IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 September 2022, accepted 7 November 2022, date of publication 17 November 2022,
date of current version 1 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222834

RESEARCH ARTICLE

DeepClean: A Robust Deep Learning Technique for Autonomous Vehicle Camera Data Privacy

OLAYINKA ADEBOYE[1], TOOSKA DARGAHI[1,2], MEISAM BABAIE[3], MOHAMAD SARAEE[1], AND CHIA-MU YU[4], (Senior Member, IEEE)

[1]School of Science, Engineering and Environment, University of Salford, M5 4WT Manchester, U.K.
[2]Computing and Mathematics Department, Manchester Metropolitan University, M15 6BH Manchester, U.K.
[3]School of Mechanical Engineering, University of Leeds, LS2 9JT Leeds, U.K.
[4]Department of Information Management and Finance, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

Corresponding author: Olayinka Adeboye (o.a.adeboye@edu.salford.ac.uk)

This work was supported in part by the U.K. Royal Society Award, under Grant IEC\R3\183047, and in part by the Ministry of Science and Technology (MOST), Taiwan, under Grant MOST 110-2636-E-009-018 and Grant 110-2927-I-009-510.

ABSTRACT Autonomous Vehicles (AVs) are equipped with several sensors which produce various forms of data, such as geo-location, distance, and camera data. The volume and utility of these data, especially camera data, have contributed to the advancement of high-performance self-driving applications. However, these vehicles and their collected data are prone to security and privacy attacks. One of the main attacks against AV-generated camera data is location inference, in which camera data is used to extract knowledge for tracking the users. A few research studies have proposed privacy-preserving approaches for analysing AV-generated camera data using powerful generative models, such as Variational Auto Encoder (VAE) and Generative Adversarial Network (GAN). However, the related work considers a weak geo-localisation attack model, which leads to weak privacy protection against stronger attack models. This paper proposes DeepClean, a robust deep-learning model that combines VAE and a private clustering technique. DeepClean learns distinct labelled object structures of the image data as clusters and generates a more visual representation of the non-private object clusters, e.g., roads. It then distorts the private object areas using a private Gaussian Mixture Model (GMM) to learn distinct cluster structures of the labelled object areas. The synthetic images generated from our model guarantee privacy and resist a robust location inference attack by less than 4% localisation accuracy. This result implies that using DeepClean for synthetic data generation makes it less likely for a subject to be localised by an attacker, even when using a robust geo-localisation attack. The overall image utility level of the generated synthetic images by DeepClean is comparable to the benchmark studies.

INDEX TERMS Autonomous vehicle, data privacy, data utility, deep clustering, generative model.

I. INTRODUCTION

Autonomous vehicles (AV) onboard sensors generate diverse datasets [1]. These datasets include camera data (for example, images and videos of street views showing road objects in a city), distance data from Lidar and Radar sensors, and Global Positioning Systems (GPS) trajectory data. The captured datasets are required for several functional and non-functional processes [2]. For instance, the captured visual images and videos can be used for accident claims and training auto-driving deep learning models (e.g., for object detection and recognition [3], [4], [5]). Also, real-time data analysis on in-vehicle data is used for performance evaluation purposes [6], [7]. This rich dataset could be held inside the vehicle or sent to external storage, such as Cloud [8].

One of the main concerns regarding AV-generated data is users' privacy [9]. Camera data contain several visual and context-rich features that can be extracted and geo-localised. Several studies have shown how over-needed location information in images, such as background buildings, landmarks, road signs and markings, and surrounding vegetation, improve image matching and geo-localisation [10]. Suppose we assume an attacker can get unauthorised access to the stored camera data in the internal or external storage.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

FIGURE 1. Example to show a geo-localisation attack. (a) Camera data of target's vehicle near Queen's Tower, (b) Camera data of target's vehicle near Boston Market, (c) Leaked trajectory information. All the images are extracted from Berkeley AV Open-source data [8].

In that case, she can perform a location inference attack using geo-localisation techniques. This attacker may be able to infer sensitive information, such as the user's home/work address and past/future travel patterns, which leads to a location privacy breach.

Figure 1 (c) shows an example of a location inference attack. If an attacker has access to both Figure 1 (a) and (b), she can learn that the target's vehicle has passed through Queen's Tower in Figure 1 (a). By getting access to more images and videos from the target's vehicle with timestamp correlations in Figure 1 (b), she can perform geo-localisation and predict the trajectories (Figure 1 (c) is an example).

Researchers have proposed several ways of distorting private objects in a dataset to mitigate location inference attacks on AV-generated camera data. Recently, *Xiong* et al. proposed ADGAN, in which they use Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) to generate privacy-preserving camera datasets [11]. They have considered a weak attack model under Multi-KNN (i.e., multiple k nearest neighbour) feature matching geo-localisation approach. Multi-KNN [12] was also used in other research studies for geo-localisation, such as [13] and [14].

Schindler et al. organised image features as a bag of words and arranged them in a vocabulary tree for image matching [13]. Their approach is inefficient for processing large image features and computationally too slow. *Zamir* et al. improved the computational efficiency of feature matching using a generalised minimum clique problem [12]. However, their formulation of a fixed nearest neighbour selection algorithm limits the number of matching features and hence does not allow for image matching improvement.

Zemene et al. [15] designed a more robust geo-localisation system to localise street view images with higher performance compared to the previous studies, and other image geo-localisation approaches [16], [17].

The improved feature matching approach in [15] is based on returning a dynamic nearest neighbour of the reference images using dominant set clustering, which outperforms the approaches based on multi-KNN with a fixed value for k. However, improved geo-location estimate increases the image matching performance with the cost of increased potential privacy threats. This motivated us to consider the robust geo-localisation method proposed in [15] as a strong attack model against AV-generated camera data and draw the following research questions: 1) To improve privacy, what features in an image could be manipulated to decrease the similarity between an original image and its distorted version? 2) can we find a privacy-preserving technique for generating synthetic AV-camera data that sufficiently balances the privacy-utility trade-off to suit several data use cases?

In this paper, we propose DeepClean, which answers the above research questions. DeepClean is a deep clustering approach which combines VAE with GMM clustering methods to improve the privacy-utility trade-off. It proposes a solution for learning and controlling the visual representation of objects in an image. We consider two labels for the objects in each image, i.e., private and non-private. Private objects are those that could significantly help in the geo-localisation process, such as buildings, pedestrians, vehicles, and road signs. We use deep clustering to separate (and then distort) those clusters that include private objects while retaining the underlying structure of the non-private areas (e.g., roads). The GMM clustering method is used for learning clusters of objects in high-dimensional image data that are well-separated to enforce the privacy/utility requirements.

DeepClean uses the VAE data generation technique to produce high-dimensional image samples without directly operating on original data. The VAE approach is flexible for 3D street view models and traffic analysis applications. A similar work, ADGAN II [11], also adopts the VAE approach to improve the data generation performance from distributional assumptions, while UNIT in ADGAN I [18] uses image-dependent processing. DeepClean utilises an encoder and decoder model. Its encoder model encodes data by partitioning it into object clusters using our private GMM algorithm. A function of the algorithm learns a supervised clustering task and accurately partitions the clusters into private and non-private object parts by using mask binary code as a key. The learned private object clusters are distorted by injecting Gaussian noise into their cluster centres. This approach ensures that we can efficiently preserve privacy in the private cluster areas without affecting too much visual quality of the non-private object areas. DeepClean's decoder model

decodes the resulting high-dimensional feature representation from the encoder network into observable samples using a deep neural network. The model optimisation is achieved by maximising the expected lower bound of the VAE system.

