多次元データ K 匿名化としての顔画像匿名化

中村 太一[1]　西 宏章[2]

概要：近年のデータ通信網の発展や計算機処理能力の進歩によって日々扱うデータの量は飛躍的に増大している．このデータを分析し，利活用することで将来的なイノベーションが起こることが期待されるが，同時にプライバシの侵害も問題視されるようになった．例えば，SNS 上では，画像の公開がそこに写る他人の肖像権の侵害へつながりかねない．従来から個人情報を保護しつつデータ公開するために，匿名化の研究が行われてきたが，画像のような高次元データの匿名化は従来の研究では次元間の複雑な関係を考慮した匿名化を行っていなかった．そのため，匿名化によるデータの意味的な損失が大きくなる傾向にあった．本論文は，高次元なデータの 1 つである顔画像を対象として匿名化を行う．顔情報が持つ高次元情報における次元間情報を用いて高精細な匿名化画像を得るため、提案手法 MIKU は StyleGAN の潜在空間を利用して匿名化する．これによって，提案手法は直接画像を匿名化する従来手法に比べて，より画像の質を保ったまま匿名化する事ができる．このことを FID による定量的な結果と出力された画像の定性的な評価を従来手法と比べることによって確かめた．FID の結果は K 匿名性が大きな匿名度を持つとき MIKU の方がより良い性能を持つことを示した．K 匿名性が小さな匿名度を持つときも FID の値が従来手法の作る匿名化画像を正しく評価できていないことを Inception モデルの線形性の影響から示した．加えて，定性的な評価では MIKU の作る匿名化画像の方が輪郭や髪の表現などが従来手法に比べてより自然に生成できていることを確かめた．

キーワード：K 匿名化，顔画像，高次元，StyleGAN，ニューラルネットワーク

Face Image Anonymization as an Application of Multidimensional Data K-Anonymizer

Taichi Nakamura [1]　Hiroaki Nishi [2]

Abstract: In recent years, the development of data communication networks and advances in computer processing capacity increase the amount of data to be handled dramatically. Though the future innovations using the data are expected, privacy violation from the data has become a problem. For example, image disclosure in SNS may lead to infringement of portrait rights. In the past, researches on anonymization have been conducted in order to disclose data for protecting personal information. However, conventional anonymization, even for high-dimensional data such as face images, uses an averaging operation in anonymization and has not considered complex relationships between dimensions. Therefore, the loss of semantic meaning increases because the loss is caused by the anonymization process assuming Euclidean data space.　In this paper, we proposed MIKU, an anonymization algorithm focusing on face image anonymization as a typical example of high dimensional data. MIKU enables anonymization that retains the quality of the images. Since the conventional method directly anonymizes images, it causes the quality deterioration in images. MIKU considers the relation between dimensions by using the latent space of StyleGAN. The effect of using the latent space was confirmed by comparing the quantitative results of FID and the qualitative evaluation of the output image with the conventional method. The quantitative results by FID proved that MIKU achieved better performance when K-anonymity is large. When K-anonymity is small, we clarified that the image quality of the conventional method such as FID could not be measured correctly and overestimated because of the effect of the linearity of the inception model. In addition, qualitative evaluation of anonymized images shows MIKU generates a more natural image in the contour and hair expressions compared with the conventional method, and the anonymized images of MIKU have no unnatural edge lines on a face which is generated in the conventional method.

Keywords: K-Anonymization, Face Image, High-Dimension, StyleGAN, Neural Network

1. 背景

計算機の計算処理能力は日々進歩しており，単位時間内に処理可能なデータ量も増大している.これによってより多くのデータの活用が可能となり, さらなる 経済発展やイノベーションなどが期待されている.利活用の対象となり得るデータは，政府機関などの団体や組織で扱われるデータから IoT (Internet of Things)デバイスのセンサ情報まで幅広く，かつ膨大である.例えば，顔画像は監視カメラやスマートフォンのカメラなど，様々なデバイスから得ることができる．しかし，この顔画像を被写体の許可なく公開することにはプライバシの問題がある．現在は，この問題に対処するために，顔をぼかす，黒塗りするなどの加工が一般

1 慶應義塾大学大学院
Graduate School of Science and Technology Keio University
taichi@west.sd.keio.ac.jp
2 慶應義塾大学大学院
Graduate School of Science and Technology Keio University
west@sd.keio.ac.jp

