International Journal of Networking and Computing – www.ijnc.org, ISSN 2185-2847
Volume 11, Number 1, pages 102–119, January 2021

Face-Image Anonymization as an Application of Multidimensional Data k-anonymizer

Taichi Nakamura
Graduate School of Science and Technology Keio University, Nishi Laboratory,
taichi@west.sd.keio.ac.jp
3-14-1 Hiyoshi, Kouhoku, Yokohama, Kanagawa 223-8522, Japan

Yuiko Sakuma
Graduate School of Science and Technology Keio University, Nishi Laboratory,
sakuma@west.sd.keio.ac.jp
3-14-1 Hiyoshi, Kouhoku, Yokohama, Kanagawa 223-8522, Japan

Hiroaki Nishi
Graduate School of Science and Technology Keio University, Nishi Laboratory, west@sd.keio.ac.jp
3-14-1 Hiyoshi, Kouhoku, Yokohama, Kanagawa 223-8522, Japan

Received: February 15, 2020
Revised: May 4, 2020
Revised: June 10, 2020
Revised: December 2, 2020
Accepted: December 29, 2020
Communicated by Kozo Ohara

Abstract

Recently, the developments of data communication networks and advancements in the processing capacity of computers have significantly increased the amount of data that can be applied to a service. These data might result in future innovations. However, the privacy violation of data has become a problem. For example, images of customers' faces captured with a surveillance camera may lead to new marketing strategies, as the reactions of the customers can be measured from their facial expressions. However, we must consider the privacy of customers when entrusting the images to a third-party data analyst. For solving this privacy problem, researchers have been developing anonymization technologies that are used to preserve privacy by deleting the private information from the original data. However, conventional anonymization techniques cannot appropriately anonymize high-dimensional data such as facial images. This is attributed to the fact that conventional anonymization techniques do not consider the complex relationships between dimensions and the semantic loss of data. However, studies regarding machine learning have been actively conducted; particularly, neural networks (NN) have developed remarkably since the advent of AlexNet[8]. Both machine learning and anonymization share the common idea in their bases to abstract statistical information from a given dataset . Therefore, the machine learning technology might enhance the functionality of anonymization techniques. In this study, we propose a method to apply the results of machine learning to anonymization based on the aforementioned common idea , and the method is named multi-input k-anonymizer unit (MIKU). Notably, MIKU has two modules called S and G maps for mapping a given data, and NNs are used for these modules to generate more natural anonymized data for humans than directly anonymized data, which is generated only by processing the pixel values of facial images. To evaluate MIKU, a direct anonymization method is used, without any NN, for the facial images. The facial images of CelebA [9] are used for both qualitative and quantitative evaluations. The qualitative evaluation is conducted by analyzing the anonymized facial images obtained using different methods, and the quantitative evaluation is performed using the Fréchet inception distance(FID) [4]. In the qualitative evaluation, there are cases where the quality of the anonymized images generated using the comparison method is low because unnatural edges and blurs are created on the anonymized facial images. However, MIKU maintains the quality and attributes of the original facial images even in those cases. In the quantitative evaluation, a different result is obtained when $k = 2$ anonymity. However, on anonymity greater than 2, the facial images anonymized using MIKU have higher quality than those anonymized using the comparison method.

Keywords: k-anonymization, Facial Image, High-Dimension, StyleGAN, Neural Network

1 Introduction

Computing and communication capacities are increasing with each year. Accordingly, it has become easy to take and share high-precision images anytime and anywhere over the Internet, even if using small terminals. In this circumstance, secondary uses of these data might result in further economic development. For example, facial images can be obtained using various devices, such as surveillance cameras and smartphone cameras. In this situation, disclosing the facial images without the permission of the individuals causes privacy concerns. Hence, the noise-insertion approach is often used. For a facial image, blurring and blacking methods are frequently used to perform data anonymization. However, the images processed using these methods are not natural, and their image quality is degraded.

For example, when publishing a group photo of individuals, it will be appropriate to blur the faces of other people in consideration of their privacy in the image. However, processing such a mosaic significantly damages not only the information but also the quality of appearance. Therefore, if there is a technique for replacing the face of another person with an anonymous face, it would be possible to make an anonymized group photo that simultaneously retains a certain amount of information of the group and has a pleasant appearance. Although there have been technologies, such as k-anonymization, to perform anonymization by replacing data, they are ineffective for data with high dimensions.

Meanwhile, the recent development of machine learning, as represented by neural networks (NNs), has been remarkable. Although these techniques appear to be different from anonymization, both extract statistical information from a dataset. Particularly, a NN has a hidden layer, which is expected to obtain the intermediate information between the input and output of the target NN.

In this study, we proposed the multi-input k-anonymizer unit (MIKU) that maps the facial image to the latent space of StyleGAN and anonymizes images in that space. For mapping facial images, MIKU has two mapping modules that names G and S map. Owing to G and S map, MIKU can apply NNs to anonymization. MIKU is evaluated qualitatively and quantitatively by comparing with the direct anonymization method that does not use any NNs. In both evaluations, the facial images of CelebA are used for pre-anonymization data. In the qualitative evaluation, we observe anonymized images and evaluate the quality of them. In the quantitative evaluation, the Fréchet inception distance(FID), which is one of the most authorized quality indices of images, was used for the index of anonymized images quality. We defined the appropriate degree of grouping as the label Euclidean distance given to face images before and after anonymization. We measure the indices of all face images anonymized by MIKU and the conventional method. Consequently, MIKU can make better quality anonymized images than the conventional method.

Table 1: Row example of data for k-anonymity

ID	ZIP-CODE	AGE	GENDER	DISEASE
t1	0123	22	FEMALE	CANCER
t2	0124	24	MALE	FLU
t3	0125	26	MALE	AIDS
t4	1220	31	MALE	COLD
t5	1221	39	MALE	FLU

Table 2: Anonymized example for Table 1 on k-anonymity

ID	ZIP-CODE	AGE	GENDER	DISEASE
t1	012*	24	*	CANCER
t2	012*	24	*	FLU
t3	012*	24	*	AIDS
t4	122*	35	MALE	COLD
t5	122*	35	MALE	FLU

2 Related research

2.1 k-anonymization

k-anonymity [15] is a typical privacy-preservation method that considers the anonymization level. Before explaining k-anonymity, the definitions of data table, index, attribute, identifier, and quasi-identifier are given as follows.

2.1.1 Data table

A data table is a table that constitutes a database. Its rows and columns are called the tuple and field, respectively. Table 1 is an example of a data table.

2.1.2 Index and Attribute

The heading of each tuple and field is called the index and attribute, respectively. Typically, an index is assigned to each user or each data sample, and an attribute indicates the data content. For example, in Table 1, ID is an index, and zip code, age, gender, and disease are attributes.

