Article

GAN-based Differential Private Image Privacy Protection Framework for Internet of Multimedia Things

Version November 9, 2020 submitted to Journal Not Specified

Abstract: With the development of Internet of Multimedia Things (IoMT), more and more image data is collected by various multimedia devices, such as smart phones, cameras, drones. These massive amount of images are widely used in each field of IoMT, which presents substantial challenges for privacy preservation. In this paper,we propose a new image privacy protection framework, with an effort to protect the sensitive personal information contained in images collected by IoMT devices. We aim to use deep neural network techniques to identify the privacy-sensitive content in images, and then protect it with synthetic content generated by generative adversarial networks (GANs) with differential privacy (DP). Our experimental results show that the proposed framework can effectively protect users' privacy while maintaining image utility.

Keywords: Internet of Multimedia Things (IoMT), image privacy, object detection, deep learning, generative adversarial network.differential privacy

1. Introduction

The recent advances in multimedia-recording devices, such as phones, cameras, drones, and other type of sensors, have greatly facilitated the collection of multimedia data, especially in the form of images and videos. In such an era of IoMT, a massive amount of images are widely used, not only by social network personal users but also by government and companies. Image data is the most representative type of data in IoMT data collection, which contain sensitive information that might be used to dig personal information. Data mining attacks on images can easily cause privacy leakage, which can cause serious consequences. The issue of privacy leakage has been paid attention by the public in recent years, which has aroused public concern about this issue. Moreover, privacy issues are no longer just personal concerns as many countries have launched privacy acts and laws. For example, the European General Data Protection Regulation (GDPR) took effect on 25 May 2018 [1]. Any violations of the regulation will trigger heavy fines and penalties. GDPR emphasizes the protection of "personal data", interpreting as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" [2]. According to this definition, images include a variety of personal identifiers such as people's faces, text and license plates. Therefore, effective image privacy protection techniques are in urgent need.

The research community has seen some effort in image privacy protection. The early works mostly focus on the access control of the data, i.e., privacy protection by safeguarding against unauthorized access. This can be achieved through setting preferences of users [3] [4] or tags control [5] [6]. However, these methods cannot be applied to the scenarios where the images are shared openly, but some sensitive information needs to be concealed. For example, in the "Google Street View" application, we have full access to photos showing the streets while people's faces and other personal identifiers have been obfuscated, e.g. by blurring. To achieve this, the privacy protection methods need to detect, and then cover/remove/replace sensitive content in images. There are some recent research in this direction [7] [8] [9] [10] [11] [12] [13] [14]. For example, Viola et al. [7] used a sliding window detector to identify and blur the license plates in Google Street View images. Yu et al. [9] used a deep multi-task learning algorithm to detect privacy-sensitive objects and provide simple protection by blurring. Overall, most of the existing work performs personal data detection as the first step of privacy protection. While on the protection part, it mostly relies on simple approaches such as blurring or pixelation. Consequently, the image utility suffers to a considerable extent. It not only makes the images look unnatural, but also makes the person who looks at the image aware that the obfuscated part is private. Moreover, such a protection mechanism is powerless in facing the emerging attacks based on advanced deep neural networks. For example, Mcpherson et al. [15] use artificial neural networks to recover hidden information from images protected by pixelation, blurring and P3.And the method obtained good results on different data sets, MINIST 80%, CIFAR-10 75%,ATT dataset 95%, FaceScurb 57%.

Moreover, the existing methods are almost discussing single object protection,such as face or text. However, most images that require privacy protection have multiple objects that need to be protected (For example, in street view images, human faces and license plates need to be protected at the same time).

Current methods are unable to find a way to quantify the tradeoff between image usability and privacy protection. To tackle this, we use DP to control the image private objects generation to mitigate privacy threats.

To overcome these obstacles, we propose the a three-stage frameworks for image privacy protection in this paper. The framework consists of three steps: 1) privacy-sensitive content detection and position extraction powered by a deep Convolutional Neural Network: We use CNN networks to detect various objects in images and classifying objects into private and non-private ones; 2) real private objects projecting into latent space: We use generative adversarial networks(GANs) to projecting the real private objects of the images into latent space and get the corresponding latent vector ω. 3) private content generation controlled by DP (de-identification): We use Laplace noise into the latent vector ω and generated de-identification content. Finally, replace the originally private objects with the synthetic ones to protect users' privacy.

In order to evaluate the performance of our proposed framework, we have conducted extensive experiments on a real-world image data set collected by cameras of IoMT, and investigated two types of personal identifier related data: license plate and face. We choose these two types of objects as they represent the two most significant categories of personal identifiers in images.

In summary, the contributions of this paper are as follows:

- We propose an image privacy protection framework that can protect the privacy in the IoMT's image.
- We propose a GAN-based method to generate the replacement content for private objects in the images.
- We use differential privacy methods to disturb generation to quantify the tradeoff between image usability and privacy protection.

The remaining of the paper is organized as follows. Section 2 reviews the related work. Section 3 give the definition and foundation of the methods. Section 4 presents our framework on multimedia privacy protection based on Mask-RCNN and synthetic content generation using GANs. Section 5 shows the experimental results of our framework for multi-object privacy protection(street view scenarios). Section 6 concludes the paper and outlines the future work.

2. RELATED WORK

Privacy protection, in general, has been extensively studied in recent years. Among all the researches, differential privacy (DP) has attracted the most attentions and applied to different application. Therefore, in this section, we will review the most relevant research works on image privacy and the related fundamental deep learning researches, including: (1) image privacy issue and protection; (2) deep learning and object detection of the images; (3) the content generation; (4)and privacy protection.

2.1. Image Privacy Issue and Protection

The image privacy issue first attracted people's attention along with the booming of social networks developing. The proliferation of social networks generated massive photos flooding on the internet that contains sensitive information. For example, Pesce et al. [16] use photo tags to attack users and get their privacy. The image privacy issue becomes more server with the widely spread of facial recognition systems, as people start to worry that their faces might be used by organizations for profiling or social control.

To combat the image privacy attack, the previous mainstream method is using access control on sensitive contents. Mannan et al. [3] use Instant Messaging (IM) networks to control personal web content sharing. Vyas et al. [4] use annotation data to predict the privacy preferences of users and control the shared content. Wang et al. [5] studied privacy control on Facebook. Moreover, Squicciarini et al. [6] proposed collaborative privacy management that can let users collaborative control their photos. Similarly, to deal with the privacy issue in facial recognition systems, the current countermeasure is simply banning [17]. The access control-based method has several limitations. It only gives "Yes" or "No" options for the use of images, while we need to use part of the information in applications such as Google Street View. And it can not automate protect privacy based on the privacy information of the image itself, requiring human participation.