The main contributions of this paper are as follows:

- We propose DeepClean, a privacy-preserving generative technique for AV camera data that combines our private Gaussian Mixture Model (GMM) with a Variational Autoencoder (VAE) to learn high-dimensional feature representations of images as a supervised private/non-private cluster task. Then trains the cluster outputs on a VAE to generate more privacy-protected samples from our model.
- We evaluate the privacy performance of DeepClean on a robust geo-localisation attack (that improves image matching of distorted images with their trained reference images) for location inference resistance.
- Our thorough experiments on real-world publicly available datasets show that DeepClean learns more features in an image, variably controls privacy/utility requirements and generates more privacy-preserved image data compared to the state-of-the-art.

The remainder of the paper is organised as follows. Section II discusses the related work and provides the required background. Section III explains the methodology and the components of DeepClean. Section IV presents the evaluation results of the experiments on image quality, utility and user privacy. Section V concludes the paper and highlights future work directions.

II. RELATED WORK AND BACKGROUND

In recent years, machine learning techniques have been widely utilised and applied along with traditional privacy techniques (such as K-anonymity and differential privacy) to address privacy challenges in data mining, publishing, and storage [19]. A fundamental part of machine learning is clustering, which involves grouping a set of similar objects in clusters [20]. Its application in computer vision tasks, e.g., object detection, face recognition, and image analysis, has been widely studied and has achieved efficient performance. Usually, efficient clustering algorithms are justified by how well they can represent data, typically performed by solving an optimisation problem. However, the more complex the features in an image or video data, the more difficult it becomes to generate a well-structured representation of the data using many existing clustering algorithms [21].

Recent works focused on deep learning-based image clustering approaches for feature representations in an unsupervised setting, which are shown to be more efficient than in supervised settings. For example, in [21], [22], and [23], the data generation process is performed using an unsupervised approach, aiming at learning a joint distribution of images in different domains by using images from the marginal distribution in individual domains. *Yang* et al. represented images using agglomerative clustering and activations of convolutional neural networks [23]. *Hsu* et al. proposed a clustering convolutional neural network to better capture the salient part of an image without providing any bounding boxes in the training stage for a better representation [24]. *Wang* et al. combined Sparse coding base pipeline into deep learning for clustering, achieving an extremely efficient inference process and high scalability of large-scale data [22]. Thus, these methods are only efficient on images with fewer features like the MNIST dataset [25] (the handwritten digits) and do not consider privacy in the image generation process.

The image translation performance of VAE and the GAN models has been remarkable recently. *Liu* et al. proposed an unsupervised image-to-image translation framework based on GAN and VAE, which is called UNIT [26]. These adversarial training objectives interact with a weight-sharing constraint, enforcing a shared latent space to generate corresponding images in two domains. At the same time, VAE relates translated images with images in the respective domain. These methods achieve high-quality image translation results for street-view images and videos. Similarly, in DeepClean, we are also taking advantage of the data generation power of VAE. However, to outperform GAN-based models, we consider a stricter attack model (the geo-localisation approach in [15]) and deliver higher privacy protection.

Recently, *Xiong* et al. [11], [18] were the first to address privacy concerns of auto-driving images and videos. The auto-driving generation neural network (ADGAN I) uses UNIT to generate data and applies noise directly to the original image to produce the synthetic samples [18]. This direct approach gives no flexibility to learn the variations of samples. Moreover, the added noise affects the whole image quality. ADGAN II [11] combines GAN with VAE to better represent street view images. With VAE in ADGAN II, the synthetic samples can now be produced by a latent vector without any original data, making ADGAN II more flexible for real applications, such as the street view image. Generally speaking, GAN-based models may lose perpetual accuracy due to the model collapse property of GAN. For this reason, several methods such as Mean Square Error, Peak Signal-to-noise Ratio, and Structure Similarity Index Measurement are used to achieve high perceptual accuracy [27].

Regarding privacy challenges, robust geo-localisation techniques can variably compute the similarity between the images generated by the methods using discriminative proximity [28]. A few other techniques proposed a more robust data generation utilising the data generative power and useful basic generative structures of VAE with deep neural networks for clustering tasks. *Acs* et al. divided data into clusters using a differentially private clustering approach [25]. Then they gave each cluster a separate Generative Neural Network to train on differentially private gradient descent. The data partitioning into general clusters led to more accurate synthetic samples than training the whole dataset as a single model. A more powerful clustering framework was proposed by [29], which combines VAE and a GMM and maximises the

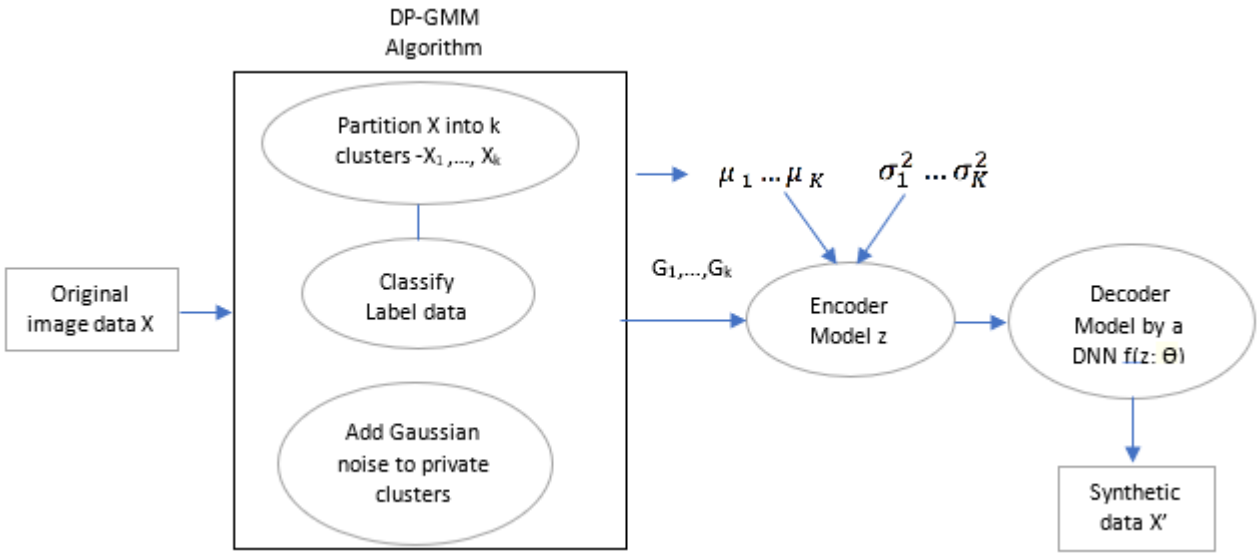

FIGURE 2. Different components of the DeepClean model (including the combination of our differentially private GMM with VAE).

Evidence Lower Bound using Stochastic Gradient Variational Bayes estimator and the reparameterisation trick. DeepClean adopts the data generation technique in [29] and optimises GMM for our specific clustering tasks. The private version of GMM is employed to inject noises in specific cluster centres.