2.1.3 Identifier, quasi-identifier, and sensitive attribute

An identifier is an attribute that is directly connected to the personal information that is unique to an individual; a social security number can be a typical example of such information. Identifiers are typically deleted when data are anonymized. A quasi-identifier is an attribute that is used to identify individuals by combining it with other quasi-identifiers, such as zip code, gender, position information, and purchase history. Quasi-identifiers are anonymized when the anonymized data completely satisfy a given anonymization level. However, sensitive attributes are not anonymized, as they are important for secondary use and data analysis. Notably, k-anonymity guarantees that the data have at least $k-1$ or more identical quasi-identifiers for any individuals. The process of obtaining k-anonymity is called k-anonymization. Here, the data of Table 1 are used for explanation. Table 2 shows an example of $k=2$ anonymized data generated using the data of Table 1. The ID is an identifier and has to be deleted when it is anonymized. Here, ID is left for the convenience of the before-and-after comparison of anonymization. In Table 2, the disease information is protected by editing the zip code, age, and gender. $(t1, t2, t3)$ and $(t4, t5)$ can be grouped together. As these groups share the same quasi-identifiers (i.e., zip code, age, and gender), an individual cannot be exactly identified using these. In k-anonymization, individuals with similar quasi- identifiers are first divided into groups of size k or more; subsequently, the quasi-identifier of each group is unified. Here, unification means replacing the operation of a quasi-identifier with a value computed using a set of quasi-identifiers in the same group. For example, the following processes are conceivable for anonymization: only the upper digits of zip code are retained via masking; age is approximated using average values; gender is erased.

In k-anonymization, grouping is critical to minimizing the information loss. However, performing the optimal grouping is an NP-hard problem [11]. Therefore, heuristics methods have been proposed for the cases in which the optimal method could not be used [7]. However, because heuristic methods only consider data that have hierarchical structures, they cannot be applied to non-hierarchical data (i.e. image data) used in this study. Therefore, in this study, Mondrian is used for performing grouping. Although various Mondrian implementations exist, the grouping process outlined in Algorithm 1 is used in this study.

Algorithm 1: Mondrian

$i\,(i \in N)$denotes an individual, and v_i denotes the data of individual $\left(v_i \in R^d\right)$. The number of dimensions to search is represented by N_s $(N_s \leq d, N_s \in N)$, and V denotes the set of individuals. Here, $v_{i,j}$ indicates the value of the j-th dimension of the data of individual i, and the initial value of V is $\{1, 2, 3, \dots\}$. k represents k-anonymity, and G denotes the groups' set that is the result of this grouping algorithm.

Step 1 : If the number of elements in V is less than $2k$, add V to G.
Step 2 : Randomly create a set A of N_s integers ranging from 1 to d without duplication.
Step 3 : On the dimension j contained in A, calculate j^* by the following equation.

$$j^* = \arg\max_{j \in A} \left(\max_{i \in V} v_{i,j} - \min_{i \in V} v_{i,j} \right)$$

Step 4 : V is sorted by the j^*-th dimension and divided into V_a and V_b as the first and second halves of V, respectively.
Step 5 : Consider V_a and V_b as V and repeat **Step 1**, recursively.

Algorithm 1 is a recursive algorithm that outputs each grouping result in Step 1. In Step 1, if the size of set V is $2k$ or more, then the subsequent steps are performed because V can be divided into two sets with each having k or more elements. If the size of set V is less than $2k$, V is added to G. This is because k-anonymity cannot be satisfied when it is further divided. In Steps 2 and 3, N_s dimensions are randomly selected. Subsequently, one of them (j^*) is selected with the largest difference between the maximum and minimum values. This random selection reduces the calculation cost compared with full selection. In Step 4, V is sorted using j^* and divided into two halves. In Step 5, we recursively perform Step 1 on each of the divided sets. Algorithm 1 can specify the number of dimensions N_s to perform grouping at high speed, even for high-dimensional data. This can result in the processing of comparatively high-dimensional data, such as the facial images used in this study ($3 \times 1024 \times 1024$), in a constant calculation time; namely, the calculation time is not affected by the size of the images.

2.2 Problem of high-dimensional anonymization

$$\begin{gathered} 0.3 \leq t \leq 1.3 \\ 0 \leq y \leq 1 \\ x = t \sin 4\pi t \\ z = t \cos 4\pi t \end{gathered} \tag{1}$$

Generally, anonymizing high-dimensional data is difficult [2], as this involves two major problems. The first is a problem in grouping, and the second is a problem in averaging. Figure 1 shows a Swiss

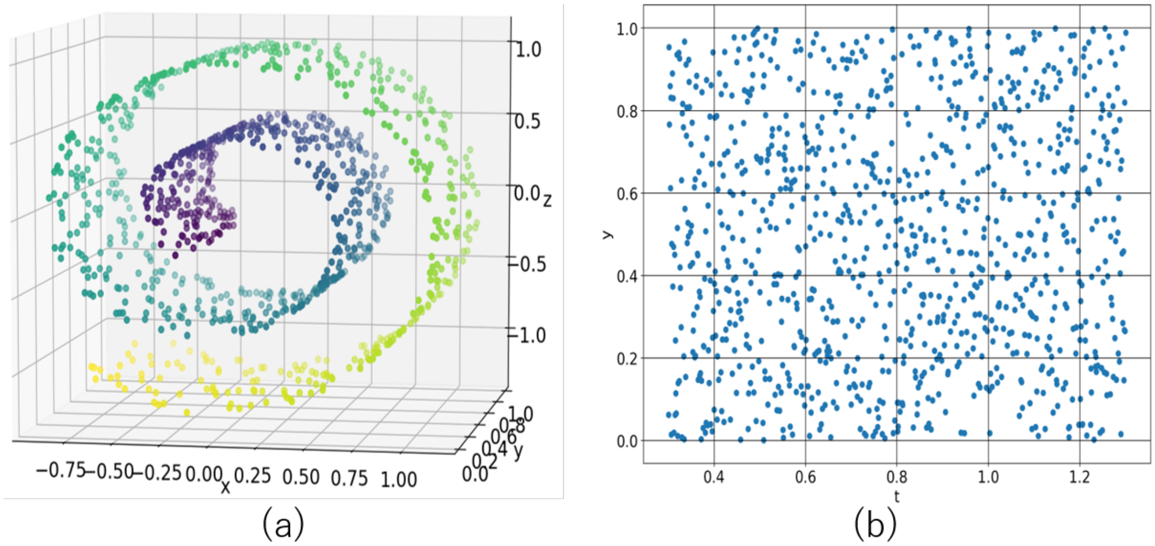


Figure 1: (a):Axes x,y, and z, (b): Axes t and y

roll, which is often used as an example of a manifold. The Swiss roll is calculated using (1).

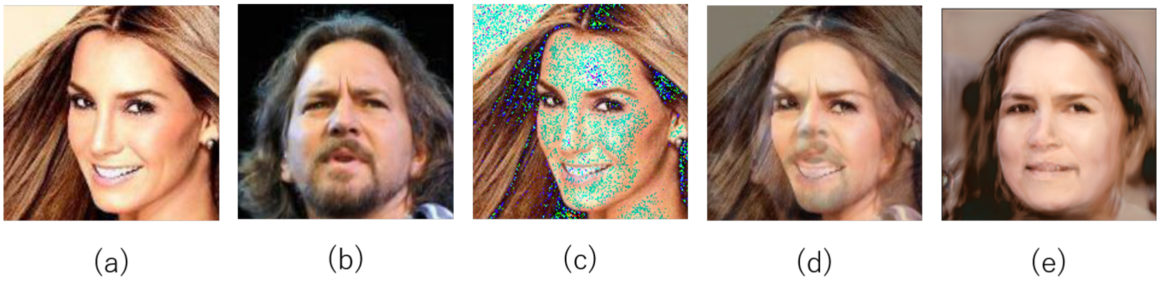


Figure 2: (a),(b) CelebA facial images. (c) Noise added to (a). (d) Mean of (a) and (b) in the pixel space. (e) Mean of (a) and (b) in the StyleGAN latent space