Some more recent image privacy researches focus on the inherently implicit information of the photos. Tonge et al. [8] explore learning models that can automatically classify the private or public parts in an image by using Deep Neural Networks. Yu et al. [9] create a new tool called "iPrivacy" that uses a deep learning algorithm to detect the privacy-sensitive objects. Yu's work can detect the privacy parts of photos, but in the step of privacy protection, they just use blur to protect privacy which is not good looking. More than blurring, Uittenbogaard's work [10] set a framework that automatically removes moving objects. However, there are two limitations, one only for moving objects and the other for missing partial information in the image. Liu's work [11] proposes a novel Stealth algorithm, which makes the automatic detector can not detect the objects in an image. However, human beings can easily get privacy information from the image.

Our framework is a further advancement compared with the researches mentioned above. It can identify the privacy part of the photos in the pixel level. Then it will generate the target replacement content based on the privacy content, not just using mosaic, blurring or removing to protect privacy. Our framework can protect privacy information from both human and machine.

2.2. Deep Learning-based Image Object Detection and Segmentation

Object detection and semantic segmentation technologies have been advancing rapidly in recent years. In the beginning, Girshick et al. [18] use high-capacity convolutional neural networks (CNNs) to bottom-up region proposals, which called R-CNN. This algorithm improves the mean average precision (mAP). In 2015, Hariharan et al. [19] define the hypercolumn at a pixel as the vector of activations of all CNN units above that pixel to improve the result of the experiment. After that, a large part of the research works are based on the Fast R-CNN [20] [21] and Fully Convolution Network (FCN) [22]. The disadvantage of Faster R-CNN is that it cannot deal with pixel-to-pixel alignment between the inputs and outputs of the network. To solve this problem, He et al. proposed a method called Mask R-CNN [23] that extends the Fast R-CNN by adding predicting segmentation masks on each Region of Interest (RoI) to get the results. As our goal is to find the privacy part of the images, so we choose to use the Mask R-CNN to get the instance segmentation results that can be used as the basis for the follow-up privacy content detection and positioning. To obtain good results for our use case, we need to re-train the network using our image dataset that includes more privacy sensitive contents.

2.3. GAN-based Content Generation

Preliminary ways for image privacy content protection include blurring, deletion, etc. In this paper, we use the replacement of content to protect privacy, i.e. generating content without identification information to replace the privacy content in the images. Traditional content generation methods such as [24] [25] [26] [27] just fill the pixels by matching and pasting based on the low-level features in the images. The effect is not very satisfactory as they often produce the failure contents and the results obtained are also not good. In 2014, Goodfellow proposed a new framework called GAN [28] can synthesize new content by training the models. Following the GAN-based method, the latest GAN-based generation content generation technology can generate very realistic content, such as faces, cats, dogs, even Airbnb rooms [29] [30] [31] [32]. In our framework, we use StyleGAN [33] to generate the replacement content. The StyleGAN can generate content which is not much different from the real image. The image content generated by StyleGAN does not exist in real life and these contents can avoid copyright disputes. With the replacement of the generated content, the privacy of the images can be protected.

2.4. Privacy Protection

In the traditional privacy protection technology, one of the most common method is data encryption, which has high security. However directly encrypted and decrypted on large-scale data such as image sets will consume a lot of computing resources. Another privacy protection methods is anonymity privacy protection technology. In 2002, Sweeney et al. proposed k-Anonymity[34] method to protect privacy.Machanavajjhala proposed l-Diversity[35] to address the limitations of k-Anonymity, and Li et al. introduced t-Closeness[36]. However, with the development of attack technology, attackers can use data mining, machine learning, background knowledge attack, and big data analysis to obtain enough useful information of the privacy. To solve this problem, Dwork[37] proposed the concept of differential privacy which has a solid mathematical theoretical foundation. Once differential privacy is proposed, it has attracted attention in the field of privacy protection, and various privacy protection algorithms based on differential privacy have been proposed. In this paper, we propose a new image privacy protection method based on the differential privacy method combined with GANs. Take advantage of the controllability of differential privacy, our method can protect the privacy of IoMT images with high controllability.

3. Preliminaries

3.1. Privacy Protection and Image Utility

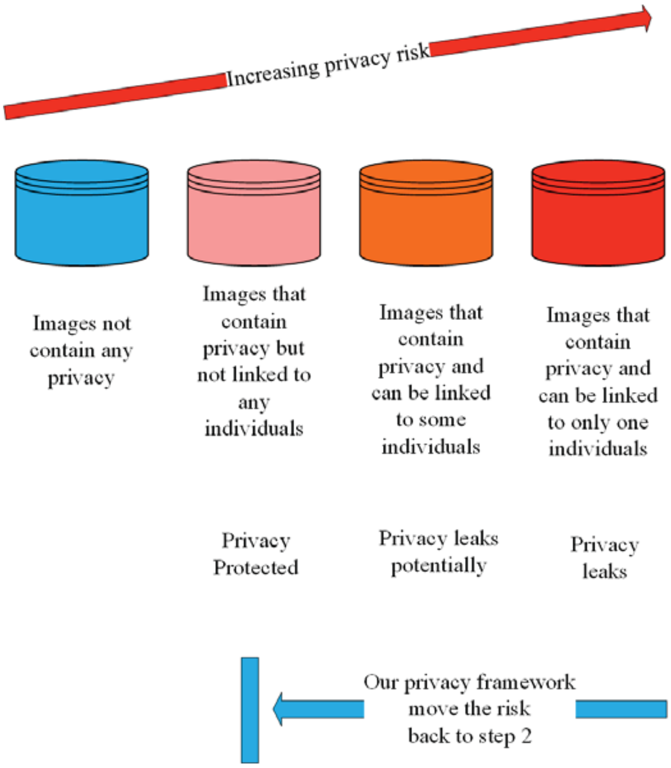

Increasing privacy risk
Images not contain any privacy
Images that contain privacy but not linked to any individuals
Images that contain privacy and can be linked to some individuals
Images that contain privacy and can be linked to only one individuals
Privacy Protected
Privacy leaks potentially
Privacy leaks
Our privacy framework move the risk back to step 2

Figure 1. The four levels of image privacy risks.

In this part, we discuss the image privacy protection and image utility. Firstly, the different levels of image privacy risk are shown in Fig. 1. On the left is images that do not contain any private information (such as a landscape photograph) and the risk of privacy leakage is zero. On the right is images that contain private information and can be linked to specific individuals which violates individuals' privacy directly. Between the two extreme cases are images that contain private information but might not leak individuals privacy. Our goal is to propose a framework to reduce the risk of privacy leak from Level 3/4 back to Level 2 in Fig. 1. It means that we can protect privacy in images so that they cannot be linked to any individual.

However, the strength of privacy protection will affect the utility of images. The common methods such as mosaic and blur, might reduce the utility of the image while image processing. The greater privacy protection, will result in lower utility of images, example shown in Fig. 2. Although mosaic or blur methods protect the privacy, it reduces the readability and usability of the images. It also make images sharing pointless. In our image privacy protection framework, we found an effective way to compromise between privacy protection and image utility.