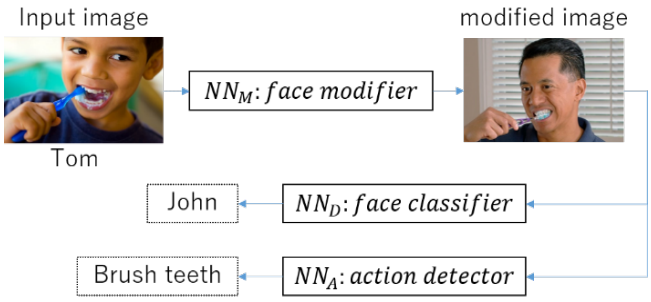

図 1　従来手法[1]の構造図

Figure 1　The architecture of the conventional face anonymization method[1]

に行われているが，これらの手法により得られる画像は自然ではなく，見た目上の画像の質の劣化が大きい．

一方で，プライバシの問題を考慮したデータ流通をおこなうために，匿名化技術の研究が行われている．匿名化技術の多くは，個人間の識別をできなくするようにデータを変更することによって，個人のプライバシを保護する．しかし，一般に顔画像データのような大きな次元数を持つデータに対して従来の匿名化技術は有効ではない．

また，昨今はニューラルネットワーク（以降 NN）に代表される機械学習器の発展が著しい．これらの技術は匿名化とは異なる技術に見えるが，匿名化と機械学習には，どちらもデータセットから，統計的な情報を抽出するという機能に共通点がある．特に NN には，隠れ層があり，ここからは NN の持つ入力と出力の間の中間的な情報を得ることが期待できる．

そこで，本論文では，顔画像をその質を保ちながら匿名化することを目的とし，顔画像を NN の 1 つである StyleGAN の潜在空間に写像し，その空間内で匿名化を行うことで，多次元データを匿名化する際の問題への対処を試みる．

2. 関連研究

2.1 顔画像匿名化

本論文では，顔画像の匿名化を行うが，例えば[1]は同様の問題に対する解決策を示している．この手法は，図 1 に示すように，顔を加工するNN_Mと，個人を識別するNN_D，個人の行動を識別するNN_Aにより構成されている．NN_Mに，NN_Dの出力が誤り，かつ，NN_Aは正しく識別できるような加工を学習させることで，顔画像の劣化なしに加工できる．しかし，匿名性の根拠は個人を識別するNN_Aであり，これはこれまでの匿名化技術研究の文脈の匿名性とは異なることは明らかである．実際，出力される画像の中にはほとんど変化が見られないような顔画像も存在し，顔画像匿名化の手法としては問題がある．そのため，匿名性には従来の匿名化手法を応用する必要がある．

表 1　生データの例

Table 1 Row example of a 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

表 2　K 匿名化後のデータの例

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

ID	Zip-code	Age	Gender	Disease
$t1$	0***	24	*	Cancer
$t2$	0***	24	*	Flu
$t3$	0***	24	*	Aids
$t4$	122*	35	Male	Cold
$t5$	122*	25	Male	Flu

2.2 K-匿名性

K 匿名性[2]は典型的な匿名性の 1 つである．まずこの匿名性の説明をする前に，いくつか言葉を定義する．

まず，属性のうち，氏名やユーザに固有な ID など個人をそのまま一意に特定できる符号を「識別子（Identifier）」と呼ぶ．これは，匿名化においては事前に削除されることが多い．本研究では識別子が既に削除されたデータベースを扱う．次に，単体での特定は困難であるが，複数組み合わせることで個人の特定が可能になる属性を「準識別子（Quasi-Identifier）」と呼ぶ．最後に，識別子でない属性のうち，データ解析にあたり，重要なため匿名化を施したくない属性を「機密情報（Sensitive Attribute）」と呼ぶ．

K 匿名性とは任意の個人について最低でもK－1個以上の同一の準識別子を持つデータがあることを保証する匿名性である．K 匿名性を得るために行う加工のことを K 匿名化と呼ぶ．例えば，表 1 のようなデータがあり，これをK＝2で匿名化したデータが表 2 である．ID は識別子であり，匿名化される際は削除される．ただし，この表では匿名化前後の比較のため便宜上残した．Zip-code, Age, Gender を準識別子，Disease を機密情報として置き，Zip-code, Age, Gender を加工することで，Disease の情報が保護されるようにする．表 2 を見ると，(t1, t2, t3)と(t4, t5)がそれぞれグループになり，同じ準識別子を共有しており，準識別子から個人を特定できないことが分かる．K 匿名化を行うには，まず，準識別子の近い個人同士を必ず K 人以上が含まれるようにグループに分け，次に，そのグループ間で準識別子を統一化する操作を行う．ここで統一化は同じグループに分けられた準識別子の集合から計算可能な値で準識別子を置き換える操作に相当する．例えば，Zip-code のように上位の桁だけを残すマスクや，Age のように値の平均で置き換える方法，Gender のようにデータ自体を消す消去が考え