The first problem is attributed to using Euclidean distance as a distance function between data. For example, if the Euclidean distance is used in Figure 1 (a), then the points in different layers are evaluated as close data. However, when it is more appropriate to evaluate in the space using y and t as shown in Figure 1 (b) than the space in Figure 1 (a), performing clustering using the Euclidean distance on the space in (a) will not be satisfactory because distant points would be classified into the same cluster. It is assumed that the same problem occurs for facial images. For example, Figures 2 (a) and (b) show two facial images selected from CelebA [9]. Gaussian noise is added to the image shown in (a) to create the image depicted in (c) to make the Euclidean distance in the pixel space equal to the distance between (a) and (b) in the StyleGAN latent space. A human can perceive that the images in (a) and (c) are of the same person. However, although (a) and (b) are of different persons, the Euclidean distance between (a) and (b) is equal to that between (a) and (c); the high-frequency components of the facial image insignificantly affect human recognition. The Euclidean distance between facial images is different from human perception. Moreover, it is desirable to use a space that can obtain the distance recognizable by humans. The second problem occurs while averaging data. In k-anonymization, we must unify the values of quasi-identifiers. One of the unification methods is to take an average value. This method is also used to ensure other anonymity, as averaged data are not sampled using actual individuals. In our proposed method, averaging is also used as an operation for unifying k-anonymization; however, a problem occurs while taking the average value in the case of multidimensional data such as facial images. For example, considering the example in Figure 1, when a group is created that straddles the Swiss roll layers, the average value is the value between the layers. Conversely, virtual points are created that do not exist in the Swiss roll. These points result in an unexpected dataset loss. Let us consider the example of the facial image depicted in Figure 2(d), which represents the average value of the images shownd in (a) and (b) in the pixel space; clearly, (d) cannot exist as a natural person. In Swiss roll, the second problem can be avoided by taking the average value in the space shown in Figure 1 (b). Similarly, in facial images, the second problem may be avoided by taking the average value in the latent space obtained using NN. Furthermore, Figure 2 (e) is an image obtained by averaging the facial images of (a) and (b) in the latent space of StyleGAN and re-mapping the image to the facial image. This facial image looks like a more average face compared with (d). When averaging facial images, it is desirable to use a space that can obtain an intermediate facial image recognizable by humans.

2.3 Natural face image anonymization

Several studies were conducted to anonymize facial images using NNs. A face-image anonymization method that generates natural images is briefly explained as a baseline in [13]. This method comprises the following three NNs: NN_M , which modifies the face; NN_D, which identifies an individual; and NN_A, which identifies the actions of the individual. This method can be used to anonymize facial images without compromising the image quality, as NN_M creates modified images that can be correctly detected using NN_A but cannot be detected using NN_D. However, the authors of the paper concluded that data are anonymized when their NN_D cannot identify the target individuals from the anonymized images. However, this does not comply with any definition of anonymity. Therefore, the resulting data obtained using the method may not be suitable for secondary use. To guarantee anonymity, we must apply the logic of the conventional anonymization method to face-image anonymization.

2.4 StyleGAN

StyleGAN [5] is an NN architecture proposed by Tero Karras et al. It generates high-resolution (1024 × 1024) facial images. It has two latent spaces, Z and W. Z is generated from a normal distribution, and W is obtained by mapping Z through a mapping network. In this study, we use StyleGAN to anonymize the facial images. StyleGAN obtains a facial image by mapping W via a synthesis network. In a conventional GAN, facial images are often directly obtained from Z, which is constrained by a normal distribution. This is because the facial images of a conventional GAN are entanglement in the Z space. In StyleGAN, this problem is solved by obtaining the latent space W by using a mapping network from Z. In [5], it was experimentally confirmed that the latent space W is a disentanglement. Therefore, it is easier to linearly obtain independent elements, such as facial expression and gender, using the latent space W of StyleGAN than using the latent space of a conventional GAN. Consequently, facial images with intermediate facial expressions and of different genders can be obtained via linear interpolation in the latent space W.

3 Proposed method

Table 3: Which function can G or S map use?

	LATENT	ATTRIBUTE	DIRECT
G map	YES	YES	YES
S map	YES	NO	YES

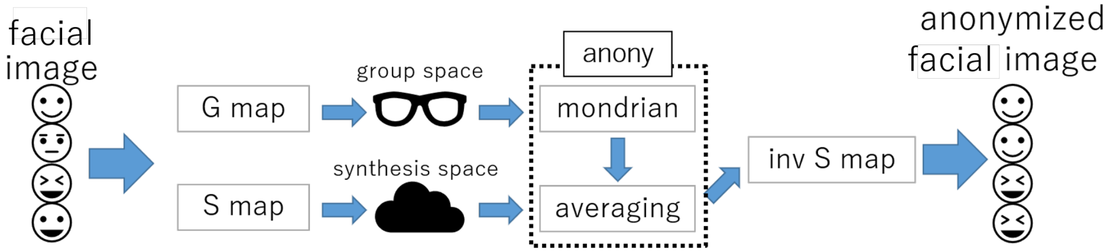


Figure 3: Architecture of

In this study, we propose a new anonymization method named multidimensional inputs k-anonymizing Unit (MIKU). The method comprises four modules, namely, G map (Group map), S map (Synthesis map), anony, and inv S map, as shown in Figure 3. MIKU maps data to another space and anonymizes all dimensions of the mapped data. First, the G map outputs the group space that is used for clustering, and the S map outputs the synthesis space that is used for generating anonymized facial images. MIKU uses NNs, such as StyleGAN, for the G and S maps. Table 3 shows the functions that the S and G maps can use. In [17], both the G and S maps used the same function. G and S maps were prepared to solve the two problems described in Section 2.2. Second, “anony” is used to perform k-anonymization, which uses Mondrian for grouping and averages to unify the values of the groups. Finally, the inv S map creates an anonymized facial image from the average in the synthesis space. According to the definition of k-anonymity used in our proposed method, the latent vectors of k face images of different individuals are used, and the latent vectors' average is calculated. In [18], the suppression method for k-anonymity was discussed, and it mentions the suppression method can consist of either blanking a value or replacing it with a locally neutral value, such as an average value. The proposed method follows the strategy of taking average because a generated anonymized face image using an average in the latent vector has all features of k concerning face images and gives less quality loss; namely, it gives better FID. The method of suppression is not limited to the proposed method and is applicable to other suppression methods.

3.1 G map

The G map outputs the group space where the distance between data can be measured using the Euclidean distance. In this study, we use an identity function (DIRECT), the mapping of StyleGAN to the latent space (LATENT), and the estimation of facial attributes (ATTRIBUTE). It is assumed that the distance between data can be measured using the Euclidean distance in group space. DIRECT outputs as group space values the pixel values of the facial images. This method is effective when the distance between data can be expressed using the difference in pixel values. LATENT outputs the facial image data in the latent space of StyleGAN. However, the latent space of StyleGAN includes noise components that do not depend on the face of an individual, for example, facial shadows and hair waves. Many noise components add noise on the Euclidean distance between data. Therefore, it may be difficult to use the latent space of StyleGAN as a space for finding effective distances between data. ATTRIBUTE outputs the labels of the facial images using NNs. It is not necessary to consider the existence of dimensions that can be the noise, as the labels are estimated on the basis of the attributes selected. However, it is generally difficult to list all the attributes that must be considered to satisfactorily express the distance between facial images. Therefore, in this study, we considered facial-image attributes, such as hair, mustache, gender, and ages, which seemed important.