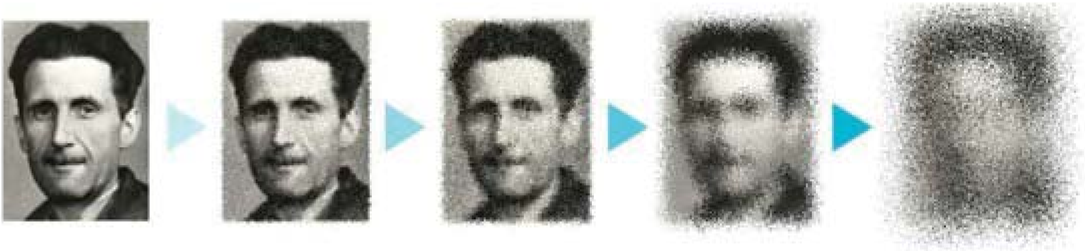


Figure 2. The privacy and utility.

3.2. Formulation of Image De-Identification

We now formally define the problem of image de-identification. This part help us to define the problem we need to deal with and build the foundation for following discussions.

Definition 3.1. (Image). An image is a matrix I of m columns, n rows and c channels. The c channels usually is 3 in common color space such as RGB and YUV. Each cell in matrix I contains a coding which ranging from 0 to 255. Image should contains multi private objects such as face or text.

Definition 3.2. (Object sets). An object set is a set of M objects images contained in image matrix I: $O_i : i = 1, 2, ..., M$.

Definition 3.3. (Privacy object sets). A private object set is a set of N objects images contained in image matrix I:$P_i : i = 1, 2, ..., N$. Which $P_i \in O_i$ and $N \leq M$.

Definition 3.4. (Privacy Object De-Identification Function). Let P and P_d be a private object set and de-identification object set.

$$f : P \rightarrow P_d \tag{1}$$

f is defined de-identification function for each P to remove their identity.

Definition 3.5. (Image De-Identification). Given image matrix I and de-identification function f, for each private object $P_i \in O_i$:

$$I_d = f(I) \tag{2}$$

which we can use de-identification function to get an image matrix I_d not contain privacy.

3.3. Differential Privacy

Definition 3.6. (Differential Privacy). The formal definition of DP is given by (3):

$$Pr[K(D_1) \in S] \leq exp(\epsilon) \times Pr[K(D_2) \in S] \tag{3}$$

Definition 3.7. (The Sensitivity of Differential Privacy). The sensitivity of DP is defined in (4), which determines how much perturbation is required in the DP mechanism.

$$\Delta f = \max_{D_1, D_2} ||f(D_1) - f(D_2)||_1 \tag{4}$$

4. Image De-identification Framework

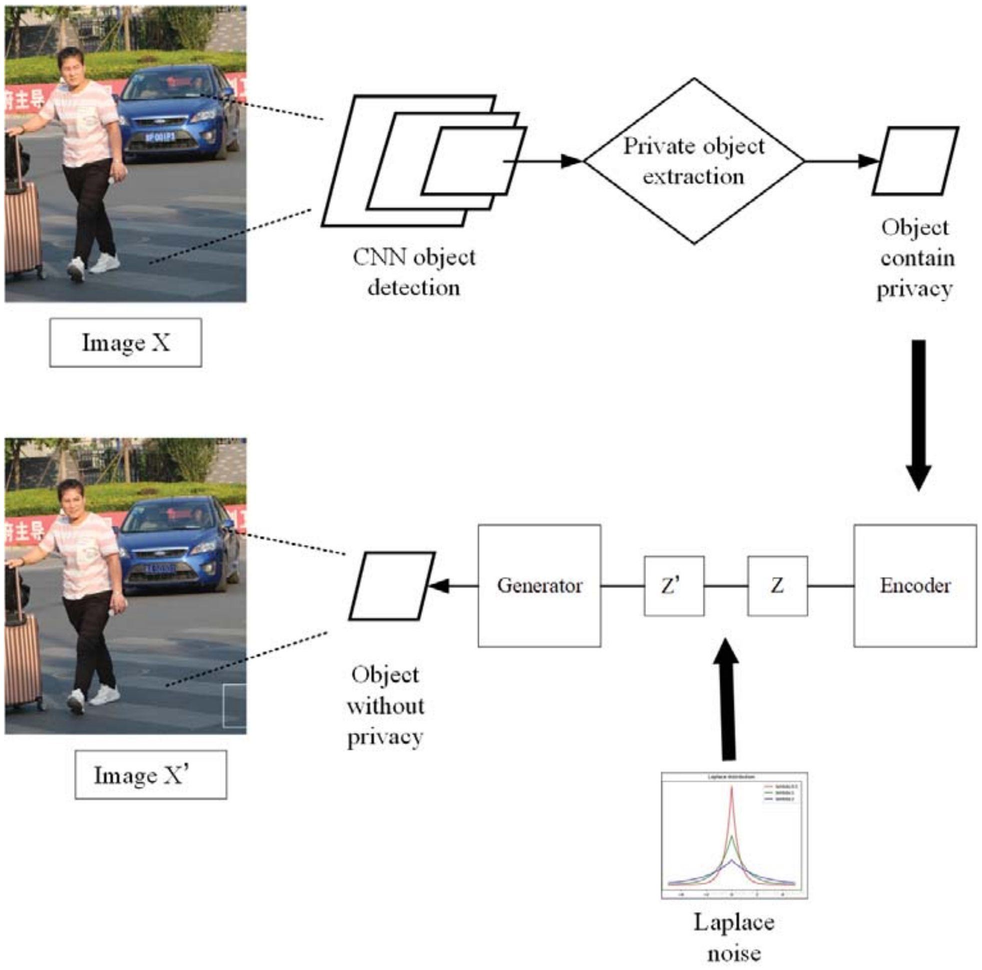


Figure 3. The diagram of the proposed image de-identification (DE-ID) framework.

In order to achieve the above goal of image privacy protection, we propose an image de-identification framework consists of three steps: (a) objects detection and private objects extraction; (b) de-identification content generation; and (c) content replacement and image privacy protection.

Fig. 3 shows the diagram of the framework. The original image X contains privacy information such as face and car plate. It is first input into a CNN to identify and extract the private objects in the image. Then we transform the extracted private objects into latent space and use differential privacy to control the de-identified content generation. Finally, we get a de-identified image X', i.e., image without any sensitive information. In the following part of this section, we will explain the framework in details.

4.1. Step-I:objects detection and private objects extraction

To protect the privacy of an image, it is necessary to detect the sensitive privacy zone in the image. We use two steps to achieve this target. First, all objects in the image are detected, and then the included private objects are extracted.

4.1.1. Objects detection

The state-of-the-art object detection algorithm Mask-RCNN is used to detect the objects in the image.

For an image I, the ROI (region of interest) vector $\boldsymbol{X}_{roi}$ of each object O_i can be detected by $R(\cdot)$:

$$\begin{aligned}\boldsymbol{X}_{roi} &= R(I) = (\boldsymbol{P}|\boldsymbol{E}_p)\\ &= \left(\begin{array}{cccc|cccc} x_1 & y_1 & w_1 & h_1 & p_{11} & p_{12} & \dots & p_{1m} \\ x_2 & y_2 & w_2 & h_2 & p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & w_n & h_n & p_{n1} & p_{n2} & \dots & p_{nm} \end{array}\right),\end{aligned} \tag{5}$$

where $\boldsymbol{P}_n = (x_n, y_n, w_n, h_n)$ is position vector including the information of up left corner coordinate (x_i, y_i), width w_i and height h_i of object O_i. The probability of objects noted as $\boldsymbol{E}_p$, the $\boldsymbol{E}_{pi}$ is the probability of Object O_i belonging to the mth class (there are m class objects in the image I).