られる．

K 匿名化を行うにあたり，統一化前後のデータで情報の劣化が最小になるようにグループ化を行うことが重要になるが，最適なグループ化は NP 困難であることが既に指摘されている[3]．そこで，最適な方法でないヒューリスティックな方法も提案されている[4]．しかし，本論文で扱う顔画像データはベクトルであり，統一化には平均化を用いるため，階層的なデータ構造を必要とすることからそのまま用いることは難しい．また，高精細な顔画像に用いると扱うベクトルが比較的高次元であるため計算時間が長くなりがちである．そこで，本論文では，グループ化として Mondrian と呼ばれる手法用いる．Mondrian の実装は様々あるが，本論文での実装は次の Algorithm 1 のようにした．

Algorithm 1 Mondrian

個人iのデータ$\mathrm{v_i} \in \mathbb{R}^d, i \in \mathbb{N}$と探索する次元の数$\mathrm{N_s}(\mathrm{N_s} \leq \mathrm{d}, \mathrm{N_s} \in \mathbb{N})$，匿名性Kが与えられ，個人の集合 V の初期値は$\mathrm{V} = \{1,2,3,\ldots\}$であるとする．ただし，$v_{\mathrm{i,j}}$は個人 i のデータの j 次元目の値を示すとする．また，このアルゴリズムのグループ化の結果であるグループの集合をGとする．

Step 1 V に含まれる個人数が2K未満である場合，VをGに追加して試行を終了する．

Step 2 1 からdまでの整数から重複せず,ランダムに$\mathrm{N_s}$個の整数を選び，これを集合 A とする．

Step 3 A に含まれる次元 j のうち

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

となるj*次元目についてVをソートする．

Step 4 ソートされた V の前半の半分を$\mathrm{V_a}$，後半を$\mathrm{V_b}$として分割する．

Step 5 $\mathrm{V_a}$および$\mathrm{V_b}$をVと置きなおして，Step 1 を再帰的に繰り返す．

Algorithm 1 は再帰的なアルゴリズムであり，Step1 でそれぞれのグループ化結果を出力する．Step1 では，集合Vの大きさが2K以上である場合，K個以上の要素を持つ 2 つの集合に分割できるので以降の Step2~5 へ進む．それ以外の場合は，これ以上分割すると K 匿名性を満たすグループが作成できないため，集合Vをグループの 1 つとして出力する．Step2 では，d以下の正整数をランダムに$\mathrm{N_s}$個重複なく選ぶ．Step3 では，ランダムに選んだ$\mathrm{N_s}$個の次元について，最大値と最小値の差が最も大きい次元j*を計算する．Step4 では，その次元J*についてVをソートし，Step5 で，Vを 2 等分に分割し，分割された集合に対して，それぞれ Step1 を再帰的に行う．

この Algorithm 1 は高次元なデータでも安定した速度でグループ化を行うために探索する次元数$\mathrm{N_s}$を指定できる．このおかげで，本論文で扱うような顔画像（$3 \times 1024 \times 1024$次元）でもデータの次元数ではなく，$\mathrm{N_s}$に依存した計算時間でグループ化ができる．

2.2.1 高次元ベクトルの匿名化の問題

本論文以前から，一般に高次元なデータの K 匿名化は困難とされている[5]．これは高次元なデータには大きく分けて 2 つの問題があるためであると考えられる．1 つ目は，実際の解析には必要のない情報の増大であり，2 つ目は，非線形的な相関を持つ情報の増大である．

1 つ目の問題は，例えば，顔画像で言えば，背景の画素の情報などが相当する．背景の情報は、顔画像の解析とはほとんど関係ないはずであるが匿名化が前処理なしで行われる場合匿名化は背景と顔を等しく扱い顔画像のグループは背景にも依存する．つまり，このグループ化は分析に必要のない情報に依存するため，事前に無関係な要素を排除した場合と比べ，出力される匿名化された情報の劣化は大きくなる．そこで，本論文では提案手法と比較に用いる従来手法ともに前処理として背景画像の排除を行っている．

2 つ目の問題が匿名化に悪い影響を与えるのは，本論文で扱う Mondrian を含め多くのグループ化の手法が暗にデータの空間にユークリッド空間を想定していることに起因している．実際の空間はユークリッド距離を用いてデータ間の意味的な距離を測れるとは限らない．データ数が十分にあるとき，ある 1 つのデータの微小な周辺についてはユークリッド空間として近似できる可能性はある．しかし，高次元なデータを扱うため，同時にデータの密度も小さくなり，微小なユークリッド空間が連結しているという仮定を置くことも難しい．そこで，提案手法では，事前に顔画像の空間について学習している StyleGAN の潜在空間を利用することで，非ユークリッド空間に属するデータの匿名化の問題への対処を試みている．