3.2 S map

The S map outputs the synthesis space for making the anonymized facial image. Similar to the G map, the S map uses DIRECT and LATENT. However, ATTRIBUTE is not used. This is because the identity function for DIRECT and the Generator of StyleGAN for LATENT play the role of decoders (inv S map). Contrastingly, it is difficult to assume a decoder of ATTRIBUTE. When DIRECT is used for the S map, the anonymized facial images are the average pixel values of the facial images. Because the average pixel value cannot be a human facial image, the quality of the anonymized facial images is deteriorated. When LATENT is used for the S map, anonymized images are produced by obtaining the average value in the StyleGAN latent space. The linear interpolation in the StyleGAN latent space is a natural morphing [1]. LATENT needs a function to map facial images to the latent space of StyleGAN, W^+. However, this function is not involved in the architecture of StyleGAN. The basic idea of this function is discussed in [1]. The same value on the latent space W is used in the 18 layers inside the synthesis network of StyleGAN. However, it is difficult to calculate the value on latent space W from the facial image. Therefore W^+ is calculated instead of W. Notably, W^+ comprises 18 different 512-dimensional vectors, each of which corresponds to one layer. There are two ways to calculate the latent value of the target facial image. One method is to change the latent value from random latent value to desired latent value by using the gradient method. Another method is to learn the inverse synthesis network that can make latent values from facial images. As a result of experimenting both the methods, the former method can make latent values that can synthesize clearer facial images compared with the latter method.

3.3 Mapping to the latent space of StyleGAN (LATENT)

StyleGAN has a function that generates facial images using a mapping network and a synthesis network from Z that follows a normal distribution. However, no function exists to generate latent spaces W and Z from facial images. In [1], a method of mapping facial images to a latent space was discussed. The latent space W of StyleGAN is given as a 512-dimensional vector, which is used by 18 different layers in the synthesis network. As reported in [1], it is difficult to calculate W from facial images. Therefore, W^+ is calculated instead of W. Notably, W^+ comprises 18 different 512-dimensional vectors, which correspond to these 18 layers. The following two methods can be used to calculate W^+ from facial images: first, obtaining a function that performs inverse mapping and, second, gradually changing from an appropriately selected W^+ to obtain the optimal W^+, which can create the target face using a synthesis network. In [1], a method to gradually change W^+ was adopted. This means that W^+ is updated using the gradient method to obtain W^{+*} as follows:

$$W^{+*} = \underset{W^+}{\arg\min}\, Loss(I, G(W^+)) \tag{2}$$

The function G represents the synthesis network of a trained StyleGAN, and I denotes the target facial image of an inverse mapping. Loss is defined as the difference between I and $G\,(W^+)$. The sum of L_2 and $L_{percept}$ was used as the loss function, as shown in equation (3), with reference to [1]. One has the following:

$$Loss(I_1, I_2) = L_2(I_1, I_2) + 0.1L_{percept}(I_1, I_2) \tag{3}$$

where L_2 denotes a function that takes the sum of the L2 norms of the pixels between images I_1 and I_2. The term $L_{percept}$ denotes the sum of the L2 norms of the difference between the output values of the output values of conv1_1, conv1_2, conv3_2, conv4_2 of VGG-16 [14]. $L_{percept}$ was incorporated because it can easily learn facial image details such as wrinkles and hair.

3.4 Estimation of face image attributes (ATTRIBUTE)

In this study, we used ResNet-50[3] to estimate the attributes of facial images. Because there are 40 attributes in CelebA, the last fully connected layer of ResNet-50 was replaced with three fully connected layers with the output dimensions of 1024,1024,40. The first and second fully connected layers' activation layers were rectified linear units , and the final one was Sigmoid for the classification of 40 attributes. While learning this ResNet-50 using CelebA images and labels, three parameters were evaluated: (a) starting with a completely random initial value, (b) fixing convolution-layer parameters using ResNet-50 convolution parameters pre-trained on ImageNet, and (c) starting with

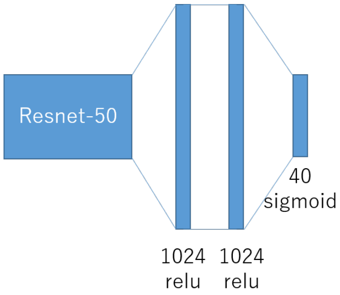


Figure 4: ATTRIBUTE architecture using RESNET-50

the ResNet-50 convolution parameters pre-trained on ImageNet. Binary cross entropy(BCE) was used as the loss function.

$$BCE = \frac{1}{40}\sum_{i=1}^{40}\left(y_i \log \tilde{y_i} + (1 - y_i)\log\left(1 - \tilde{y_i}\right)\right) \tag{4}$$

Here, y_i denotes a teacher value that becomes 1 when the input image has label i and 0 otherwise. $\tilde{y_i}$ denotes the estimated probability that the input facial image has label i. Momentum was used for the optimization as follows:

$$\begin{aligned} v_{t+1} &= \mu \times v_t + g_{t+1} \\ p_{t+1} &= p_t - lr \times v_{t+1} \end{aligned} \tag{5}$$

where v_{t+1} is the value inspired by the inertia term of the model of the physical system called moment. v_{t+1} is updated with the gradient value g_{t+1} of the loss function and $\mu \times v_t$. lr denotes the learning rate. In this study, the learning rate lr is 0.01, and the coefficient mu is 0.9. Figure 5 shows the trajectory of the loss value for each step during the experiment. The result of experiment (c) showed the least loss values for both training and testing. Therefore, the ResNet-50 model obtained through the initialization in (c) was used as ATTRIBUTE.

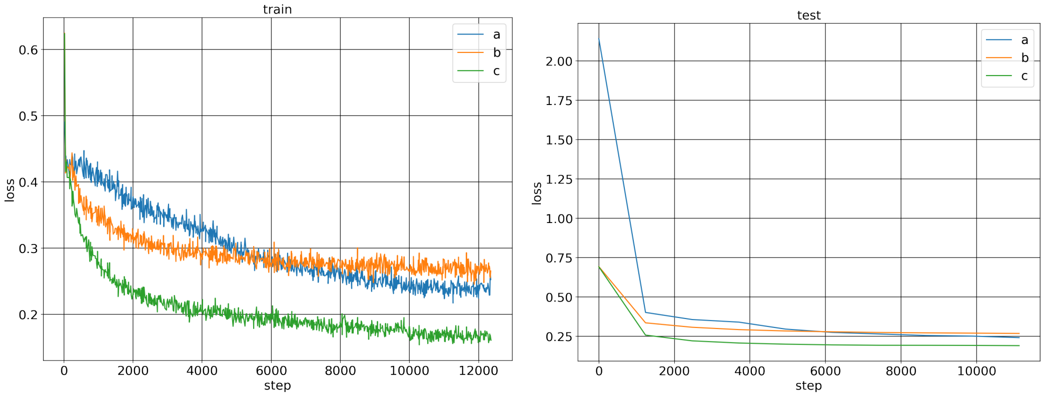


Figure 5: Loss value when training by CelebA

Table 4 shows the accuracy of ATTRIBUTE for each label in the test data. From the table, it is evident that most attributes show an accuracy of 80% or more. Furthermore, false positives are low even for labels with a low probability, such as Eyeglasses. There are labels with large false positives, such as Big_lips. However, such labels do not result in a problem unless they are used as the output of the G map. Considering the accuracy and the size of the impression given to the face, we selected 14 labels (Bald, Bangs, Black_Hair, Blond_Hair, Brown_Hair, Eyeglasses, Goatee, Gray_Hair, Heavy_Makeup, Male, No_Beard, Smiling, Wearing_Hat, and Young), and we estimated probability output using ResNet-50 as the G map.