In (5), we choose the maximum probability c_i in each $\boldsymbol{E}_{pi}$, so the output of the object detection shown as blew:

$$X_c = (\boldsymbol{P}|\boldsymbol{C}_p) = \left(\begin{array}{cccc|c} x_{p1} & y_{p1} & w_{p1} & h_{p1} & c_1 \\ x_{p2} & y_{p2} & w_{p2} & h_{p2} & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{pn} & y_{pn} & w_{pn} & h_{pn} & c_n \end{array}\right), \tag{6}$$

where $\forall i \in (1, n)$:

$c_i = \begin{cases} \text{arg}\, max(p_{ij}), 1 \le j \le m; \text{if } max(p_{ij}) > \delta \\ \qquad\qquad c_{bg}, \text{if } max(p_{ij}) \le \delta \end{cases}.$

In Mask-RCNN, if the maximum probability is smaller than a thresholdδ, this object will be treated as the background class, otherwise the object belongs to class i.

4.1.2. Private Objects extraction

After getting the objects' information and position, we set a classifier to classify the objects as either private of non-private. In our the street View experiment scene, the private objects can be human face, car plates, etc. The non-private objects can be as background, tree, traffic lights.

The extraction process is finished by $D(\cdot)$ accordingly as shown in (7).

$$\begin{aligned} D(X_c) &= D\left(\begin{array}{cccc|c} x_{p1} & y_{p1} & w_{p1} & h_{p1} & c_{p1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p\alpha} & y_{p\alpha} & w_{p\alpha} & h_{p\alpha} & c_{p\alpha} \\ \hline x_{np1} & y_{np1} & w_{np1} & h_{np1} & c_{np1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{np\beta} & y_{np\beta} & w_{np\beta} & h_{np\beta} & c_{np\beta} \end{array}\right). \\ &= \left(\begin{array}{cccc|c} x_{p1} & y_{p1} & w_{p1} & h_{p1} & c_{p1} \\ x_{p2} & y_{p2} & w_{p2} & h_{p2} & c_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p\alpha} & y_{p\alpha} & w_{p\alpha} & h_{p\alpha} & c_{p\alpha} \end{array}\right). \end{aligned} \tag{7}$$

So we got the private objects' position, class, and pixel information. The private objects' information is represented as follows:

$$\boldsymbol{X}_i = D(\boldsymbol{P}|\boldsymbol{C}_p) = \left(\begin{array}{cccc|c} x_{p1} & y_{p1} & w_{p1} & h_{p1} & c_{p1} \\ x_{p2} & y_{p2} & w_{p2} & h_{p2} & c_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p\alpha} & y_{p\alpha} & w_{p\alpha} & h_{p\alpha} & c_{p\alpha} \end{array}\right) \tag{8}$$

4.2. STEP-II: De-identification content generation

In the second step, we use a content generator $G(\cdot)$ and the differential privacy method to generate the de-identification content. The algorithm shown as below:

Algorithm 1: Image De-identification Content Generation

```
**Input:** The original image $I \in \Re^{n \times m \times 3}$ to de-identify; A pre-trained generator $G(\cdot)$.
**Output:** The de-identified image $I_{de}$ optimized via $G(\cdot)$
Initialize latent vector $\omega$, differential privacy Laplace noise with $\Delta f$ and $\epsilon$;
while not converged do
    $I \simeq I' = G(\omega^*)$;
end
$I_{de} = G(\omega^* + Lap(\frac{\Delta f}{\epsilon}))$ ;
```

Firstly, we find the latent vector ω^* of each input image I. Initialize a latent vector ω and search for a optimized vector ω^* minimizes the loss function (9) that measures the similarity between the private object image and image generated by latent vector ω^*. This step enables the image editable.

$$\omega^* = \arg\min_{\omega} \mathcal{L}_{percept}(G(\omega^*), I) + \frac{\lambda_{mse}}{N} ||G(\omega^*) - I||_2^2, \tag{9}$$

where image $I \in \Re^{n \times m \times 3}$ is the input privacy image. $G()$ is the pre-trained generator, N is the number of scalars in the image, ω is the latent code to optimize, $\lambda_{mse} = 1$.

Secondly, after we got latent vector ω^* of each private objects, we put the Laplace noise on latent vector ω^*. Then put the new latent vector into the generator $G(\cdot)$ and got the de-identify content.

$$I_{de} = G(\omega^* + Lap(\frac{\Delta f}{\epsilon})) \tag{10}$$

In equation (10), we used the DP criterion to protect the sensitivity information of the image. The Laplace mechanism was used. Generally speaking, the Laplace mechanism adds a controlled Laplace noise to a query result before returning it to the user. Here, the Laplace noise is sampled from a Laplace distribution, which is showed in (11).

$$Lap(x) = \frac{1}{2b} exp(-\frac{|x|}{b}) \tag{11}$$

To sum up, the Laplace mechanism can be summarized as

$$M(D) = f(D) + Lap(\frac{\triangle f}{\epsilon}) \tag{12}$$

The Laplace mechanism in (12) indicates that the size of the Laplace noise is related to the sensitivity of query f and the privacy budget ϵ. A larger sensitivity leads to a higher noise. In our method, we use privacy budget ϵ to control our GAN generator to generate the synthetic de-identify content.

4.3. STEP-III: De-identification content replacement

After de-identification contents generated, we use the generated content to replace the original private object images. The algorithm is shown in ??.

Finally, we get the de-identified image I_{de}.

Algorithm 2: Image Protected by de-identification content swapping

```
Input: The original image $I \in \Re^{n \times m \times 3}$ contains private content $X_i, i = 1, 2, ..., N$;
de-identified contents in the image: $X_i^{de}, i = 1, 2, ..., N$
Output: The protected image $I_d \in \Re^{n \times m \times 3}$
for each $X_i^{de}$ in $\mathbf{X^{de}}$ do
    $\mathrm{X} \xleftarrow{swapping} X_i^{de}$
end
$\mathrm{I}_{de} = I(X \xleftarrow{swapping} X^{de})$
```

5. Experiments and Discussions

5.1. Experiment Setup

First of all, we set up an experiment database contains amount of street view images collected by IoMT technology. The street view images contains human faces, car license plates, road signs, traffic lights and more. In these images, the sensitive private information are human faces and car license plates. In our test database, the human faces and car plates are the private objects, and the road sign, the traffic light and background are the non-private objects. We use the camera to collect over 4000 typical street view images as the test database.