Some approaches to privacy preservation ensure that data features must be selectively distorted to balance privacy/utility trade-offs. In response, Chong [30] proposed a generative adversarial network that aims at reducing privacy risks by removing location-relevant information, e.g., background buildings, from the camera data, before being used for analysis. The location-relevant information in the camera data was analysed and reported as a threat to privacy when providing the data for analysis. Location-relevant information in the camera data was highlighted as a privacy threat to the data. Trajectories of a vehicle could be formed or traced by extracting the location hints from image data and matching them with reference data to geolocate them. However, camera data may also contain other quasi-identifiers (QIDs such as the human face and vehicle plate number) besides location-related ones, putting users' privacy at risk.

To the best of our knowledge, only two research studies (previously explained in this section [11], [14]) addressed location inference threats for AV-generated camera data. Their solution to the problem involves using VAE and GAN-based models to generate privacy-preserving datasets. We argue that using GAN in their approach has two practical limitations. One is that the privacy achieved by the discriminative distance measurement cannot guarantee the location privacy of the image objects. Secondly, a robust geo-localisation tool can exploit the discriminative distance value of the original and distorted images to estimate the geo-location of the target image.

In comparison, DeepClean clusters different parts of an image into private and non-private objects. It then adds noise to specific private objects without affecting the underlying structure of the non-private objects. It achieves a better privacy-utility trade-off compared to the state-of-the-art.

III. PROPOSED APPROACH

In this section, we present DeepClean, a privacy-preserving generative model for AV-generated camera data to address a balanced privacy-utility trade-off in the presence of a potential location privacy threat. This section first explains the considered system and attack model, while Section III-B presents the details of the proposed approach.

A. SYSTEM MODEL

We assume that we have a set of raw camera data, which is generated by an AV. We want to generate a synthetic dataset resilient against location inference attacks. The original camera data is passed to the DeepClean model as an input, and synthetic data is generated as the output. Different components of the DeepClean model are presented in Figure 2. The first component is the labelled DP-GMM algorithm to partition the image into k clusters, learn and predict the labelled clusters, and add Gaussian noise to the learned private object clusters. The output of this component is a noisy partitioned cluster. These clusters are then trained in the encoder $g(x, \phi)$ to produce a latent representation z. A decoder network $f(z; \theta)$ interprets z, such that a synthetic sample can be drawn from the model θ.

Let x be a real camera image such that $x \in I$, where I is a set of raw images from real AV camera data. An image x is fed into the model M consisting of inference and generative processes, and an observable image sample $\hat{x} = M(x)$ is generated. Our private Gaussian mixture model is applied to the sensitive clusters during inference, and the generative model produces a privacy-preserved image $\hat{x}$.

In the inference process, the private GMM component partitions the labelled image objects into k clusters, $X_1, X_2, \ldots, X_k$ where each cluster is a group of similar objects in X. The GMM is trained in a supervised setting to classify the objects in the clusters. Then the GMM trains separately on each cluster; if the cluster is classified as sensitive, Gaussian noise is applied to the cluster centre, else it retains its accurate visual representation (without noise). The VAE encoder trains separately on the cluster outputs and maximises the expected lower bound (ELBO) for optimisation. In the generative process, the decoder, a deep neural network $f(z; \theta)$ decodes the embeddings to an observable, where θ is the parameter of the resulting model.

1) ATTACK MODEL

We consider an attacker or a curious analyst who can access a target's camera data. A vehicle user or vehicle is regarded as a target. A location inference attack can be mounted on the data with or without external multi-source information such as trajectory and distance data. The core task of the attack relies on extracting visual and contextual features, e.g., landmarks, background buildings, surrounding vegetation, and surrounding objects, from query image data. Then the features extracted from the query data are compared with the features of an already trained reference image data of a city or a group of cities (e.g., Google Street View images). If there is a match of features, the geo-localisation system returns the nearest neighbour (NN) image reference with matching features. Then a scheme is used to estimate the location of the most matching NN or even evaluate the location proximity of the multi-NN. A robust geo-localisation system must improve image feature matching and geo-location estimates.

As explained in the introduction section, our attack model is based on the scheme that is proposed by *Zemene* et al. [15]. It uses discriminative values from the image features in the NN selection phase, dominant set clustering for feature matching and constrained dominant set for localising the best matching reference images. This geo-localisation system improves image localisation accuracy by 21% compared to [13] and [12], which are used as the attack model in the related work. We assume that if the attacker can access some AV camera data and, using this sophisticated geo-localisation system [15], she can infer vehicle location information. More so, the attacker can still learn estimated location information from the less privacy-preserved datasets that the state-of-the-art has generated (e.g., ADGAN [11], [14]). Figure 3 shows the matching reference images of a given distorted query image data (ADGAN-generated image). The exact matching image is the nearest neighbour with the most frequent occurrence (the NN with the yellow-coloured ID and frequency of 6). In contrast, the geo-localisation technique in [13] and [12] cannot locate the exact match because of its fixed NN selection constraint.

The attacker is motivated to learn users' sensitive information, such as trajectories that link the target's past travel patterns, places of interest, home/work address, and even predict future journey patterns. The attack resulting from tracking the victim's vehicle could be as severe as physical damage and theft. To control the impact of this attack, privacy analysts would reduce the precision of extracting sensitive features and side-channel location information from the data. A typical privacy-preserving approach would remove or blur sensitive objects, which is not trivial to achieve. However, the data utility for analytics purposes, such as auto-driving navigation analysis, will be affected, and the generated data may become entirely useless.

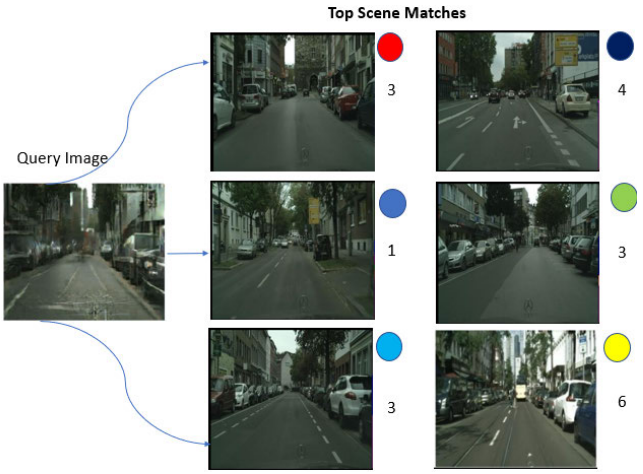

FIGURE 3. Image matching of a distorted query image by the Dominant Set framework [15]. The reference data with the yellow colour ID occurs most frequently.

This creates a challenge in balancing the privacy-utility trade-off. Thus, the transformed data must retain statistical structure in various non-private areas yet preserve the privacy of the private object areas in the data, which is achieved through DeepClean.

B. DeepClean DESCRIPTION

As shown in Algorithm 1, a private GMM partitions X as a mixture of Gaussians with labelled clusters $G_\rho = ((G_1, \rho_1), \ldots, (G_z, \rho_z))$, where we can choose G_i from the mixture component $G(\rho_i)$, such that ρ_i is a labelled image object in the independent and identically distributed (i.i.d) clusters. An output of the private Gaussian mixture partitioning algorithm on X is a cluster G_i. Then computes an estimation of the Gaussian mixture parameters $z \sim \mathcal{N}(\mu_j, \Sigma_j)$. Finally, a DNN model $f(z; \theta)$ takes z as an input with the model parameter θ. Model θ is a privacy-preserving model that can produce synthetic samples.