4 Experiment

4.1 Preprocessing

Facial images often include not only faces but also background images. However, background information is not required to determine the latent space of StyleGAN. To remove the background of an image, we used the learned FCN-Resnet 101 [10] and masked it such that the background was not included in the learning of W^{+*} in (2). Subsequently, StyleGAN generated facial images in which the positions of the eye, nose, and mouth were fixed. Therefore, referring to [1, 6], the face was clipped into a rectangle, and the positions of the eyes and nose were fixed. Furthermore, the rectangle was resized to 1024 × 1024.

4.2 Experiment environment

A total of 202,599 images of CelebA were used for the evaluations. However, after preprocessing, only 197,016 facial images were used in the experiments, because the faces in the remaining images were not recognized. The parameters were $k = [2, 4, 8, 16, 32, 64, 128]$ for k-anonymization, and $N_s = 9216$ for Mondrian. The experiment used a machine with one NVIDIA V100 and another machine with four NVIDIA P100s.

4.3 Comparison method

The conventional method [13] does not perform anonymization with k-anonymity and, therefore, cannot be compared with the proposed method in terms of k -anonymity. Therefore, a method of directly anonymizing facial images was used as a comparison method. However, this is equivalent to the case in which both S and G maps are DIRECT. Therefore, when S and G maps are changed from DIRECT, the manner in which the results are affected is compared via experiments. In addition, in [17], MIKU used the same function for G and S maps and did not use ATTRIBUTE for the G map. Therefore, when G and S maps use LATENT, it is equivalent to MIKU in [17].

4.4 Qualitative evaluation

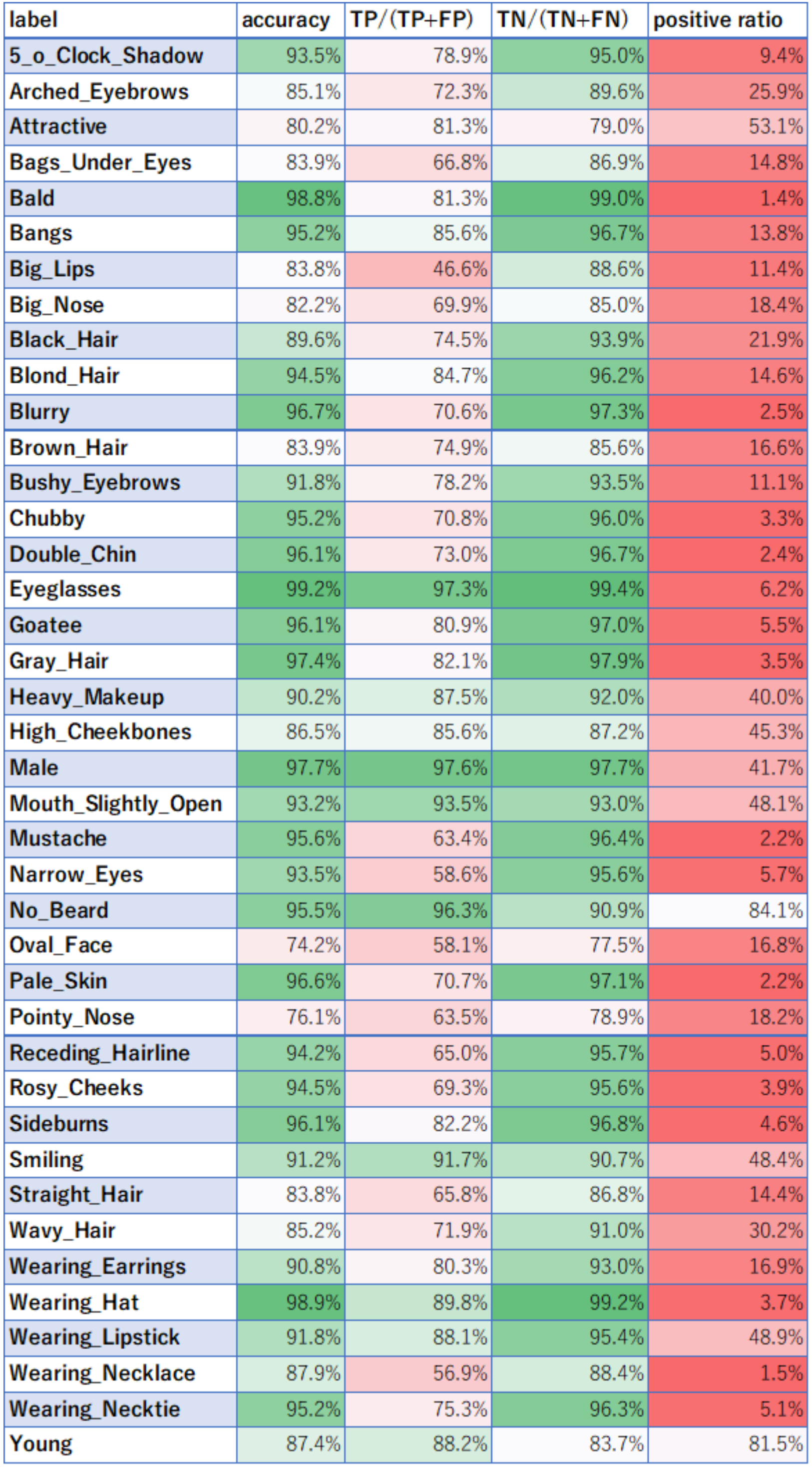
An anonymized facial image created using MIKU was randomly displayed to men and women one by one. A qualitative evaluation was performed on the anonymized facial images while changing the S map, G map, and k-anonymity.

4.5 Quantitative evaluation

4.5.1 Euclidean distance between labels before and after anonymization

CelebA has attributes given by humans. Using these attributes , the attribute vector (AV) of the facial image can be defined as a one-hot vector. Moreover, A certain group AV (gAV) can be defined as mean AVs that are in the group. The Euclidean distance between AV and gAV was used to quantitatively evaluate the G map. Here, the following three cases were experimented; [All]: using all the labels (40 types) of CelebA, [Male]: using only the label “Male,” and [Oval_face]: using only the label “Oval_face.” [All] was used for a comprehensive evaluation. [Male] was used for gender evaluation, which is important for human recognition. [Oval_face] is used for evaluating the label with lowest accuracy as presented in Table 4, as such label evaluation is disadvantageous to ATTRIBUTE.