5.2. Performance Evaluation Metrics

5.2.1. Privacy metrics

Confidence Score. In the privacy protection metric for human face, we use the open-source "face recognition" platform to evaluate the confidence in face privacy. This platform was built using dlib's state-of-the-art face recognition which was built with deep learning. The model has an accuracy of 99.38% on the Labeled Faces in the Wild benchmark. The output of the platform is the facial distance between each unrecognized face and the recognized face. By setting the corresponding threshold, the distance metric can judge whether the face is protected. This means after the face photo is processed by our method, whether the general third-party platform still considers the same person. The default threshold is 0.3.

Distance. In the privacy protection metric for the car license plate, due to the license plate is a set of characters, we believe that the distance between the original license plate and the processed license plate is the privacy metric. In the experiment, we set the threshold of the car license plate for 3. This means that the sensitive information of the license plates is protected when the distance is greater than 3.

5.2.2. Image utility metrics

The quantitative judgment is necessary for the degree of modification between the original image and the protected image. So we use several metrics to calculate the degree of modification, these metrics include L_0, L_2, ALD_p, Structural similarity index(SSIM), and difference value hash(Dhash). Deciding there are two images: processed image X^a and original image X, the utility image metrics are:

The L_0 calculate the number of pixels changed.

$$L_0 = num(X^a, X) \tag{13}$$

where num is calculated the number of pixels changed between X^a and X.

The L_2 calculate Euclidean distance between the original image and protected image.

$$L_2 = ||X^a - X||_2 = \sqrt{\sum_{i=1}^{N}(X_i^a - X_i)^2} \tag{14}$$

The ALD calculate the average L distance between the images.

$$ALD_p = \frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i^a - X_i\|_p}{\|X_i\|_p} \tag{15}$$

The SSIM is the common method to evaluate the similarity between the original image and the protected image.

$$SSIM(X^a, X) = \frac{1}{n}\sum_{i=1}^{n} SSIM(X_i^a, X_i) \tag{16}$$

The Dhash use the difference hash to evaluate the degree of modification which value is the smaller the better.

$$Dhash(X^a, X) = hash(X^a) - hash(X) \tag{17}$$

5.3. Street view image protection

5.3.1. Human face privacy protection

Human face is the most sensitivity information of the IoMT images, which can straight leak personal identification. Therefore, we use our method to protect the human face privacy in the street view experimental scene. Firstly, we use Mask-RCNN to extra the human face images I from the experimental street view images. Secondly, initialize a latent vector ω and use the loss function (18) to find the latent vector ω^* of human face I. The algorithm to find the latent vector ω^* was shown in algorithm (3).

```
Algorithm 3: Human face Image Projecting into Latent Space
    Input: A human face image $I \in \Re^{n \times m \times 3}$ to project; a pre-trained generator $G(\cdot)$
    Output: The latent code $\omega^*$ and the projected image $G(\omega^*)$ optimized via $F'$
    Initialize latent code $\omega^* = \omega$
    while not converged do
        L $\leftarrow L_{precept}(G(\omega^*), I) + \frac{\lambda}{N}||G(\omega^*) - I||_2^2$
        $\omega^* \leftarrow \omega^* - \eta F'({}_{\omega}L)$
    end
```

In algorithm (3), the loss function was show in (18).

$$\omega^* = min_{\omega} L_{percept}(G(\omega^*), I) + \frac{\lambda_{mse}}{N}||G(\omega^*) - I||_2^2 \tag{18}$$

where image $I \in \Re^{n \times m \times 3}$ is the input privacy image. $G()$ is the pre-trained generator, N is the number of scalars in the image,ω is the latent code to optimize,$\lambda_{mse} = 1$. For the loss term $L_{percept}$, shows as below:

$$L_{percept} I_1, I_2 = \sum_{j=1}^{4}\frac{\lambda_j}{N_j}||F_j(I_1) - F_j(I_2)||_2^2 \tag{19}$$

where $I_1, I_2 \in \Re^{n \times m \times 3}$ are the input images, F_j is the feature output of VGG-16 layers conv1_1, conv1_2,conv3_2,conv4_2.N_j is the number of scalars in the jth layer output,$\lambda_j = 1$ for all js are empirically obtained for good performance.

Fig. 4 is an example of the original human face image and the human face generated by GAN with no modify.

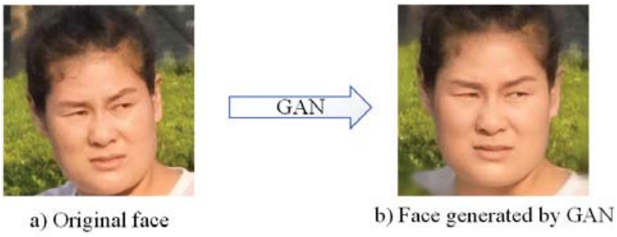


Figure 4. The original face image projecting into StyleGAN

Thirdly, put the Laplace noise on the latent vector ω^* and use the generator $G(\cdot)$ to generate the de-identify human face image.

$$I_{de} = G(\omega^* + Lap(\frac{\Delta f}{\epsilon})) \tag{20}$$

Finally, use the de-identify human face image to swap the original human face image. In this step, we use Dlib, which is a toolbox in Opencv based on key-point face detection, to get the 68 key points of the human faces and use seamless cloning to swap the face. The face swapping algorithm can transfer the input face features to the target face without obtrusive. An example result is shown in Fig.5 (d). Intuitively speaking, a larger Laplace noise leads to a more different human face compared with the original human face.

In our experiments, we use Laplace noise parameter ϵ to control the distance between de-identify human face images and private human face images. In addition, we use the open-source "face recognition" platform to determine if the synthetic face and the original face represent the same person.

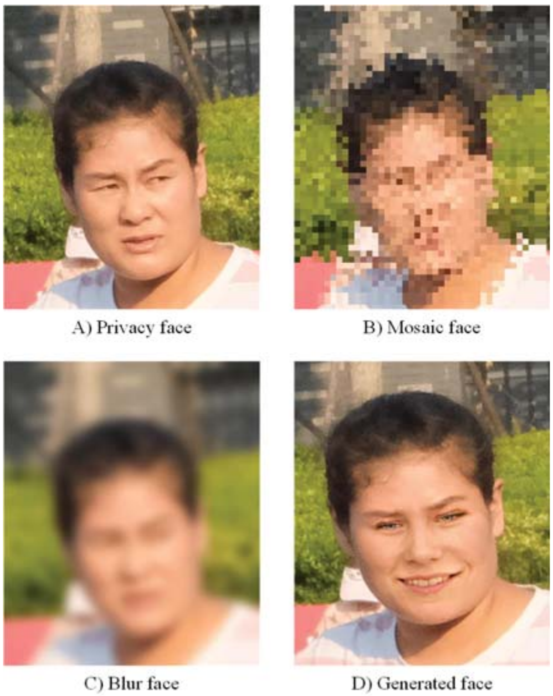


Figure 5. Face images comparison: A) Face in street view, B) Mosaic methods, C) Blur method, D) Our method

Fig.5 shows the original private face, the mosaic face, the blurred face, the the new face generated by StyleGAN. It can be seen that it is not easy for both human and machine to recognize the de-identify generated face image Fig.5 D as Fig.5 A.

