

Stable Diffusion と GHOST を用いたプライバシー保護のための顔変換システムの実装に関する研究

池田 麻輝 山富 龍 マハブービ シェヘラザード 二宮 洋
 湘南工科大学工学部情報工学科

1. はじめに

近年、人工知能の急速な発展により、様々な技術が開発された。中でも、DeepFake などのように写真や動画の人物の顔情報を偽り、社会から批判を受けた技術も開発されている。本研究では、画像や動画の人物の顔情報を変更することを偽る行為ではなく、プライバシー保護の手段として着目し、匿名性のある顔変換システムの実装を試みた。本研究では、DeepFake 技術である Generative High-fidelity One Shot Transfer(GHOST)[1]とテキストから画像を生成する Stable Diffusion[2]を組み合わせたシステムを提案する。DeepFake はターゲットの顔をソースの顔と入れ替えるシステムであり、実在する人物同士で顔を変換するため、プライバシーに問題がある。本システムでは、Stable Diffusion で生成した実在しない人物の画像をソース画像として用いることで、匿名性を保持する。このシステムにより従来のぼかしやモザイク処理よりもターゲットの状況をリアルに表現した画像及び動画を生成することが可能となる。

2. GHOST[1]

GHOST は画像から画像、画像から動画へ顔を交換する画像認識モデルである。GHOST の構造を図 1 に示す。図 1 では、 X_s はソース画像、 X_t はターゲット画像であり、 X_s と X_t が GHOST の入力となる。 X_s は、事前学習済みの ArcFace モデルである Identity Encoder のエンコーダに入力され、ソース画像の人物のアイデンティティ特徴を保持するベクトル Z_{id} に変換される。 X_t は Attribute Encoder のエンコーダに入力される。Attribute Encoder は U-Net 構造を持つモデルであり、属性特徴を抽出したベクトル $Z_{att}^1, Z_{att}^2 \dots Z_{att}^{n-1}, Z_{att}^n$ に変換される。AAD Generator は複数の AAD block からなる AAD ResBlock から構成されており、AAD block は

Identity Encoder, Attribute Encoder の情報を抽出する。AAD Generator では、Identity Encoder, Attribute Encoder, AAD ResBlock の出力を順次入力することで、ソース画像のアイデンティティ情報及び、ターゲット画像の属性特徴を持つ新しい顔 $\hat{Y}_{s,t}$ を生成するモデルである。

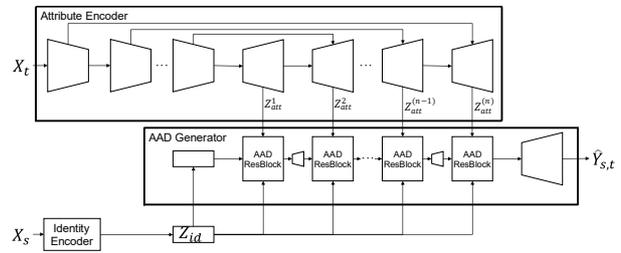


図 1 GHOST の構造

3. 顔変換システム

本研究では、入力データが画像-画像、画像-動画の一組から出力を生成するワンショットパイプラインである点と、画像と動画両方の交換が可能である GHOST に着目する。また、高精度な画像を短時間で生成できる点と、入力された任意のテキストに基づいた画像を生成するテキストエンコーダーであることに着目し、仮想の顔画像生成に Stable Diffusion を使用する。具体的には、プライバシー保護の観点から、DeepFake 技術である GHOST のソースとなる入力に Stable Diffusion により生成した顔画像を用いた新たな顔変換システムを提案する。提案システムの構造を図 3 に示す。ここで、Keywords, X_t が入力、 $\hat{Y}_{s,t}$ が出力となる。Keywords は表 1 に示す選択式のキーワードによって構成される。Keywords を Stable Diffusion の入力として与え、Source となる X_s' を生成する。 X_t をターゲット(Target)画像として、 X_s' をソース画像として GHOST に入力として与え、Blending となる $\hat{Y}_{s,t}$ を出力する。図 4 に提案システムの完成例を示す。

次に、提案システムについて説明する。提案システムの X_s' を生成するためのキーワード選択画面を図 4 左部に示す。年齢、眼鏡の有無のチェックボックスにチェックを解除した場合や、その他

A Study on the Implementation of Face Transformation System for Privacy Protection Using Stable Diffusion and GHOST
 Asaki IKEDA, Ryo YAMATOMI, Shahrzad MAHBOUBI, Hiroshi NINOMIYA
 Department of Computer Science, Faculty of Engineering, Shonan Institute of Technology