With the labelled image as input, we run a private version of Principal Component Analysis (PCA) [31] to project onto the top k principal components, containing the means of the components. Our implementation differs because we pass a label sample image as an input. Then privately run the subfunction to individually locate components that find a small ball containing many points. This ensures that all the points generated from a single Gaussian lie in the same cluster.

Algorithm 1 DeepClean: Deep Clustering Generative Model

Require: Image Data : x, # of clusters : k, Bounds on the GMM parameters $w_{min}, \sigma_{min}, \sigma_{max}$, learning parameter α, β, Privacy parameters $\epsilon, \delta > 0$

Ensure: A Privacy-Preserving Model θ

1: $[G_1, G_2, \ldots, G_k] \leftarrow PGMM$ $(x, k, R, w_{min}, \sigma_{min}, \sigma_{max}, \epsilon, \delta)$

2: **for** j from 1 to k **do**

3: $(\mu_j, \Sigma_j) \leftarrow PGE((G_j);\ R, w_{min}, \sigma_{min}, \sigma_{max}, \epsilon, \delta)$; *Comment: Proof of PGE [31];* $\pi_j \leftarrow |G_j| + 2\sqrt{2\ln(1.25/\delta)}/\epsilon$;

4: **end for**

5: set weight such that for all; j, $w_j \leftarrow \pi_j/(\Sigma\pi_j)$

6: $z \leftarrow (\mu_j, \Sigma_j, w_j)_{j=1}^{k}$

7: $\hat{x} \leftarrow f(z, \theta)$

We then estimate the mean and variance of the corresponding Gaussian component privately. Next, we dive into the formal analysis and justification of the version of the algorithm used to design DeepClean.

1) GAUSSIAN MIXTURE MODEL

Assuming the underlying distribution G is a mixture of k Gaussian in high-dimension d, $\{G_i \in R^d\}_{i=1}^{k}$ is a k distinct Gaussian distribution with dimension d. The cluster component G_i is chosen with probability $w_i \in [0, 1]$, and the mean $\mu_i \in R^d$ and variance $\Sigma \in R^{d*d}$ are the parameters of the distributed Gaussian. The mixture can be written as the tuple $\{(w_i, \mu_i, \Sigma_i)\}_{i\in[k]}$. We can accurately recover the tuple $\{(\hat{w}_i, \hat{\mu}_i, \hat{\Sigma}_i)\}i \in [k]$ for a mixture $\hat{G}$. Where $\|\hat{w}-w\|_1$, $\|\hat{\mu}_i - \mu_i\|_{Zi}$, and $\|\hat{\Sigma} - \Sigma\|_{Zi}$ are small for every $i \in [k]$. The vector $\|.\|_Z$ approximately ensures that $\mathcal{N}(\mu_i, \Sigma_i)$ and $\mathcal{N}(\hat{\mu}_i, \hat{\Sigma}_i)$ are close in total variation distance and likewise $\|.\|_1$ ensures the same for comparing the weights.

To learn from the GMM with n samples, independent identically distributed (i.i.d.) samples can be obtained from the mixture D and roughly approximate the parameters of a mixture $\hat{D}$ by a probability $\pi : [k] \longrightarrow [k]$ and satisfying two conditions. One is a separate condition that measures the learning guarantees of the clustering and shows how the clusters are well-separated. In our case, it will ensure that privacy is adequately controlled within the clusters and limit privacy loss due to distributional assumptions. Secondly, certain boundedness of the mixture components is assumed to control the output. Let the separation condition satisfy

$$\forall 1 \leq i < j \leq k, \|\mu_i - \mu_j\|_2 \geq s.max\sigma_i, \sigma_j$$

For $s > 0$ the Gaussian mixture $D \in G(d, k)$ is s-separated. Depending on the number of mixtures and independent of the dimension d. Assuming some large known quantities $R, \sigma_{max}, \sigma_{min}$ such that

$$\forall i \in [k]\|\mu_i\|_2 \leq Rand\sigma_{min}^2 \leq \|\Sigma_i\|_2 \leq \sigma_{max}^2$$

Definition 1((α, β)-Learning): Let the parameters of a Gaussian mixtures $D \in G(d, k)$ be $\{(\mu_1, \Sigma_1, w_1),\ldots, (\mu_k, \Sigma_k, w_k)\}$, an algorithm (α, β)-learns a distribution D and outputs a distribution $\hat{D} \in G(d, k)$ parameterized by $\{(\hat{\mu}_1, \hat{\Sigma}_1, \hat{w}_k),\ldots,(\hat{\mu}_k, \hat{\Sigma}_k, \hat{w}_k)\}$, with a probability of at least $1 - \beta$ and a permutation $\pi : [k] \longrightarrow [k]$. The following conditions will hold

1. $1 \leq i \leq kd_{Tv}(\mathcal{N}(\mu_i, \Sigma_i), \mathcal{N}(\hat{\mu}\pi(i), (\hat{\Sigma}\pi(i)) \leq O(\alpha)$

2. $\forall 1 \leq i \leq k, |w_i - \hat{w}\pi(i)| \leq O(\alpha/k)$

Both conditions imply that $d_T v(D, \hat{D}) \leq \alpha$

Definition 2 (Learning Labelled Clusters): We learn the mixture of Gaussian, where we can choose G_i from a mixture component $G_{\rho i}$. Such that ρ_i is a label to predict the mixture component in the i.i.d. clusters. A labeled cluster is a set of tuples $G_\rho = ((G_1, \rho_1), \ldots, (G_m, \rho_m))$ sampled from a distribution D, where

$$D \in G(d, k, \sigma_{min}, \sigma_{max}, R, w_{min}, s)$$

The label ρ is composed of a matrix $\rho = \rho(i, j)$ which is the same size as D. Each element $\rho(i, j)$ is a label of corresponding pixels in the original data X. Let p_t denote the label of sensitive clusters in G. The classification result maps of the non-sensitive clusters in the original distribution $\hat{D}$ should be similar.

We aim to locate the clusters distinctly so sensitive clusters are perturbed, and non-sensitive clusters are unperturbed. So, we divide the image into sensitive and non-sensitive parts using masking, where M_t and M_o denote the parts, respectively. M_t is $0 - 1$ binary matrix which equals $M_t(i, j)$, where $M_t(i, j) = 1$ *iff* $\rho(i, j) = \rho_t$ and $M_o = 1 - M_t$ where 1 is an all 1 matrix with the same size as M_t. Our GMM algorithm locates the object clusters by their binary number label.

2) VARIATIONAL AUTO-ENCODER

In the inference process of the VAE, the encoded latent variable z is obtained from sampling the output of the Gaussian mixture $z \sim \mathcal{N}(\mu_j, \sigma_j^2)$. The reparameterisation trick is used to adapt the recognition model $q(z|G_i)$ to approximate the time posterior distribution $p_\theta(z|G_i)$. So, make z be a deterministic function of ϕ and some noise ϵ, where $z = f(\phi, \epsilon)$. A sample can be drawn from a normal distribution like $z = \mu + \sigma\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$.