5.3.2. Car license plate privacy protection

The car license plates are another kind of sensitive objects of IoMT images. As for the privacy protection of the car license plates, we use Chinese car license plates as our experimental objects. The car plate should be generated according to the rules enforced by the vehicle management authority. The rules of a valid Chinese car license plate are: 1) the first character is a Chinese character, representing a province; 2) the second symbol is an English letter; 3) the last five symbols forms a random string of letters and numbers, and 4) the background of a license plate is dark blue.

After getting the car license plates images from the street images, we use OCR to recognize the characters and symbols of the car license plates, and then map the car plate into a sequence of numbers. According to the Chinese car plate rules, the first character will be one of 31 Chinese province abbreviation characters (except special district). Because of the first Chinese character represents location information, we map them into 2-digit numbers 00-30 based on the sorted distances from each province to the capital city Beijing. The mapping table for the first character is shown in Table ??.

Next, the numerical values 0-9 will be translated into 2-digit codes 00-09, and the English symbols will be translated into 2-digit codes 10-33. For example, a car plate "Beijing A132B3" will be mapped to a sequence of numbers "00 100103021103". After we translate each car plate into a sequence of numbers, we add Laplace noise onto the number sequence and obtain a synthetic number sequence satisfying DP. In Laplace noise generate, we let the $\Delta f = 1$ and control the ϵ to generate the Laplace noise. For example, if we add a random Laplace noise on the above car plate "00 100103021103", and obtain a perturbed sequence as "03 130214231502", which can be translated to a synthetic car plate "Hebei D2ENF2". The above example is illustrated in Fig.6. And there is a cyclic shifting if the Laplace noise makes the value out of the bounds, e.g. the province code > 33.

Then, we use the generator to generate a synthetic car plate image according to the car plate code. Finally, we swap the car plate with the synthetic car plate image. The synthetic car plate is protected by the DP criterion.

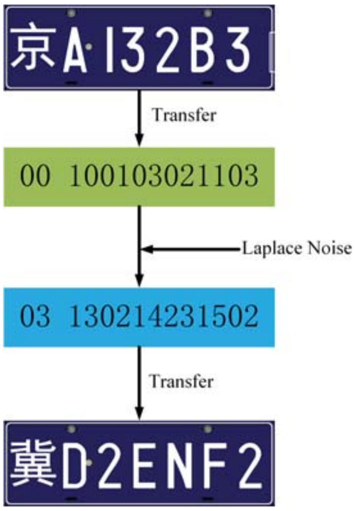
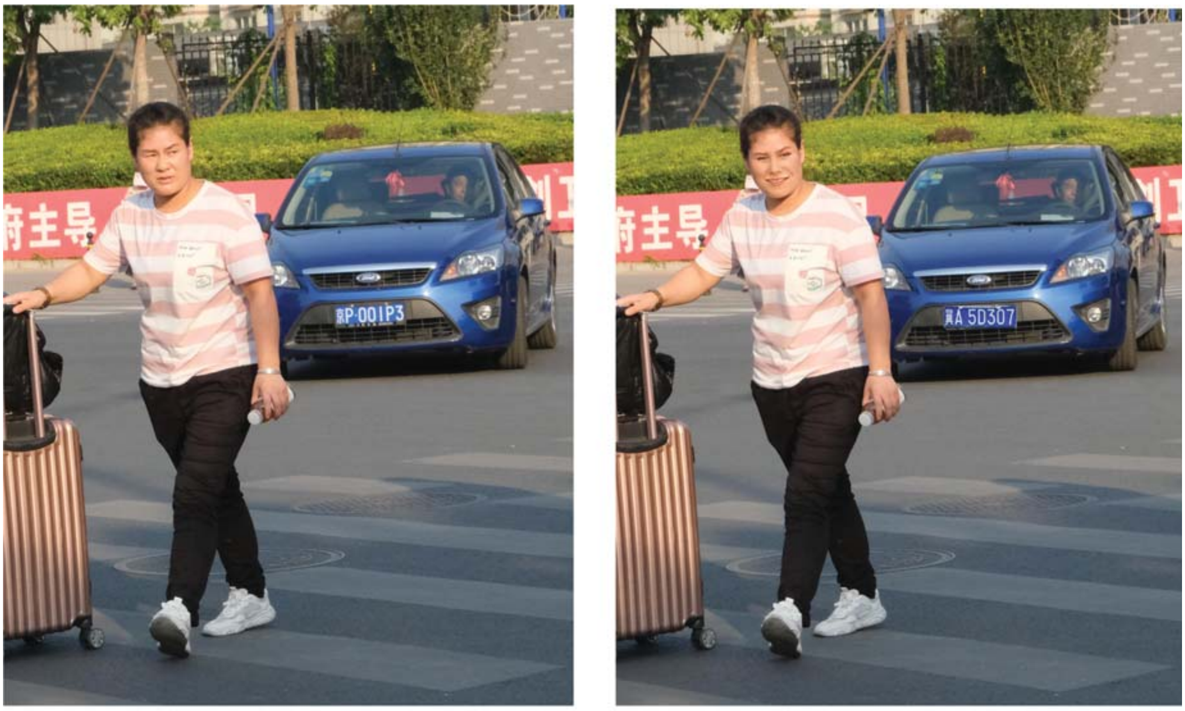


Figure 6. A new car plate content created by DP

In the car number transfer, the larger the noise, the longer distance(original car number as origin) car number will be generated. For example, if a province name is Jilin in a car plate, the province codes should be generated for Jilin based on the distance from the other provinces to Jilin.

Our method uses the synthetic DP car plate to protect the private car plate information. As shown in Fig. 7, we can see that the car plate is smoothly replaced by our the synthetic car plate.


A) Original street view image

B) Our method

Figure 7. A typical Chinese car plate swap to protect the street view image

It is very important to note that the replacement of the privacy content in a image is not simply a copy-and-paste job. Instead, it needs to transform the synthetic content by generator into an image that fits into the original image area with a correct orientation.

Therefore, the synthetic image is generally not perceptible to human eyes.

5.4. Performance Evaluation

5.4.1. Privacy protection metrics

In this part, we calculate the distance between the original private image and protected image to measure the degree of privacy protection.

In human face, the average facial distance between the same person is 0.12,which confidence score is 88. After using our method processed, the average facial distance is 0.45 and confidence score is 55, which is over the threshold of confidence score 70. This experiment result means our method can remove the identity of the human face, which means our method can protect the privacy of the human face image.

