technology. Prior to joining GSU, Dr. Cai was a research faculty in the School of Electrical and Computer Engineering at Georgia Institute of Technology. Dr. Cai's research areas focus on Internet of Things, Machine Learning, Cyber-Security, Privacy, Networking and Big data. Dr. Cai is the recipient of an NSF CAREER Award. He served as a Steering Committee Co-Chair and a Steering Committee Member for WASA and IPCCC. Dr. Cai also served as a Technical Program Committee Member for more than 20 conferences, including INFOCOM, ICDE, ICDCS. Dr. Cai has been serving as an Associate Editor-in-Chief for Elsevier High-Confidence Computing Journal (HCC), and an Associate Editor for more than 10 international journals, including IEEE Internet of Things Journal (IoT-J), IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Transactions on Vehicular Technology (TVT).

Qilong Han is a Professor and Deputy Dean in the College of Computer Science and Technology, Harbin Engineering University. His research interests include data security and privacy, mobile computing, distributed and networked systems. He has more than 60 publications as edited books and proceedings, invited book chapters, and technical papers in refereed journals and conferences. He is a senior member of CCF, and the Chair of CCF YOCSEF Harbin. He has served as program committee members and co-chairs of a number of international conferences/workshops for areas including web intelligence, e-commerce, data mining, intelligent systems, etc.

Arwa Alrawais (M'19) received the M.S. degree in computer science and the Ph.D. degree from the Department of Computer Science, The George Washington University, Washington, DC, USA, in 2011 and 2017, respectively. She holds a patent in system and method for remote authentication with dynamic usernames. Her current research interests include network security, wireless and mobile security, and algorithm design and analysis. She serves as a Professional Reviewer for several conferences and journals of the IEEE and ACM.

Wei Li (M'16) is currently an Assistant Professor in the Department of Computer Science at Georgia State University. Dr. Li received her Ph.D. degree in computer science, from The George Washington University, in 2016 and M.S. degree in Computer Science from Beijing University of Posts and Telecommunications, in 2011. She won the Best Paper Awards in ACM MobiCom Workshop CRAB 2013 and international conference WASA 2011, respectively. Her current research spans the areas of blockchain technology, security and privacy for the Internet of Things and Cyber-Physical Systems, secure and privacy-aware computing, Big Data, game theory, and algorithm design and analysis. She is a member of IEEE and a member of ACM.