In the generative process, the obtained latent variable z is decoded to obtain another distribution $p_\theta(z)$, where the synthetic image $\hat{x}$ can be sampled. The DNN parameters ϕ and θ are jointly learned by optimising the ELBO using the Stochastic gradient descent of the DNN. The ELBO is computed as the difference between the latent variable distribution and the observed variable distribution as follows;

$$\log p(x) \geq \mathcal{L}(x) = Eq_\phi(z|x)[\log p_\theta(x|z)] - KL(q_\phi(z|x)\|p_\theta(z))$$

where the first term of the difference is the expected log-likelihood, and the second term is the KL divergence.

To improve the visual quality of the non-private areas, we inject information about the non-private clusters into the generative process of the decoder. The conditional information ρ' has the same size as ρ and only holds information about the non-private objects. Hence, the conditional VAE reconstruct most labelled non-private areas to preserve utility. The loss function for the conditional VAE based on the generative model is stated as

$$\mathcal{L}_c(x) = Eq_\phi(z|x)[logp_\theta(x|z, \rho')] - KL(q_\phi(z|x)\|p_\theta(z))$$

3) DIFFERENTIAL PRIVACY

A randomised mechanism M will satisfy (ϵ, δ)-differential privacy $((\epsilon, \delta) - DP)$ for learning mixtures of Gaussian if it takes two pair of image data $(X, \hat{X})$ that differ in one single item (pixel), the distributions $M(X)$ and $M(\hat{X})$ are precisely (ϵ, δ)-close. If the image data is partitioned into cluster distributions $X_1, \ldots, X_k \sim$ D for a mixture D satisfying separation and boundedness, $M(X)$ produces an approximate output to the parameter of G. The images $X, \hat{X} \in M$ and every set of output O, if M satisfies

$$Pr[M(X) \in O] \leq e^\epsilon . Pr[M(\hat{X}) \in O] + \delta$$

where $Pr[.]$ denotes the probability of an event, and δ bounds the probability of the privacy guarantee not holding, which is often better set to be less than $1/|D|$. Specifically, the distribution of $A(D)$ and A$(\hat{D})$ are (ϵ, δ)-close.

Let's define the global Lp-sensitivity of the feature vector $f(x)$, as we inject noise into the cluster centres of specific locations in the image. If the images consist of n pixels, such that $X = (x_1, \ldots, x_n)$ and $\hat{X} = (\hat{x}_1, \ldots, \hat{x}_n)$, the function f maps the image to feature space, and the sensitivity Δf is defined as

$$\Delta pf = \max_{X,\hat{X}} ||f(X) - f(\hat{X})||p$$

where $X, \hat{X}$ are neighbouring datasets, Δ_f is the maximum differences in $f(x)$ generated by two different images, and $\|.\|p$ denotes the $Lp - norm$.

Our private GMM achieves differential privacy by injecting Gaussian noise, defined in the following.

Gaussian Mechanism (GM): The GM with parameter σ adds noise scaled to $\mathcal{N}(0, \sigma^2)$ to each of the private components of the output. For any $G(X) = f(X) + [N1(0, \Delta 2f.\alpha)],\ldots, Nd(0, \Delta 2f.\sigma]$ where $Ni(0, \Delta 2f.\sigma)$ are i.i.d. normal random variables with zero mean and variance $(\Delta 2f.\sigma)^2$. Let $\epsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the Gaussian mechanism with parameter $\sigma \geq c\Delta 2f/\epsilon$ is $(\epsilon, \delta) - DP$.

To learn our differentially private GMM with well-separated and bounded image object clusters, we describe the private GMM conditions in the following theorem (the proof is available in [31]).

Theorem 1: A (ϵ, δ)-differentially private algorithm takes n samples from an unknown mixture of k Gaussians $D \in R^d$ satisfying the above conditions of separation and boundedness.

$$n = \left(\frac{d^2}{\alpha^2 w_{min}} + \frac{d^2}{\alpha w_{min}\epsilon} + \frac{poly(k)d^{3/2}}{w_{min}\epsilon}\right) \cdot poly\log\left(\frac{dkR(\sigma_{max}/\sigma_{min}}{\alpha\beta\epsilon\delta}\right)$$

where $W_{min} = min_i w_i$, with probability at least $1 - \beta$, learning the parameters of D up to error α. The parameters $\alpha, \beta, \epsilon, \delta$ are the estimator accuracy of variation distance, failure probability, and privacy parameters, respectively. R is the radius of a ball at the centre containing all means, and k is the ratio of the variances' upper and lower bound.

Under Theorem 1, we transform data to a lower dimension space and recursively cluster the data with a Principal Component Analysis (PCA) [32]. This approach ensures the maximum effect of the injected noise. The PCA projection privately learns under the following assumptions: (i) All components being spherically Gaussian such that each component's variances lie in a small known range (with bounder ratio by a constant factor), (ii) The means of the Gaussian lie in a small ball around the origin. Making the PCA private by injecting noise into the covariance matrix makes the algorithm private. The projection shifts each component mean by the complexity of $\mathcal{O}(\sqrt{k\sigma_{min}})$ under the already stated assumptions and preserves the separation of data because all variances are within a constant factor of one another. Finally, cluster data using the 1-cluster method of [33] and learn each component's parameters using a simplified version of [34].

IV. EXPERIMENTAL ANALYSIS

To evaluate the performance of DeepClean, we use a dataset which is a high-dimensional street view scene from Cityscapes [35]. The image data consists of 2975 training sets, 500 validation, and 1525 test sets showing street views of different cities at different times. The images have a size of $256 * 256$ and are trained with no data augmentation because the DNN learnt more patterns and trained faster without it. We set up our deep-learning Python and Tensorflow on a Colab playbook.

Training Method – In all the experiments, we follow the same experimental setup of the VAE network in ADGAN-II [11] by set epochs to 150 and batch-size of 1. For DeepClean, the latent dim is 128, the label dim of 64, beta $\beta = 0.65$, and the learning rate of 0.001.

For our comparative analysis, we evaluate the performance of DeepClean in comparison with two benchmark techniques for AV camera data, i.e., ADGAN [11] and VAE+DP-Kmeans [25]. We chose these two techniques due to their balanced privacy/utility claims and their use of VAE models (similar to DeepClean). Regarding the chosen dataset, ADGAN and VAE+DP-kmeans models were evaluated using the Cityscapes dataset.

The comparison results (provided in this section) show that DeepClean outperforms the considered benchmark techniques by preserving the better visual quality of the

TABLE 1. FCN-score comparisons of various generative models on the cityscape dataset.

Model	Image Quality (IQ)		Image Privacy (IP)		Image Utility (IU)	
	PA	IoU	PA	IoU	PA	IoU
ADGAN [11]	70.69%	17.39%	11.65%	4.72%	77.53%	21.06%
VAE+DP-Kmeans [25]	64.60%	15.86%	6.27%	2.37%	60.54%	16.53%
DeepClean	68.30%	17.15%	6.35%	2.76%	77.58%	23.04%

non-private object parts of an image while resisting location inference attacks. A brief explanation of these three techniques is provided in the following to improve readability.