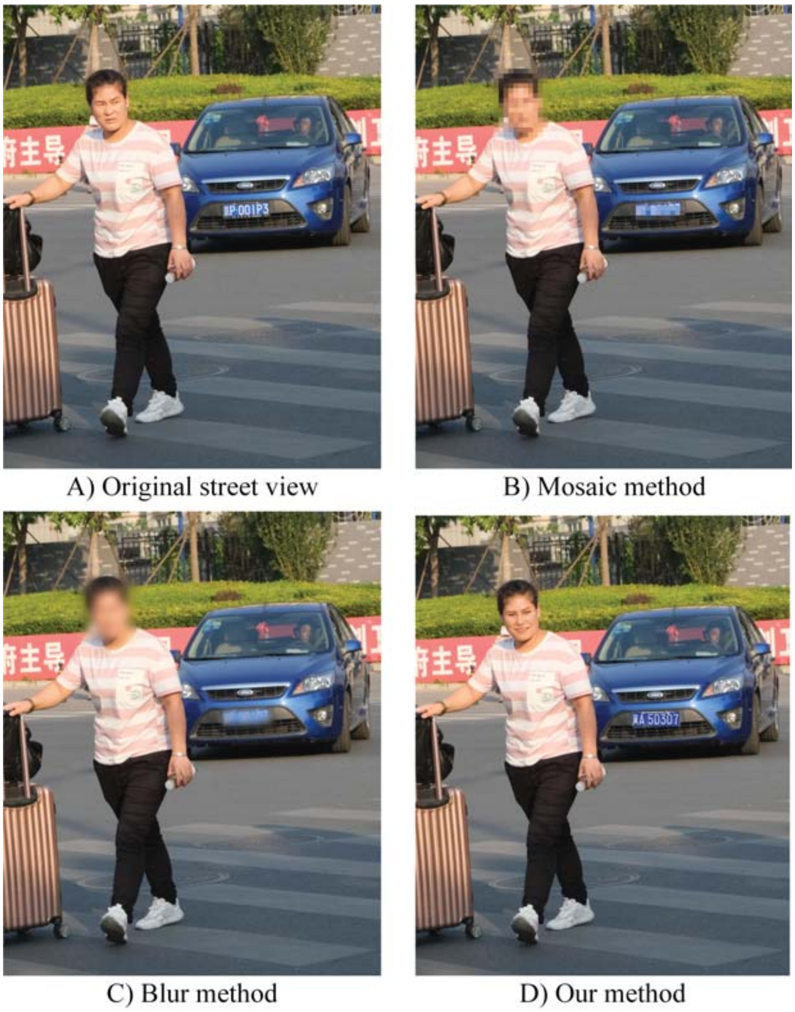
In car license plate, because of the license plates are strings, their distances are integers. In the experiment, the distance between the same license plate is 0. After using our method processed, the distance is 3, which we can consider that the sensitive information of the license plate is protected.

5.4.2. Image utility metrics

In this part, we set an automatic evaluate module to calculate the degree of image modification by different metrics through L_0, L_2, ALD_p, $SSIM$, and $Dhash$. We compare our method with the Blur and Mosaic methods. As shown in Fig 8, the Blur and Mosaic remove the sensitive area of privacy.

However, a human can easily notice the blur and mosaic in the image. Hence, the computer can easily recover the information from the processed image.[38][39][40]

In our method, we control the generator to generate the de-identify content image with DP Laplace noise. The de-identify images make both human eyes and computer vision detection methods not easily to see the difference and get the privacy information on sensitive private objects. The result of the street view image shown in Fig. 8, as we can see, human and computer can easily detect the sensitive information in unprotected street view image in Fig. 8 A). And in Fig. 8 B) and 8 C), the algorithm can not detect the face and the car plate after being blurred, but human can easily see there are blur or mosaic in the image. In Fig.8 D), the computer algorithm and human detect the fake sensitive information which had already swapped by our method, so both human and computer can not get the real sensitive information of the face and the car plate. The privacy in the image is protected under our method.


A) Original street view

B) Mosaic method

C) Blur method

D) Our method

Figure 8. The result of 4 street view image: A) unprotected image, B) image processed with blur, C) image processed with mosaic, D) image processed with our method

Next, we use metrics to evaluate the efforts of our methods. Table 1 shows the performance of our method, blur, and mosaic. The metrics are DHash, SSIM, L_0, L_2 and ALD_p. The blur and Mosaic have been modified to change the sensitive area in our experiment images.

First, compared with other methods, our methods change the minim pixels to protect the privacy part of the image. In Dhash, Our method is better than the others. Compared with blur and mosaic, our method decreases 95.02% and 95.2%. In SSIM, our method is better than others in 1.17% and 1.67%.

In L_0, Our method decreases 73.6% and 72.97%. In L_2, 86.25% and 25.99%. In ALD_p, our method is higher than blur and mosaic, which is 160.65% and 98.85%. It is shown that our method is better than the other two methods in the SSIM, Dhash and L_0. However, the results show that in the L_2 and ALD_p, our method is not the best. After analysis, we found that the L_2 and ALD_p are more suitable in big area modification. These metrics are not sensitive to minor modifications. So we use the face swap as an example to show the metrics in the minor modification in a small area. So we choose 89 face swap images to analysis, the result shows in Table 2. In Dhash, compared with blur and mosaic decreases 96.68% and 96.97%. In SSIM, increase 50.67% and 102.24%. In the L_0, decrease 76.55% and 76.84%. In L_2, decrease 64.93% and 81.08%. In ALD_p, decrease 65.11% and 79.68%. As we can see, our method is the best in all metrics.

Table 1. Average result of 4000 images with the metrics: Dhash, SSIM, L_0, L_2, ALD_p

Methods	Original	Blur	Mosaic	Our methods
Dhash	0	12873.65	13370.19	641.71
SSIM(10^{-2})	100	98.18	97.70	99.33
$L_0(10^2)$	0	1692.25	1652.57	446.74
L_2	0	9983.06	14757.19	18593.41
$ALD_p(10^{-2})$	0	3.99	5.23	10.4

6. Conclusion

This paper proposes a new image privacy protection method based on GAN and DP. Our method can protect the sensitive private information contained in the images of IoMT. We use the deep neural network to identify the private data in the images and de-identified it with the GAN-based content. Compared with traditional blur or mosaic methods, the proposed method can protect the sensitive information of image data, avoid the privacy leakage. The experimental results of IoMT collection image data show that our privacy protection method can protect the privacy with high efficient and controlabilty. In future work, we will study the privacy protection on video of IoMT and improve the real-time nature of our method. Propose high effectively privacy protection method for the privacy of IoMT.

Appendix A.

Appendix A.1.

References

1. EU. The EU General Data Protection Regulation . https://eugdpr.org/, 2019. [Online; accessed 19-July-2019].
2. EU. What is considered personal data? https://gdpr.eu/eu-gdpr-personal-data/, 2019. [Online; accessed 19-July-2019].
3. Mannan, M.; van Oorschot, P.C. Privacy-enhanced sharing of personal content on the web. Proceeding of the 17th international conference on World Wide Web - WWW '08; ACM Press: Beijing, China, 2008; p. 487. doi:10.1145/1367497.1367564.