2.3 StyleGAN

StyleGAN [6]は Tero Karras らによって提唱された高精細な1024×1024のサイズの顔画像を生成する NN のアーキテクチャである．StyleGAN には，潜在空間が 2 つある．1 つ目は，正規分布から生成されるZであり，2 つ目は，ZをMapping network によって，写像することによって得られるWである．StyleGAN は W を Synthesis network によって，さらに写像することで顔画像を得る．

従来の GAN では Z から直接顔画像を写像してえることが多かった，Z が正規分布に従うため，Z 空間内での要素のエンタングルメントが防げなかった．そこで，StyleGAN では，潜在空間 W を Mapping network による Z から写像で得ることで，その問題を解決できると考えられる．論文[6]

中では, 潜在空間 W の方が独立した要素を持っていることが実験的に確かめられている．つまり，従来の GAN の潜在区間を用いるのに比べ，StyleGAN の潜在区間 W を利用した方が，表情や性別と言った独立した要素が線形に得やすい．これにより，複数の顔画像の表情や性別の表現が中間的な顔画像を W 内での線形な補間によって得ることができる．

本論文では，すでに学習済みの StyleGAN を使い，StyleGAN の潜在空間 W に線形に独立した成分を持っているという性質に注目し，これを匿名化に応用する．この詳しい説明は 3.1 節で述べる．

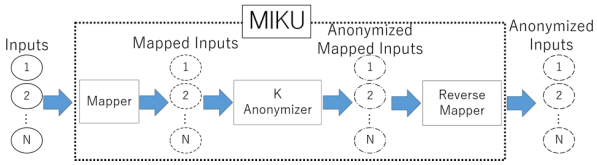

図 2　Multidimensional Inputs K-anonymizing Unit(MIKU)の構造図

Figure 2　Structure of Multidimensional Inputs K-anonymizing Unit

3. 提案手法

3.1 Multidimensional Inputs K-anonymizing Unit (MIKU)

本論文では, 新しい K 匿名化手法として Multidimensional Inputs K-anonymizing Unit（以降 MIKU）を提案する．MIKU は図 2 のように，Mapper，K-Anonymizer，Reverse Mapper の大きく分けて 3 つのモジュールによって成立する. まず，Mapper では，MIKU が K 匿名化するデータの写像を行う．この写像は，データが次元間で複雑な相関を持っている空間から独立した成分で表される空間へ移ることを期待して行う．例えば，顔画像では，ピクセルの RGB 値を表す次元から，性別，表情，年齢といった独立した成分によってなる空間へ写像されると予想される. 次に, K-Anonymizer は，従来の匿名化と同様な手法によって写像されたデータを K 匿名化する．2.2.1 で述べた通り，従来の K 匿名化手法は高次元のデータに対してうまく働きづらいが，写像されたデータは独立した成分による空間にあるため，写像せずに行う場合に比べ良い匿名化ができることが期待できる．最後に，Reverse Mapper は写像された空間で匿名化されたデータをもとの空間に戻す．このモジュールは，元のデータ形式がデータを公開するのに都合の良い場合のために存在する．そのため，データの形式を問わない匿名化を行う場合は省くことができる.

MIKU は，比較的に研究の進んだ機械学習の研究成果を匿名化に取り入れることができる手法であり，機械学習の分野におけるカーネル法に似た方法である.

本論文では，Mapper に学習済み StyleGAN の潜在空間 W を拡張したW^+へ写像を行う．W^+の解説は 3.2 節で行う．K-Anonymizer では Mondrian でグループ化を行い平均値によって値を統一化する．Reverse Mapper では，学習済み StyleGAN の Synthesis network を利用する.

3.2 StyleGAN の潜在空間への写像

StyleGAN は正規分布に従う Z から Mapping network, Synthesis network を使って顔画像を生成する機能はあるが，顔画像から潜在空間 W や Z を生成する機能はない．[7]では，顔画像から潜在空間へ写像する方法について議論している．StyleGAN の潜在空間 W は 512 次元のベクトルとして与えられ，Synthesis network 内部では 18 個のレイヤで同一の値 W が使用される．[7]では，顔画像から W を計算することが困難であるため，W の代わりに W を拡張したW^+を計算している．W^+は 18 個の異なる 512 次元のベクトルで構成され，それぞれが 1 つのレイヤに対応している．顔画像からW^+を計算するには，逆写像を行う関数を発見する方法と適当に選んだW^+から徐々に目的の顔画像を出力できるように変化させていく方法の 2 つが考えられるが，[7]では後者を採用した. これは(1)式となるようなW^{+*}を発見するために，勾配法によってW^+を更新していくことを表す．ここで，関数 G は学習済み StyleGAN の Synthesis network を表し, I は逆写像の対象となる顔画像を表す. Loss は I と$G(W^+)$との差を求める関数であり, 4.2 節で詳説する.