- ADGAN [11] – combines VAE and GAN. The synthetic image is generated by the generator transformation $\hat{x} = G(x)$ and applies a privacy loss function $L_{pri}(G)$ to make $\hat{x}$ privacy-preserving.
- VAE + DP-kmeans [25] – combines VAE and private Kmeans. The synthetic image is generated by adding differentially private Kmeans on the data points $D = x_1, \ldots, x_N$, the results of the cluster data is produced by a DPKmeans (Parameters) $= D_1, D_2, \ldots, D_k$. The output of the parameters is used to learn the VAE generator.
- DeepClean (VAE + DP-GMM) – is our proposed method to combine VAE and a private GMM. The GMM is applied to the latent distribution to learn sensitive and non-sensitive objects in clusters. Gaussian noise is applied to the sensitive clusters, while the noise does not impact non-sensitive objects. The clusters are then trained in a conditional VAE system.

A. EVALUATING IMAGE QUALITY, PRIVACY AND OVERALL UTILITY

To measure the efficiency of the techniques, we adopt the FCN score to quantify the features in the generated synthetic images. FCN score is efficiently adopted to evaluate generative models quantitatively [36]. Two indicators from the FCN score are used for the evaluation: pixel accuracy (PA) and interaction over union (IoU). The PA value estimates how well the image pixels are represented in percentages. In contrast, the IoU value estimates the overlap between the predicted segmentation and the ground truth over the area of the union between the predicted segmentation and the ground truth. We run the semantic segmentation model to compute the PA and IoU values of the generated images.

The evaluation of the indicators is defined as comparing performances using three metrics, i.e., image quality (IQ), image utility (IU), and image privacy (IP). IQ is estimated by taking the average PA and IoU over the whole image, IU is calculated by averaging PA and IoU over non-private objects in the image, and IP is estimated by averaging PA and IoU over the private objects in the image data. As for the metrics IQ and IU, the higher their value, the better the image representation performance of the technique. While for IP, the lower the value, the more privacy is preserved and the more difficult it is to recognise an object from the image.

We initially show DeepClean produces better IQ and IP than the other techniques. Table 1 shows the FCN-scores comparison of DeepClean with the other techniques using the Cityscapes dataset. DeepClean achieves a global IQ accuracy of 68.30% PA and 17.15% IoU, slightly as good as ADGAN, 70.69% PA and 17.39% IoU, and VAE+DP-kmeans with 64.60% PA and 15.86% IoU. The drop in performance of DeepClean compared to ADGAN is due to achieving better privacy preservation in the private areas of the images. However, the overall IQ performance can be improved by reducing the number of noisy scales on the IP. DeepClean preserves more privacy by achieving a lower IP value, 6.36% PA and 2.76% IoU, compared to the other models. By this, DeepClean shows better resistance to privacy attacks. The goal to preserve more utility around the non-private object areas is achieved, with IU measurement of 77.75% PA and 21.20% IoU for DeepClean, which is better than the other models. The good performance of DeepClean is due to the good clustering proficiency of GMM on the distributions. However, the two deep clustering models show the effectiveness of good clustering in better controlling the image quality of specific locations in the images.

Figure 4 shows the accuracy of the clustering technique over some epochs in training the Cityscapes dataset. The number of clusters k was initially set to 10 to achieve high clustering performance. Setting the privacy parameters for the benchmark techniques, we use the default settings in the K-means model [25], and for the clustering models, noise scales for clustering σ_k is set as 1.0 and SGD noise scale σ_G as 40. The privacy metric result shows that the DeepClean model achieves reasonable privacy protection better than ADGAN concerning the utility gained in the non-private object areas.

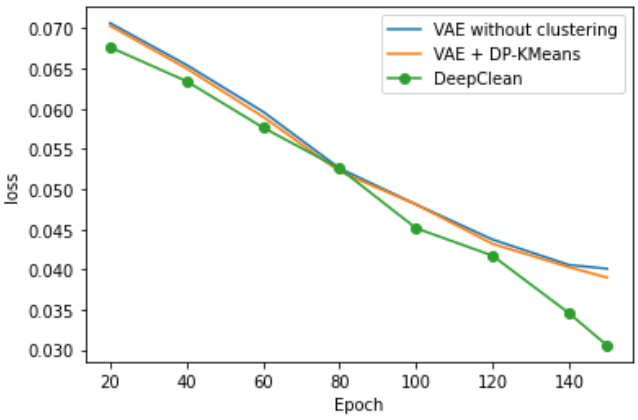

FIGURE 4. Clustering accuracy over some epochs during training on the Cityscapes dataset.

B. PRIVACY PERFORMANCE

To validate the performance of privacy protection achieved by our proposed technique, we run the geo-localisation attack using dominant set clusters (DSC) to localise the query image data. The reference dataset used for the experiment is 102k google street view images covering different cities in Europe. We select 500 sets from Section IV Cityscape test set for the query image set. The DSC quantifies the percentages of images that can be localised at 300m from their actual locations. Localisation above the 300m range is regarded as non-matching nearest neighbours. Using the DSC and the constrained DSC post-processing step for feature matching and geolocating the best matching reference image, respectively, improves the performance of geo-localisation than the Multi-KNN approach used in other studies for privacy performance.

Figure 5 shows the privacy performance of DeepClean on the images compared with the benchmark studies. The X-axis is the error threshold in meters, and Y-axis is the percentage of the test set localised within the error threshold. DSC localises the original query images at 74%, about 300m better than Multi-KNN 60%. The higher percentage result proves a higher risk of location inference threats on the image data. On the other hand, using distorted images of ADGAN models as the query image, localisation improves from 5% to 20% within the error threshold of 60m – 300m. This improvement indicates that DSC can still match some features to the produced dynamic corresponding reference data set. DeepClean reduces localisation accuracy to about 3% - 7%, which is relatively minimal compared to the other techniques. With this result, there are possibilities that the original reference images are not included among the matching nearest neighbour images. Both local and global features present around the classified private object areas are well distorted to confuse the DSC from detecting stable features. Only a few images with more stable features around features such as road signs, vegetation and structures, apart from buildings, likely make the matching step. However, the image is unlikely to return as the best matching image. This result makes Deepclean images immune to location inference attacks.

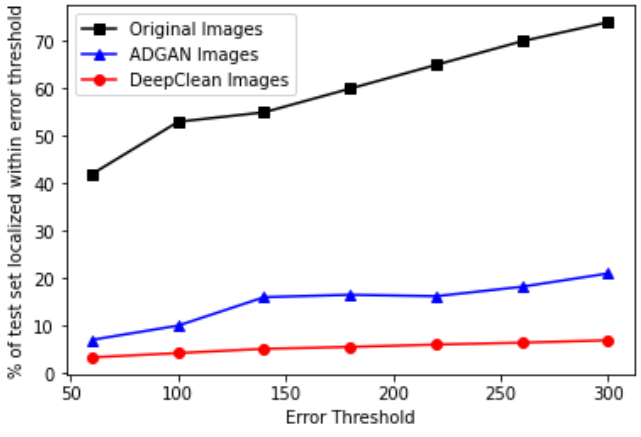

FIGURE 5. Privacy performance of DeepClean compared with ADGAN.

As seen in Figure 6, we tested a fixed Multi-KNN to examine the performance of the DSC on different numbers of nearest neighbours. Although Multi-KNN used in previous works drops in performance when k is $\geq$ 4, DSC improves the chances of selecting the original image data as the nearest neighbour increases. The first 4 NNs retrieved by the multi-KNN method assume the NNs are the stable features detected from the image. These detected features show that they contribute more to the localisation accuracy.