Table 4: Label accuracy of ResNet-50 (c)



label	accuracy	TP/(TP+FP)	TN/(TN+FN)	positive ratio
5_o_Clock_Shadow	93.5%	78.9%	95.0%	9.4%
Arched_Eyebrows	85.1%	72.3%	89.6%	25.9%
Attractive	80.2%	81.3%	79.0%	53.1%
Bags_Under_Eyes	83.9%	66.8%	86.9%	14.8%
Bald	98.8%	81.3%	99.0%	1.4%
Bangs	95.2%	85.6%	96.7%	13.8%
Big_Lips	83.8%	46.6%	88.6%	11.4%
Big_Nose	82.2%	69.9%	85.0%	18.4%
Black_Hair	89.6%	74.5%	93.9%	21.9%
Blond_Hair	94.5%	84.7%	96.2%	14.6%
Blurry	96.7%	70.6%	97.3%	2.5%
Brown_Hair	83.9%	74.9%	85.6%	16.6%
Bushy_Eyebrows	91.8%	78.2%	93.5%	11.1%
Chubby	95.2%	70.8%	96.0%	3.3%
Double_Chin	96.1%	73.0%	96.7%	2.4%
Eyeglasses	99.2%	97.3%	99.4%	6.2%
Goatee	96.1%	80.9%	97.0%	5.5%
Gray_Hair	97.4%	82.1%	97.9%	3.5%
Heavy_Makeup	90.2%	87.5%	92.0%	40.0%
High_Cheekbones	86.5%	85.6%	87.2%	45.3%
Male	97.7%	97.6%	97.7%	41.7%
Mouth_Slightly_Open	93.2%	93.5%	93.0%	48.1%
Mustache	95.6%	63.4%	96.4%	2.2%
Narrow_Eyes	93.5%	58.6%	95.6%	5.7%
No_Beard	95.5%	96.3%	90.9%	84.1%
Oval_Face	74.2%	58.1%	77.5%	16.8%
Pale_Skin	96.6%	70.7%	97.1%	2.2%
Pointy_Nose	76.1%	63.5%	78.9%	18.2%
Receding_Hairline	94.2%	65.0%	95.7%	5.0%
Rosy_Cheeks	94.5%	69.3%	95.6%	3.9%
Sideburns	96.1%	82.2%	96.8%	4.6%
Smiling	91.2%	91.7%	90.7%	48.4%
Straight_Hair	83.8%	65.8%	86.8%	14.4%
Wavy_Hair	85.2%	71.9%	91.0%	30.2%
Wearing_Earrings	90.8%	80.3%	93.0%	16.9%
Wearing_Hat	98.9%	89.8%	99.2%	3.7%
Wearing_Lipstick	91.8%	88.1%	95.4%	48.9%
Wearing_Necklace	87.9%	56.9%	88.4%	1.5%
Wearing_Necktie	95.2%	75.3%	96.3%	5.1%
Young	87.4%	88.2%	83.7%	81.5%

4.6 Fréchet inception distance

The purpose of MIKU is to generate natural face images. When giving the same k-anonymity, different face images can be generated. It is better to provide an image that is more natural in actual use. Therefore, we need to evaluate these anonymized images on their quality. Here, we use the Fréchet inception distance (FID) [4], which is also used as a performance indicator for StyleGAN. FID is an index that evaluates the perceptual difference between two image sets using the output h (2048 dimensions) of the pool in the last hidden layer of the learned inception model [16]. It is calculated as follows:

$$\begin{aligned} \mu_{diff} &= |\mu_A - \mu_B|^2 \\ \Sigma_{diff} &= tr(\Sigma_A + \Sigma_B - 2(\Sigma_A \Sigma_B)^{1/2}) \\ FID &= \mu_{diff} + \Sigma_{diff} \end{aligned} \tag{6}$$

Here, the mean and variance-covariance matrix of h are μ_A, μ_B, Σ_A, and Σ_B, respectively. Typically, while measuring the FID, A or B represents a set of images used for learning. In this study, sets A and B refer to the sets of facial images before and after anonymization, respectively. Therefore, when the FID is low, B comprises images that show the same distribution as that shown by the images before anonymization. Additionally, it is clear that set B is a set of images of higher quality compared with set A.

5 Results

5.1 Qualitative evaluation

Figure 6 shows an anonymized image when the S and G maps of MIKU are changed. The top row shows the original images, and the next row shows the anonymized facial images with $k = 2, 4, 8, 16, 32, 64$, and 128. The left three columns use LATENT, and the right three use DIRECT for the S map. The G map repeatedly uses LATENT, DIRECT, and ATTRIBUTE from left to right. When the S map is DIRECT, the positions of eyes, nose, and mouth are natural, as they remain fixed upon preprocessing. However, unnatural edges are observed on the face when k is low, because the pixel values of the facial images are taken in the synthesis space. When k is high, although the eyes, nose, and mouth are clear, the outline is blurred and unnatural. When the S map is LATENT, unnatural edges are not observed on the anonymized facial images if k is low and high. However, some images show unnatural blue color jumps because of the normalizing layers of StyleGAN. When the G map is LATENT, the anonymized facial images are the same for both the woman and man for $k = 128$. This is caused by the effect of the noise described in Section 2.2. Because the value output by LATENT contains significant noise, the result of grouping achieved using LATENT is close to that of random grouping. Therefore, the anonymized facial images were the same irrespective of the original facial image. When the G map is DIRECT, a random grouping did not occur. DIRECT, which uses the difference in pixel values as the distance between images, performs grouping according to the facial image. Consequently, DIRECT can perform grouping on the basis of human attributes that affect the pixel value, such as on the basis o f the difference between a woman with long hair and a man and the difference in skin color. For example, when the S map is DIRECT, the face regions of the anonymized and original images were almost the same. This property is not observed in other G maps. When the G map is ATTRIBUTE, the gender of the anonymized facial images always matches that of the original facial images. When the S map is LATENT, the anonymized facial images are closer to the original ones compared with other G maps. From these results, it is inferred that ATTRIBUTE was able to express a space, thereby possessing human-like sensitivity. A quantitative discussion of this assumption is made in Section 5.2.1.