$$W^{+*} = \underset{W^+}{\operatorname{argmin}}\, Loss\left(I, G(W^+)\right) \tag{1}$$

4. 実験

4.1 前処理

証明写真と異なり，多くの場合顔画像は顔だけでなく背景の画像も含む場合がほとんどである．しかし，StyleGAN の潜在空間を求めるにあたって, 背景の情報は不要である．そこで，まず[8]を参考にして，顔の位置を特定し，目，鼻口の位置を固定した上で正方形にクリッピングし，1024×1024の画像として画像のサイズを固定した．そして，画像から背景を削除するために，学習済みの FCN-Resnet101[9]を使って, (1)の学習に背景が含まれないようにマスクした.

4.2 Loss 関数

Loss 関数には[7]を参考にして(2)式のように，L_2と$L_{percept}$の和を用いた.

$$\mathrm{Loss}(I_1, I_2) = L_2(I_1, I_2) + L_{percept}(I_1, I_2) \tag{2}$$

ここで，L_2は画像間の画素の差の L2 ノルムの総和を取る関数であり，$L_{percept}$はI_1,I_2それぞれを入力に取ったときの VGG-16[10]の conv1_1, conv1_2, conv3_2, conv4_2 の出力の差の L2 ノルムの総和である．L_2のみを Loss に使ったときに比べて，$L_{percept}$を加えた方がしわ等の細かな顔画像の表現が学習しやすいため，$L_{percept}$を加えている.

4.3 Fréchet Inception Distance (FID)

提案手法の性能を評価するために，顔画像を直接匿名化した場合の方法と比べて出力される匿名化画像を評価する．このとき，評価には画像の質を用いたいので，StyleGAN の性能指標としても使われる Fréchet Inception Distance(以降 FID)[11]を用いる．FID は，学習済みの Inception モデル[12]の最後の隠れ層の pool の出力 h(2048 次元)を使って，生成された画像と真の画像との間で h の分布の差を示す指標である．画像の集合 A と B について，それぞれ h の値を求め，その平均と分散共分散行列を$\mu_A, \mu_B, \Sigma_A, \Sigma_B$としたとき，FID は以下の(3)式のように求まる．

$$\mu_{diff} = |\mu_A - \mu_B|^2$$
$$\Sigma_{diff} = tr\left(\Sigma_A + \Sigma_B - 2(\Sigma_A\Sigma_B)^{\frac{1}{2}}\right)$$
$$\mathrm{FID} = \mu_{\mathrm{diff}} + \Sigma_{diff} \tag{3}$$

通常，FID を測る際は A または B は学習に使われる画像の集合を表し，本論文でも集合 A は匿名化前の顔画像を示す．すなわち，FID が低いとき，B は匿名化前の画像と同じ分布を示す画像で構成されており，B がより質の高い画像の集合であることが分かる．FID の計算は[13]を参考にした．

4.4 実験環境

実験に用いたデータは，LFW[13]の個人 5749 人に対してそれぞれ 1 人につき 1 枚の顔画像を対象とした．ただし，前処理の結果，顔として認識されない顔画像が存在し，実際にこの実験で用いた顔画像は 5722 人の顔画像である．

実験には NVIDIA V100 が 1 枚搭載されたマシンと NVIDIA P100 が 4 枚搭載されたマシンを用いた．

実験は前処理を行った画像に MIKU による匿名化を行った場合と従来手法による場合の 2 つについて行う．従来手法には，直接顔画像を Mondrian によってグループ化し，平均値によって統一化を行う手法を採用した．ここで，用いたパラメータは，$K = [2,3,6,8,16,32,64,128]$ であり，Mondrian における探索する次元数$a = 9216$とした．

表 3 DIRECT と MIKU のそれぞれの K 匿名度での FID 値
MIKU は提案手法，DIRECT は従来手法を表す
Table 3 FID for DIRECT method and MIKU. MIKU is the proposed method, Direct is the conventional method. The leftmost column shows the level of K anonymity.