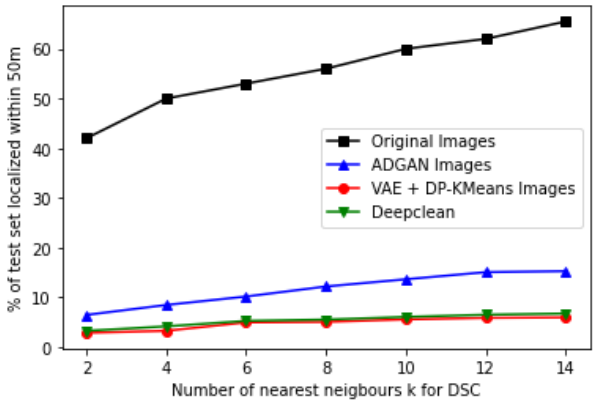

FIGURE 6. Performance comparisons of the techniques with fixed nearest neighbour.

C. UTILITY PERFORMANCE

To evaluate the utility performance of the DeepClean model, we measure the structural similarity index (SSIM) of the generated images. SSIM measures image recognition utility very close to human visibility [37]. It measures the similarity between the original and distorted data by a number greater or equal to 0 and less or equal to 1, where 0 means completely different, and 1 means the same. Table 2 shows that DeepClean achieves 0.6012 on the Cityscape data, which is closer to the value achieved by ADGAN. The slight drop in utility performance of DeepClean compared to ADGAN considers the stricter privacy requirements enforced in the private object areas. This performance only highlights the challenge of simultaneously achieving a balanced privacy-utility trade-off in images. Thus, the privacy-utility performance results show that a balanced trade-off may not be achievable to suit all requirements. Therefore, it explains our approach to achieving more utility in the non-private object areas. The results produced by DeepClean, as shown

TABLE 2. SSIM measurement on Cityscapes dataset.

Model	SSIM Measurement
ADGAN	0.6210
VAE+DP-Kmeans	0.4560
DeepClean	0.6012

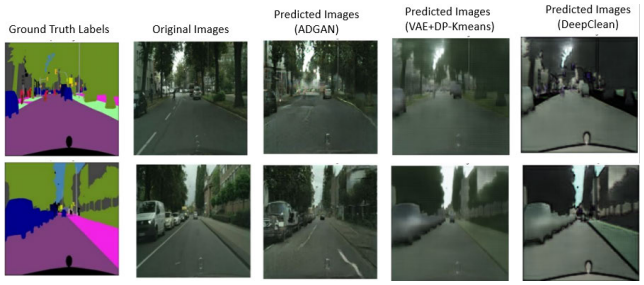

FIGURE 7. Visual quality of non-sensitive object areas and privacy-preserved sensitive areas using three techniques on the Cityscapes data.

in Figure 7 (among other image data generated by the other techniques), generate a more balanced privacy-utility trade-off regarding more privacy preservation in private object areas and utility preserved in non-private object areas. DeepClean generated data can be used to train AV driving navigation models.

V. CONCLUSION

Location inference attacks threaten the privacy of Autonomous vehicle camera data. For this reason, a reasonable level of security and privacy is required to enhance data storage and sensitive image protections, respectively. Focusing on the privacy-preservation of AV camera data, this study has addressed the privacy/utility trade-off for efficient data analysis and storage. Our proposed generative model approach integrates a differentially private technique to guarantee privacy instead of relying on masking or reconstruction loss for privacy protection by prior works. The comparative analysis of the models showed that DeepClean achieves better privacy preservation and comparable utility performance to benchmark models. Future research on AV camera data privacy preservation could formulate a GAN-based model amenable to differential privacy. Aiming to utilise generative and discriminative models for an improved image utility with a provable privacy guarantee.

REFERENCES

[1] *U.K. Connected & Autonomous Vehicle Research & Development Projects 2018*, CCAV, Beijing, China, 2017, p. 64.
[2] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.
[3] X. Li, L. Ding, L. Wang, and F. Cao, "FPGA accelerates deep residual learning for image recognition," in *Proc. IEEE 2nd Inf. Technol., Netw., Electron. Automat. Control Conf. (ITNEC)*, Dec. 2017, pp. 837–840.
[4] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, Feb. 2016.
[5] W. Zhang, C. Witharana, W. Li, C. Zhang, X. Li, and J. Parent, "Using deep learning to identify utility poles with crossarms and estimate their locations from Google street view images," *Sensors*, vol. 18, no. 8, p. 2484, Aug. 2018.
[6] A. Reich, N. A. Krämer, and R. Lenninger, "Vehicle data mangement a standardized access as the basis of new business models," *ATZelektronik worldwide*, vol. 13, no. 2, pp. 38–43, Apr. 2018.
[7] B. Martens and F. Mueller-Langer, "Access to digital car data and competition in aftersales services," pp. 1–24, Oct. 2018. [Online]. Available: https://ssrn.com/abstract=3262807, doi: 10.2139/ssrn.3262807.
[8] V. K. Veitas and S. Delaere, "In-vehicle data recording, storage and access management in autonomous vehicles," 2018, *arXiv:1806.03243*.
[9] V. Dhar, "Equity, safety, and privacy in the autonomous vehicle era," *Computer*, vol. 49, no. 11, pp. 80–83, Nov. 2016.
[10] J. Hays and A. A. Efros, *Large-Scale Image Geolocalization*. Cham, Switzerland: Springer, Jan. 2015.
[11] Z. Xiong, Z. Cai, Q. Han, A. Alrawais, and W. Li, "ADGAN: Protect your location privacy in camera data of auto-driving vehicles," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6200–6210, Sep. 2021.
[12] A. R. Zamir and M. Shah, "Image geo-localization based on MultipleNearest neighbor feature matching UsingGeneralized graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1546–1558, Aug. 2014.
[13] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–7.
[14] Z. Xiong, W. Li, Q. Han, and Z. Cai, "Privacy-preserving auto-driving: A GAN-based approach to protect vehicular camera data," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 668–677.
[15] E. Zemene, Y. T. Tesfaye, H. Idrees, A. Prati, M. Pelillo, and M. Shah, "Large-scale image geo-localization using dominant sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 148–161, Jan. 2019.
[16] T. Sattler, M. Havlena, K. Schindler, and M. Pollefeys, "Large-scale location recognition and the geometric burstiness problem," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2016, pp. 1582–1590.
[17] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1437–1451, Jun. 2018.
[18] Z. Xiong, W. Li, Q. Han, and Z. Cai, "Privacy-preserving auto-driving: A GAN-based approach to protect vehicular camera data," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 668–677.
[19] M. Amiri-Zarandi, R. A. Dara, and E. Fraser, "A survey of machine learning-based solutions to protect privacy in the Internet of Things," *Comput. Secur.*, vol. 96, Sep. 2020, Art. no. 101921.
[20] Y. A. A. S. Aldeen, M. Salleh, and M. A. Razzaque, "A comprehensive review on privacy preserving data mining," *SpringerPlus*, vol. 4, no. 1, pp. 1–36, Dec. 2015.
[21] Z. Wang, S. Chang, J. Zhou, M. Wang, and T. S. Huang, "Learning a task-specific deep architecture for clustering," in *Proc. SIAM Int. Conf. Data Mining*, Jun. 2016, pp. 369–377.
[22] J. Yang, D. Parikh, and D. Batra, "Joint unsupervised learning of deep representations and image clusters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5147–5156.
[23] C.-C. Hsu and C.-W. Lin, "CNN-based joint clustering and representation learning with feature drift compensation for large-scale image data," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 421–429, Feb. 2018.
[24] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," 2017, *arXiv:1703.00848*, doi: 10.48550/arXiv.1703.00848.
[25] G. Acs, L. Melis, C. Castelluccia, and E. D. Cristofaro, "Differentially private mixture of generative neural networks," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1109–1121, Jun. 2019.
[26] Z. Xiong, W. Li, Q. Han, and Z. Cai, "Privacy-preserving auto-driving: A GAN-based approach to protect vehicular camera data," pp. 668–677, 2019.
[27] K. Lata, M. Dave, and K. N. Nishanth, "Image-to-image translation using generative adversarial network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 668–677.
[28] D. Cremers, I. Reid, H. Saito, and M.-H. Yang, Eds., *Computer Vision ACCV 2014*, vol. 9006. Cham, Switzerland: Springer, 2015.
[29] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational deep embedding: An unsupervised and generative approach to clustering," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1965–1972.
[30] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Context-aware generative adversarial privacy," *Entropy*, vol. 19, no. 12, p. 656, Dec. 2017.
[31] G. Kamath, O. Sheffet, V. Singhal, and J. Ullman, "Differentially private algorithms for learning mixtures of separated Gaussians," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2020, pp. 1–13.
[32] S. Vempala and G. Wang, "A spectral algorithm for learning mixtures of distributions," in *Proc. 43rd Annu. IEEE Symp. Found. Comput. Sci.*, Nov. 2002, pp. 113–122.