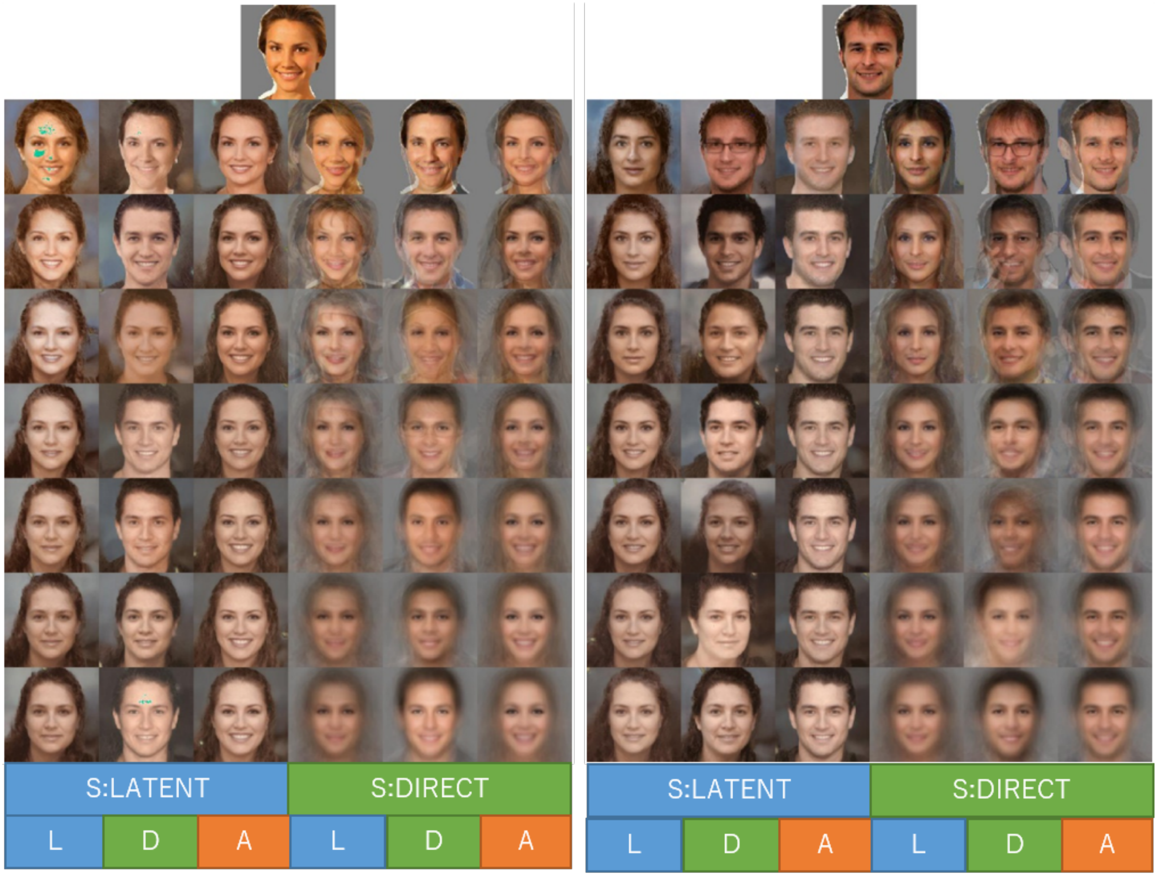


Figure 6: First row shows the original images, and the results are shown for $k = [2, 4, 8, 16, 32, 64, 128]$ in order from the second row. The three left most columns show the use of LATENT for the S map, and the three right most columns show the use of DIRECT. The G map repeatedly uses LATENT, DIRECT, and ATTRIBUTE from left to right.

5.2 Quantitative evaluation

5.2.1 Euclidean distance between labels before and after anonymization

The average Euclidean distance between the labels before and after anonymization is shown in Figures 7,8, and 9. In Figure 7, the Euclidean distance of ATTRIBUTE was least in three G maps. Grouping performed using ATTRIBUTE can reduce more semantic loss of facial images between before and after anonymization compared with using other G maps. The difference between DIRECT and ATTRIBUTE is greater than that between LATENT and DIRECT. This fact means that ATTRIBUTE can contribute to reducing the semantic loss of the anonymized faces. Notably, DIRECT has less loss than LATENT. Therefore, images anonymized using DIRECT as the G map are clearer than those anonymized using LATENT as the G map in Section 5.1. Figure 8 shows the result of the only Male, which is an attribute that significantly affects human perception. In this figure, ATTRIBUTE can reduce the gender ([Male]) loss of the facial image more strongly than in Figure 7. However, these results contain the attributes used by ATTRIBUTE. Therefore, it should be disadvantageous for other G maps. Figure 9 shows the result of the label Oval_Face, which is disadvantageous for ATTRIBUTE, as Oval_Face is the least accurate in Table 4 and is not included in the labels that are guessed using ATTRIBUTE. Furthermore, Table 4 shows the correlation between Oval_Face and other attributes in the CelebA data. From the table, it is evident that Oval_Face does not correlate with the 14 types of labels used by ATTRIBUTE. Interestingly, ATTRIBUTE shows the lowest loss in Figure 9. From these results, it is inferred that ATTRIBUTE

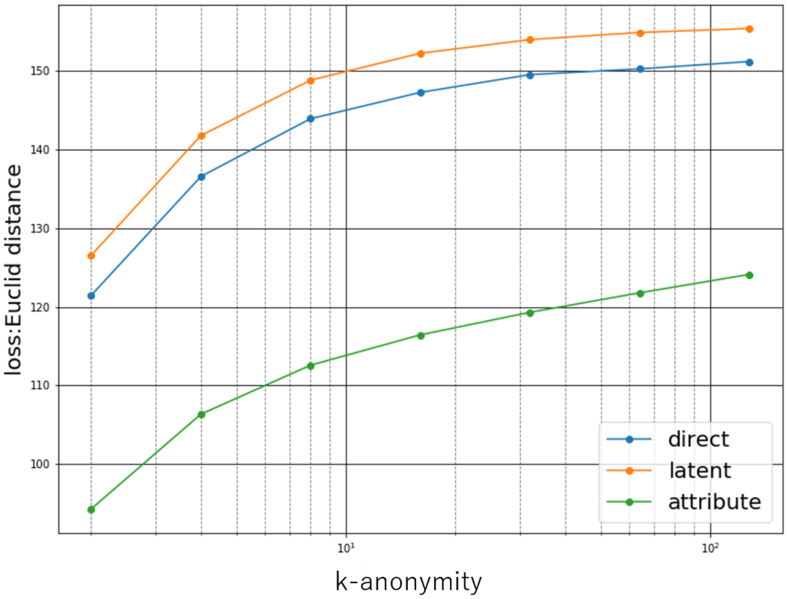


Figure 7: Euclidean distance between labels before and after anonymization [All]

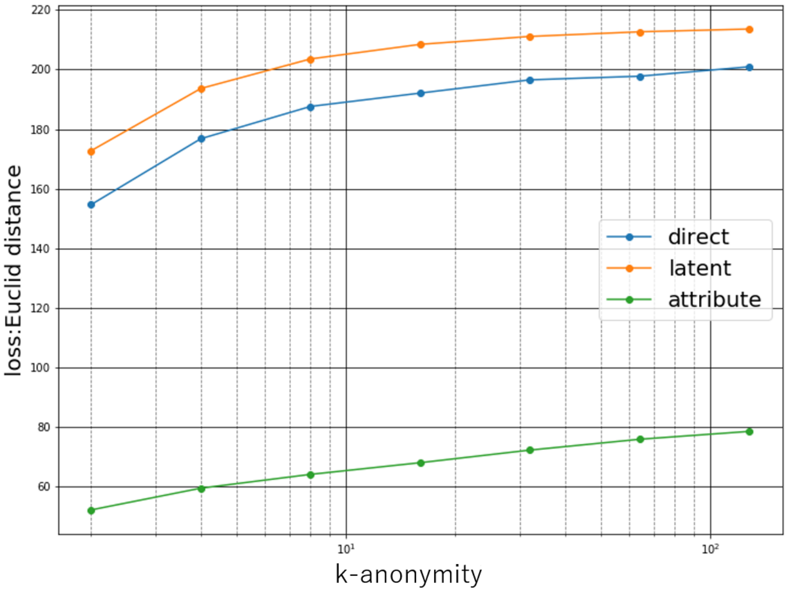


Figure 8: Euclidean distance between labels before and after anonymization [Male]

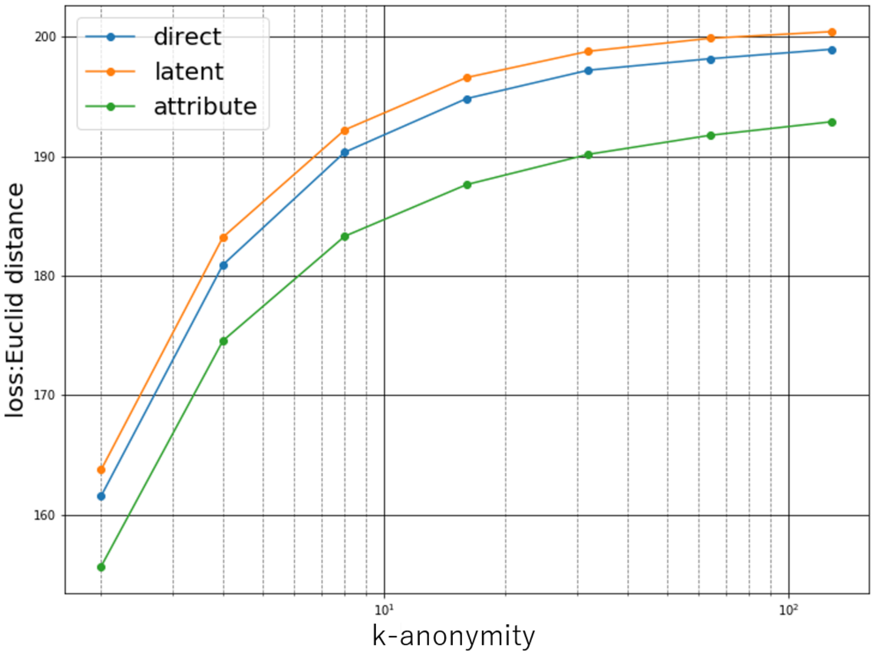


Figure 9: Euclidean distance between labels before and after anonymization [Oval_Face]

learns a space that is close to the space recognized by humans.