K	DIRECT	MIKU
2	61.17	111.8
4	128.8	127.6
8	178.0	146.1
16	190.6	162.8
32	200.5	176.1
64	227.9	191.7
128	245.1	202.6

5. 結果

5.1 FID

FID を計測した結果は表 3 のようになった．表 3 の DIRECT は従来手法を表し，MIKU は本論文の提案手法である．これを見ると，$K = 2$のときを除けば MIKU の方が従来手法に比べて FID が低く，$K = 2$のときのみ従来手法が MIKU に比べて低い FID を取っていると分かる．

しかし，$K = 2$のときも，5.2 節で述べるように定性的には提案手法が良い性能を示している．そこで，なぜ定性的な評価と FID による評価に差があるのかについて議論する．

仮に Inception モデルの h を出力する関数f(.)が線形である場合を考えると，(4)式に示すように，従来手法によって匿名化された画像の h の空間上での平均値は，関数 f の線形性から匿名化前の顔画像の h の空間上での平均値に一致する．

$$\frac{1}{N}\sum_{q\in Q}|q|f\left(\frac{1}{|q|}\sum_{i\in q}I_i\right) = \frac{1}{N}\sum_{q\in Q}\sum_{i\in q}f(I_i)$$
$$= \frac{1}{N}\sum_{i}^{N}f(I_i) \tag{4}$$

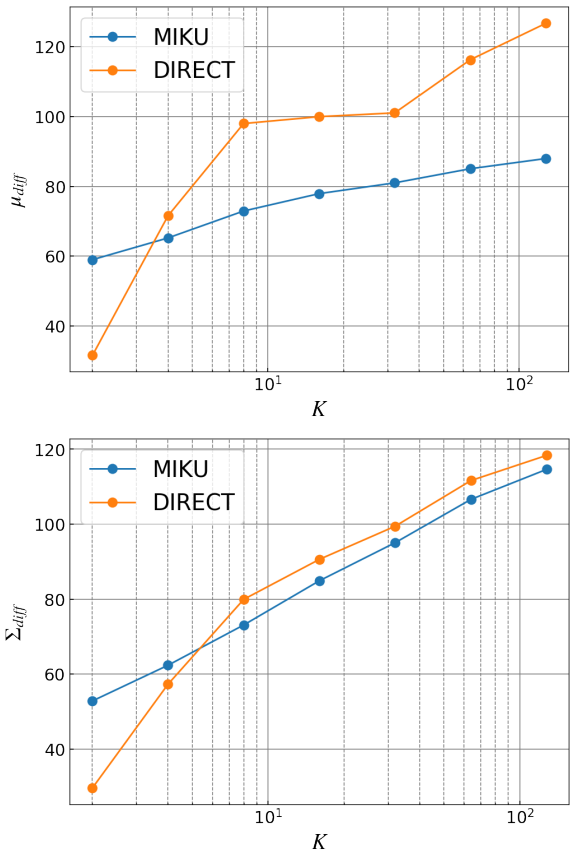

図 3 K に対するμ_{diff}とΣ_{diff}の変化
Figure 3 How μ_{diff} and Σ_{diff} grow relative to K