Table 2. Average result of 4000 face images with the metrics: Dhash, SSIM, L_0, L_2, ALD_p

Methods	Original	Blur	Mosaic	Our methods
Dhash	0	4047.80	4427.79	134.25
SSIM(10^{-2})	100	64.63	48.15	97.38
$L_0(10^2)$	0	1009.4	1022.25	236.72
L_2	0	5832.96	10812.2	2045.48
$ALD_p(10^{-2})$	0	16.68	28.64	5.82

Table A1. The transfer 2-digit code based on distance between Beijing and each province of China

Province Name	Distance to Beijing (km)	2-digit code
Beijing	0	00
Tianjin	96.07188	01
Hebei	239.4603	02
Shandong	356.9375	03
Shanxi	407.3106	04
Neimengu	424.5428	05
Henan	620.2232	06
Liaoning	630.724	07
Jiangsu	860.7032	08
Jilin	867.213	09
Ningxia	884.2019	10
Anhui	897.8403	11
Shanxi	907.8513	12
Hubei	1041.318	13
Shanghai	1041.987	14
Heilongjiang	1056.846	15
Zhejiang	1102.843	16
Gansu	1184.73	17
Jiangxi	1242.833	18
Hunan	1316.041	19
Qinghai	1340.82	20
Chongqing	1419.309	21
Sichuan	1505.931	22
Fujian	1527.525	23
Guizhou	1729.627	24
Guangdong	1856.641	25
Guangxi	2047.263	26
Yunnan	2068.306	27
Hainan	2249.545	28
Xinjiang	2433.955	29
Xizang	2559.149	30

4. Vyas, N.; Squicciarini, A.C.; Chang, C.C.; Yao, D. Towards automatic privacy management in Web 2.0 with semantic analysis on annotations. Proceedings of the 5th International ICST Conference on Collaborative Computing: Networking, Applications, Worksharing; IEEE: Crystal City, Washington DC, USA, 2009. doi:10.4108/ICST.COLLABORATECOM2009.8340.
5. Wang, N.; Xu, H.; Grossklags, J. Third-party apps on Facebook: privacy and the illusion of control. Proceedings of the 5th ACM Symposium on Computer Human Interaction for Management of Information Technology - CHIMIT '11; ACM Press: Cambridge, Massachusetts, 2011; pp. 1–10. doi:10.1145/2076444.2076448.
6. Squicciarini, A.C.; Xu, H.; Zhang, X.L. CoPE: Enabling collaborative privacy management in online social networks. *Journal of the American Society for Information Science and Technology* **2011**, pp. n/a–n/a. doi:10.1002/asi.21473.
7. Viola, P.; Jones, M.J. Robust real-time face detection. *International journal of computer vision* **2004**, *57*, 137–154.
8. Tonge, A.; Caragea, C. Image privacy prediction using deep neural networks. *ACM Transactions on the Web (TWEB)* **2020**, *14*, 1–32.
9. Yu, J.; Zhang, B.; Kuang, Z.; Lin, D.; Fan, J. iPrivacy: Image Privacy Protection by Identifying Sensitive Objects via Deep Multi-Task Learning. *IEEE Transactions on Information Forensics and Security* **2017**, *12*, 1005–1016. doi:10.1109/TIFS.2016.2636090.
10. Uittenbogaard, R.; Sebastian, C.; Vijverberg, J.; Boom, B.; Gavrila, D.; others. Privacy Protection in Street-View Panoramas Using Depth and Multi-View Imagery. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10573–10582.
11. Liu, Y.; Zhang, W.; Yu, N. Protecting Privacy in Shared Photos via Adversarial Examples Based Stealth. *Security and Communication Networks* **2017**, *2017*, 1–15. doi:10.1155/2017/1897438.
12. Liu, B.; Xiong, J.; Wu, Y.; Ding, M.; Wu, C.M. Protecting Multimedia Privacy from Both Humans and AI. 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). IEEE, 2019, pp. 1–6.
13. Liu, B.; Ding, M.; Zhu, T.; Xiang, Y.; Zhou, W. Adversaries or allies? Privacy and deep learning in big data era. *Concurrency and Computation: Practice and Experience* **2019**, *31*, e5102.
14. Xue, H.; Liu, B.; Ding, M.; Song, L.; Zhu, T. Hiding Private Information in Images From AI. 2020 IEEE International Conference on Communications (ICC). IEEE, 2020.
15. McPherson, R.; Shokri, R.; Shmatikov, V. Defeating image obfuscation with deep learning. *arXiv preprint arXiv:1609.00408* **2016**.
16. Pesce, J.P.; Casas, D.L. Privacy Attacks in Social Media Using Photo Tagging Networks: A Case Study with Facebook. p. 8.
17. Times, T.N.Y. San Francisco Bans Facial Recognition Technology. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html, 2019. [Online; accessed 19-July-2019].
18. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition; IEEE: Columbus, OH, USA, 2014; pp. 580–587. doi:10.1109/CVPR.2014.81.
19. Hariharan, B.; Arbelaez, P.; Girshick, R.; Malik, J. Hypercolumns for object segmentation and fine-grained localization. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE: Boston, MA, USA, 2015; pp. 447–456. doi:10.1109/CVPR.2015.7298642.
20. Girshick, R. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015, pp. 1440–1448.
21. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015, pp. 91–99.
22. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. p. 10.
23. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
24. Efros, A.; Leung, T. Texture synthesis by non-parametric sampling. Proceedings of the Seventh IEEE International Conference on Computer Vision; IEEE: Kerkyra, Greece, 1999; pp. 1033–1038 vol.2. doi:10.1109/ICCV.1999.790383.
25. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D.B. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. p. 10.

26. Pnevmatikakis, E.A.; Maragos, P. An inpainting system for automatic image structure - texture restoration with text removal. 2008 15th IEEE International Conference on Image Processing; IEEE: San Diego, CA, USA, 2008; pp. 2616–2619. doi:10.1109/ICIP.2008.4712330.
27. Bertalmio, M.; Vese, L.; Sapiro, G.; Osher, S. Simultaneous Structure and Texture Image Inpainting. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2003**, p. 6.
28. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]* **2014**. arXiv: 1406.2661.
29. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. Advances in neural information processing systems, 2017, pp. 971–980.
30. Mao, X.; Li, Q.; Xie, H.; Lau, R.; Wang, Z.; Smolley, S. Least Squares Generative Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2813–2821.
31. Xiong, W.; Yu, J.; Lin, Z.; Yang, J.; Lu, X.; Barnes, C.; Luo, J. Foreground-Aware Image Inpainting. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019, pp. 5833–5841.
32. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. Advances in neural information processing systems, 2017, pp. 5767–5777.
33. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019, pp. 4396–4405.
34. Sweeney, L. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **2002**, *10*, 557–570.
35. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **2007**, *1*, 3–es.
36. Li, N.; Li, T.; Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. 2007 IEEE 23rd International Conference on Data Engineering. IEEE, 2007, pp. 106–115.
37. Dwork, C. Differential privacy. Proceedings of the 33rd international conference on Automata, Languages and Programming-Volume Part II. Springer-Verlag, 2006, pp. 1–12.
38. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8183–8192.
39. Nah, S.; Kim, T.H.; Lee, K.M. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 257–265.
40. Menon, S.; Damian, A.; Hu, S.; Ravi, N.; Rudin, C. PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2437–2445.