5.2.2 FID

Figure 10 shows the results of FID. In the figure, g and s represent G and S maps, respectively, and the string that follows indicates the one used for each map. From the figure, it is evident that using LATENT as the S map reduced FID. For the G map, ATTRIBUTE could lower FID the most when the S map was LATENT, and DIRECT could lower FID when the S map was DIRECT. Overall, using LATENT as the S map and ATTRIBUTE as the G map could reduce the FID value the most compared with other maps. However, when $k = 2$ and the S map was DIRECT, FID was smaller than when S map was LATENT. Even when $k = 2$ in Figure 6, a more naturally anonymized facial image could be generated when the S map was LATENT. Considering that FID is an index based on the hidden layer of the inception and does not always match human perception, we can conclude that it is more natural to use LATENT for the S map than DIRECT even when $k = 2$.

6 Conclusion

In this study, we proposed the MIKU method, which successfully anonymized multi-dimensional data such as facial images with minimal loss by performing k-anonymization in a space different from the space of the given data. To evaluate MIKU, CelebA was anonymized, and qualitative and quantitative evaluations were performed using FID. Consequently, when $k >= 2$, it was confirmed that the use of LATENT for the S map could generate a more natural anonymized image in the qualitative evaluation. For the G map, both ATTRIBUTE and DIRECT provided good results. However, the combination that exhibited the best FID was the use of LATENT for the S map and ATTRIBUTE for the G map. It can be said that the result of machine learning contributed to the anonymization.

Table 5: Correlation between Oval_Face and other labels that are used by ATTRIBUTE

label	correlation
Heavy_Makeup	0.2134
Smiling	0.2058
Male	0.1203
Young	0.1127
No_Beard	0.0618
Eyeglasses	0.0606
Gray_Hair	0.0587
Blond_Hair	0.0499
Brown_Hair	0.0462
Wearing_Hat	0.0462
Black_Hair	0.0319
Goatee	0.0212
Bald	0.0108
Bangs	0.0016

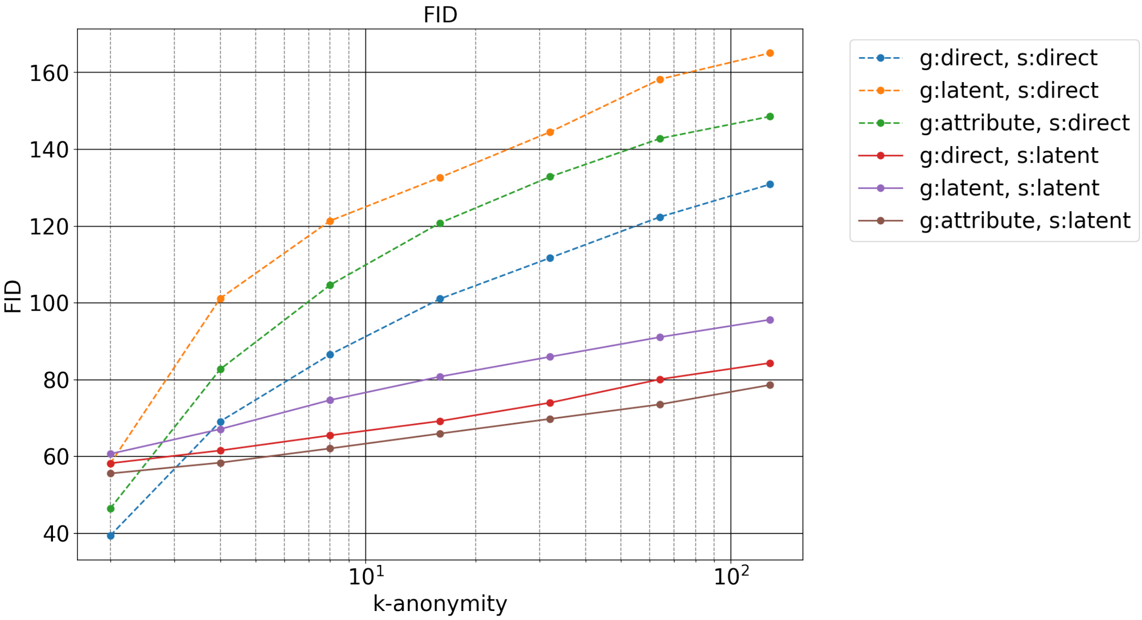


Figure 10: FID and k-anonymity (g:G map, s:S map)

7 Future works

In this study, we only focused on face-image anonymization, although our method can be applied to other high-dimensional data. For example, if an NN such as word2vec [12] is used as the S or G map, MIKU can be applied to natural language anonymization. Thus, MIKU contributed to solving the problems discussed in Section 2.2, and provided a methodology for general-purpose multidimensional anonymization solutions. Furthermore, the machine-learning architecture that will be studied in the future can be introduced into MIKU and applied as anonymization. We expect that the realization of higher-quality anonymization in domains other than facial images will be possible using the MIKU architecture.

8 Acknowledgment

This work was supported by JST CREST Grant Number JPMJCR19K1, and the commissioned research by National Institute of Information and Communications Technology (NICT, Grant Number 22004) , JAPAN.

References

[1] Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space? *arxiv*, Apr 2019.

[2] Charu C. Aggarwal. On k -anonymity and the curse of dimensionality. *Proceedings of the 31st VLDB Conference*, pages 901–909, 2005.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-Decem, pages 770–778. IEEE Computer Society, dec 2016.

[4] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 2017-Decem, pages 6627–6638, jun 2017.

[5] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[6] Vahid Kazemi and Josephine Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees. In *CVPR*, 2014.

[7] Batya Kenig and Tamir Tassa. A practical approximation algorithm for optimal k-anonymity. *Data Mining and Knowledge Discovery*, 25(1):134–168, jul 2012.

[8] Alex Krizhevsky, Ilya Sutskever, and Hinton Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pages 1–9, 2012.

[9] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, Apr 2015.

[11] Adam Meyerson and Ryan Williams. On the complexity of optimal K-anonymity. In *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '04*, page 223, New York, New York, USA, 2005. ACM Press.

[12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR, 2013.

[13] Zhongzheng Ren, Yong Jae Lee, and Michael S. Ryoo. Learning to Anonymize Faces for Privacy Preserving Action Detection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11205 LNCS:639–655, 2018.

[14] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arxiv*, Sep 2014.

[15] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, 10(5):557–570, oct 2002.

[16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2016-Decem, pages 2818–2826. IEEE Computer Society, dec 2016.

[17] Nakamura Taichi, Yuiko Sakuma, and Nishi Hiroaki. Face Image Anonymization as an Application of Multidimensional Data K-Anonymization. In *Seventh International Symposium on Computing and Networking Workshops (CANDARW)*, 2019.

[18] Vicenç Torra and Domingo-Ferrer Josep. Ordinal, Continuous and Heterogeneous k-Anonymity Through Microaggregation. In *Data Mining and Knowledge Discovery*, 2005.