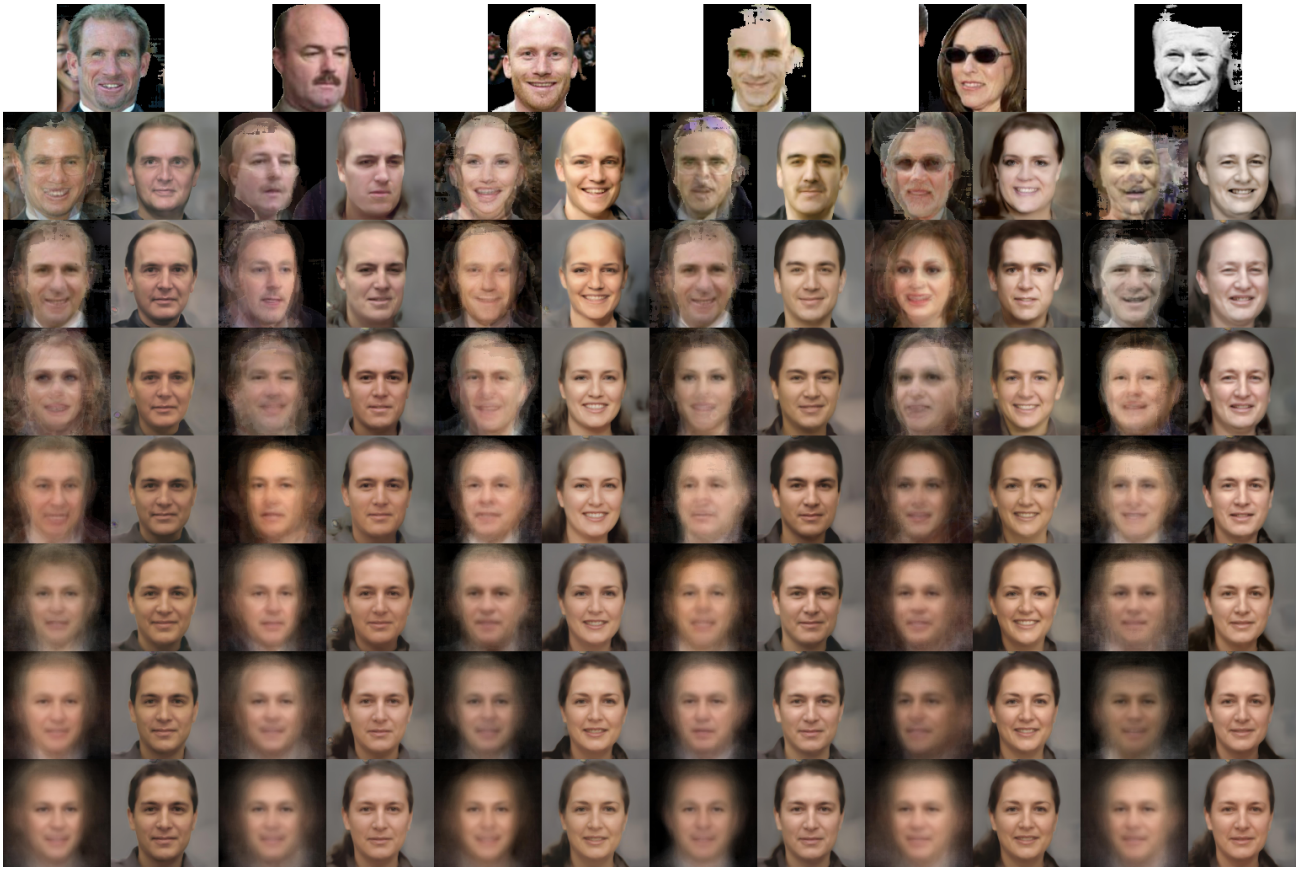

図 4　従来手法と提案手法の匿名化画像の出力の例．最初の行が元画像を表し，2 行目から上から順に K=[2,4,8,16,32,64,128]のときの結果を示している．奇数列は従来手法，偶数列は提案手法 MIKU の結果を表す．

Figure 4　Output result of conventional method and proposed method: The top row represents the original face images, and the following lines represent the anonymized image with K = [2, 4, 8, 16, 32, 64, 128] in order from the top. The odd column is the anonymized image by the conventional method, and the even column is the anonymized image by the proposed method.

ただし，N は匿名化する画像の総数であり，Q は匿名化の結果同じ準識別子を共有する顔画像の ID のグループの集合であり，qはそのグループの 1 つを表す．I_iは ID が i の顔画像をである．

この式から，h を求める関数 f が線形である場合，FID を求めるのに使うμ_{diff}が 0 になることが明らかである．ここで問題になるのは，関数 f の線形性である．通常 NN は非線形な関数を学習しており，Inception モデルもそのうちの 1 つである．しかし，Inception モデルを構成する畳み込み演算と線形写像はともに線形な処理であり，Inception モデルが非線形な関数足りうるのは，活性化関数として使われている Relu 関数の働きによるものである．そこで，Relu 関数の線形性ついて議論する．

Relu 関数は以下の(5)式で定義される関数である．

$$\mathrm{Relu}(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5}$$

この式を見て分かるように，定義域が$x > 0$または$x \leq 0$である場合は，Relu 関数は線形な関数になる．ここで，N 個の値$(X_1, X_2, \ldots, X_N)$の平均値を Relu の前後で等号が成り立つときについて考える．

$$\frac{1}{N}\mathrm{Relu}\left(\sum_i X_i\right) = \sum_i \mathrm{Relu}\left(\frac{1}{N}X_i\right)$$

この等号が成り立つときは，すなわち，N 個の値すべてが同符号の場合である．例えば，X_iが正の値を持つ確率を p としたとき，すべての値が同符号となる確率は

$$p^N + (1-p)^N$$

で与えられ，$p < 1$であるから N が増大すると，この確率は小さくなる．つまり，N が増大すると Relu 関数の線形性は失われていくと言える．ここで，従来手法について考えると Relu の線形性の議論における N の増大とは K の増大に相当する．図 3 は(3)式におけるμ_{diff}とΣ_{diff}の K に対する推移を示した図である．これを見ると，K が 2 から 4 へ増大するときのμ_{diff}の増加は，MIKU での増加に比べて，従来手法での増大の方が大きいことが分かる．これは，先の議論の Relu の線形性が崩壊したことによって，関数 f の線形性が失われたためだと推測できる．つまり，K=2 のときは，