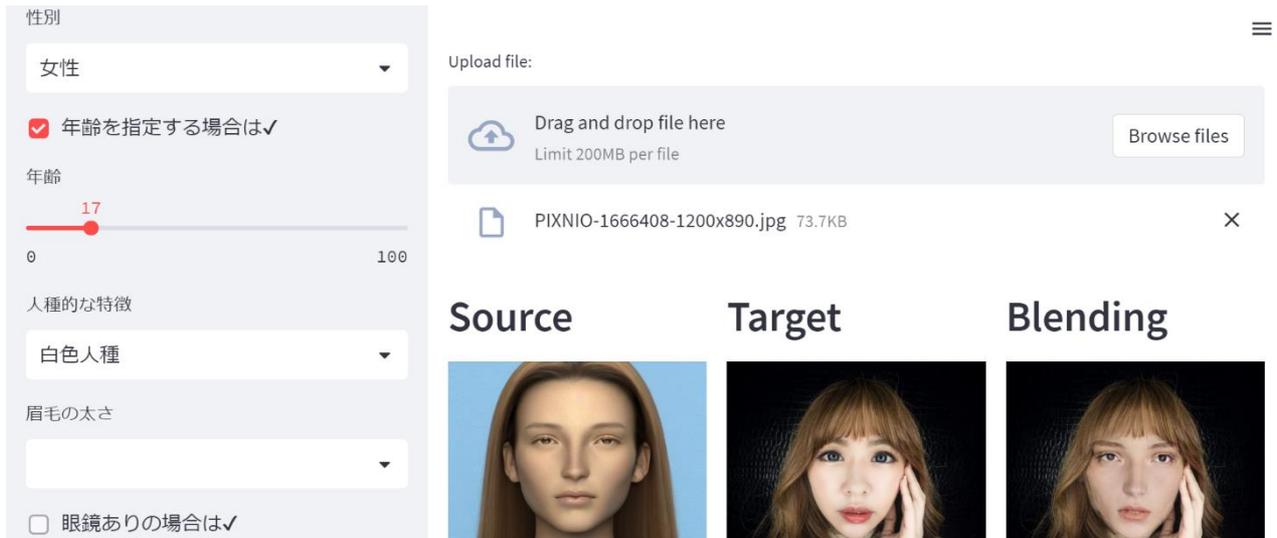


図4 提案システムの画面

の項目で空欄を選んだ場合は Stable Diffusion の入力には該当するキーワード自体が入力されない。選択したキーワードは該当する英単語に変換し、1つの文章として結合する。この際、人物の顔を生成しやすいように、文章の末尾に「face front, realistic」と追記しておく。その為、キーワードを全て選択されていない場合でも人物の顔画像が生成される。



図5 実験結果

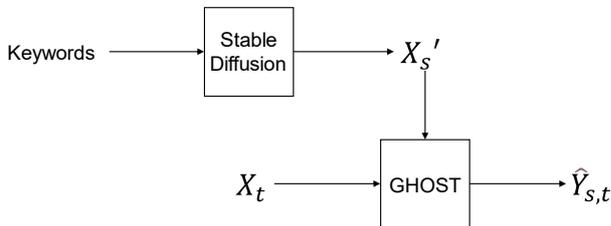


図3 提案システムの構造

表1 キーワード一覧

項目	選択肢				
	性別	選択なし	男性	女性	
人種的な特徴	選択なし	黒色人種	白色人種	黄色人種	黒褐色人種
眉毛の太さ	選択なし	太い	細い		
眼鏡の有無	選択なし	眼鏡あり			
年齢	選択なし	1 - 100までの任意の整数			

4. 実験

本研究の有効性を検証するため、作成したシステム内でキーワードを選択し、文章の作成を行った。選択したキーワードは、「性別：男性，年齢：30，人種的な特徴：黒色人種(Negroid)，眉毛の太さ：太眉，眼鏡の有無：選択なし」となっており，生成された文章は「Negroid thick eyebrows 30 years old male face front, realistic」である。上記の条件で実行した結果を図5に示す。

図5より，生成した文章から実在しない人物の顔画像の生成が問題なく行なえる事が確認できた。その上で，目的としていた実在する人物の顔に対して，生成した画像を自然に入れ替えることが出来た。

5. まとめ

本研究では，1枚の画像から実在しない人物の顔に置き換えるシステムを提案した。人物画像の生成に Stable Diffusion を，画像同士の合成に GHOST を使用し，匿名性を保持し，ターゲットの状況をリアルに表現した画像を生成することを実現した。

参考文献

- [1] A. Groshev, et al., "GHOST—A New Face Swap Approach for Image and Video Domains," *Proc. IEEE Access*, vol. 10, pp. 83452-83462, (2022).
- [2] Stability AI, "stablediffusion", GitHub, 2022; <https://github.com/Stability-AI/stablediffusion> (最終確認日:2023年1月10日)
- [3] A.Radford, et al., "Learning transferable visual models from natural language supervision.", *Proc. PMLR* (2021).