[33] K. Nissim, U. Stemmer, and S. Vadhan, "Locating a small cluster privately," in *Proc. 35th ACM SIGMOD-SIGACT-SIGAI Symp. Princ. Database Syst.*, Jun. 2016, pp. 413–427.
[34] G. Kamath, J. Li, V. Singhal, and J. Ullman, "Privately learning high-dimensional distributions," 2019.
[35] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.
[36] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

OLAYINKA ADEBOYE received the M.Sc. degree in cybersecurity and threat intelligence with merits and a dissertation on the critical security and privacy concerns on social networks.

He is currently a Research Student with the University of Salford, with a focus on research interests in cybersecurity, data privacy, data analytics, and artificial intelligence. His Ph.D. research focuses on the privacy-preservation of connected and autonomous vehicle data, intending to analyze the privacy risks associated with sensitive data and highlight potential threats to the system. As a solution to efficient camera data privacy-preservation for autonomous vehicle data storage and processing, he developed a private generative technique to address a balanced privacy-utility tradeoff. He delivered talks at the Salford postgraduate research conferences about his work and was awarded the best presentation. He also engages as a teaching assistant to support his supervisors in delivering workshops and hands-on experiments for level five and seven students.

TOOSKA DARGAHI received the Ph.D. degree in computer engineering from Azad University, Tehran Science and Research Branch, Iran, in 2014.

She is a Senior Lecturer in Cyber Security at Manchester Metropolitan University (Manchester Met), U.K. She was a Visiting Researcher at the University of Padua, Italy, in 2014, and a Research Fellow at the University of Rome Tor Vergata, Italy, between 2015 and 2017. Prior to joining Manchester Met, she was a Lecturer in Cyber Security at the University of Salford (U.K.), between 2017 and 2022. She has an established track record in the field of security and privacy in distributed systems, including the Internet of Things (IoT) and smart cities. She contributes to public engagement activities to increase awareness about the security and privacy challenges of modern digital society. She has also served as a guest editor for several special issues in reputable journals and an Associate Editor of *Cyber Threat Intelligence* (Springer Book). She has also organized several (IEEE/ACM) international conferences and workshops and served as a reviewer for high-ranked journals and conferences.

MEISAM BABAIE is currently a Lecturer in the School of Mechanical Engineering at the University of Leeds, U.K. He completed his Ph.D. in Mechanical Engineering at the Queensland University of Technology (QUT), Australia, in 2014. He is an experienced mechanical engineer and scientist, with employment history in both industry and academia. He worked in different industries such as automotive manufacturing, disposable products manufacturing and cement industry and in international academic institutions and laboratories such as Biofuel Engine Lab Facility in Australia, Gunma University in Japan, and Automotive and Autonomous Vehicle Technology Laboratory in the U.K. His research interest encompasses different areas of modern automotive engineering including connected and autonomous vehicles. He has an established track record in the field, and his research outcome has been presented in reputable conferences, industrial reports, invited talks, and prestigious journals of the field.

As a Mechanical Engineer, my disciplines of expertise are in Automotive and Energy domains with a special interest on energy analytics and low emission and smart vehicles. Due to the multidisciplinary nature of my research, I also have worked with several universities and industries in other disciplines such as petroleum and gas engineering, aerospace engineering, environmental science, power engineering, robotics and computer science. My research has led to many publications in different prestigious journal of the field (e.g., *International Journal of Hydrogen Energy*, *Energy Conversion and Management*, *Renewable and Sustainable Energy Review*, *Chemical Engineering Journal*, *Energy*, *International Journal of Environmental Science and Technology*, *PloS one*, IEEE Transactions on Plasma Science, etc.) and several conference presentations, industrial reports, and invited talks.

Since the establishment of Automotive and Autonomous Vehicle Technology (AAVT) Laboratory at the University of Salford, I was the lead academic developing the collaborative research in Modern Automotive Engineering in different areas for conventional, EV/Hybrid, and Connected and Autonomous Vehicles (CAVs).

MOHAMAD (MO) SARAEE holds a chair in Data Science, and he is the program leader of the MSc Data Science and MSc IoT with Data Science courses at the University of Salford-Manchester. He received his Ph.D. in Computer Science from the University of Manchester. He leads a Data Science research group that focuses on Machine and Deep Learning, Data and Text Mining, NLP, Big Data, and Medical Informatics. His research addressed multi-disciplinary, cross-school topics with transformative impact, benefiting local communities, dedication to action through real-world application of research including developing and integrating innovative data mining approaches to improve human health, in collaboration with both Salford City Council and the NHS. He has secured and led funded projects with income totalling over £1.7m. He is the Editor in Chief for *International Journal of Web Research* and a reviewer for several high-tier international journals.

CHIA-MU YU (Senior Member, IEEE) is an Associate Professor at the National Yang Ming Chiao Tung University, Taiwan. He had academic visits to IBM Thomas J. Watson Research Center, Harvard University, Imperial College London, University of Padova, and the University of Illinois at Chicago. He received Hwa Tse Roger Liang Junior Chair Professor, MOST Yong Scholar Fellowship, ACM/IICM K. T. Li Young Researcher Award, Observational Research Scholarship from Pan Wen Yuan Foundation, and MOST Project for Excellent Junior Research Investigators, Taiwan. He is an Associate Editor of IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and IEEE INTERNET OF THINGS JOURNAL, and IEEE *Consumer Electronics Magazine*. His research interests include differentially private mechanism design, cloud storage security, and IoT security.