従来手法は Relu 関数の線形性のため，異常に低い FID を得ており，FID では，従来手法の質の高さを正しく判定できない．また，先の議論ではΣ_{diff}については未検証であるが，図 3 を見るとμ_{diff}同様の原理がΣ_{diff}にも働いていることが予想される．

5.2 定性評価

FID での評価に加えて，定性的に匿名化画像を評価する．図 4 は従来手法と提案手法の MIKU それぞれによる匿名化画像を示す．一番上の行が元画像を表し，次の行からは上から順に$\mathrm{K} = 2, 4, 8, 16, 32, 64, 128$で匿名化された顔画像を表し，奇数列目が従来手法，偶数列目が提案手法 MIKU のものを表している．まず従来手法について図 4 を見ると，従来手法においても，前処理で目と鼻，口の位置を合わせているため，それらの要素に関しては良く表現できている．しかし，顔の輪郭や髪の形状は不自然な表現になっている．また，平均値を取っているため，特に K が小さいときに，顔に不自然なエッジが入っている．このような点で従来手法は匿名化画像の質を下げているのが分かる．対して，提案手法の MIKU による匿名化画像は，目鼻口が正しく表現できているだけでなく，髪や輪郭も自然に表現し，顔の上に不自然なエッジが乗ることもない．これは，StyleGAN の潜在空間が自然な顔画像を出力するという目的に適した空間を持っているためである．

6. 結論

本論文では，はじめに高次元での匿名化に問題があることを確認した．これを解決するために，特定の関数によってデータを写像し，その写像された空間上で匿名化を行う MIKU を提案した．本論文では，MIKU の Mapper 部に StyleGAN の潜在空間を用いることとし，顔画像を匿名化の対象として匿名化を行った．作成した匿名化顔画像の品質を評価するために、Inception モデルによる FID を従来法と MIKU で測定し、K＝2 の場合を除いて、MIKU が従来法よりも高品質の画像を出力することを確認した。また，$\mathrm{K} = 2$ の場合についても，Inception モデルの K が小さいときの線形性について議論し，FID の値が従来手法に対して，低くなる理由が画像の質のためではないことを明らかにした．最後に定性的な評価も行った．また，定性的な評価も行い，不自然な輪郭や髪，顔のエッジなどの問題が従来手法には存在するが，MIKU には存在しないことを確かめた．

7. 展望

本論文では，顔画像の評価に質の評価として FID の値のみを用いたが，表情や性別といった，人が顔画像を見たときに明らかに予想する情報の変化を見る方法も検討していきたい．また，現在 1 人の顔画像を StyleGAN の潜在空間へ写像するのに，30 秒ほどの時間を要するため，大きな顔のデータベースには適応するのが難しい．よって，逆写像の方法について見直したい．例えば，勾配法を使わずに逆関数そのものを発見するなどして，高速化を行いたい．

8. 謝辞

本研究は、文部科学省科学技術研究費基盤研究 B (JP16H04455, JP17H01739)の一環として実施された。

9. 参考文献

[1] Z. Ren, Y. J. Lee, and M. S. Ryoo, “Learning to Anonymize Faces for Privacy Preserving Action Detection,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11205 LNCS, pp. 639–655, 2018.

[2] L. SWEENEY, “k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY,” Int. J. Uncertainty, Fuzziness Knowledge-Based Syst., vol. 10, no. 05, pp. 557–570, Oct. 2002.

[3] A. Meyerson and R. 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, 2005, p. 223.

[4] B. Kenig and T. Tassa, “A practical approximation algorithm for optimal k-anonymity,” Data Min. Knowl. Discov., vol. 25, no. 1, pp. 134–168, Jul. 2012.

[5] C. C. Aggarwal, “On k -anonymity and the curse of dimensionality,” Proc. 31st VLDB Conf., no. January 2005, pp. 901–909, 2005.

[6] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” arxiv, Dec. 2018.

[7] R. Abdal, Y. Qin, and P. Wonka, “Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?,” arxiv, Apr. 2019.

[8] “ffhq-dataset/download_ffhq.py at master · NVlabs/ffhq-dataset · GitHub.” [Online]. Available: https://github.com/NVlabs/ffhq-dataset/blob/master/download_ffhq.py. [Accessed: 05-Jul-2019].

[9] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” Nov. 2014.

[10] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arxiv, Sep. 2014.

[11] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” Jun. 2017.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” Dec. 2015.

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments,” 2007.