② Deepfake生成と検出の現状

Trung-Nghia Le†, Huy Hong Nguyen†, 山岸順一†, 正会員 越前 功†

キーワード：deepfake

❶ まえがき

深層学習はデータからディープニューラルネットワーク（Deep Neural Network: DNN）を学習する機械学習手法の一種であり，コンピュータビジョン，音声認識，自然言語処理，バイオインフォマティクス，医療画像解析などさまざまな分野に応用されている．加えて，AlexaやSiriなどのディジタルアシスタント，自動運転，がん検出，機械翻訳など，身近なアプリやサービスですでに利用されている．

この深層学習技術の発展により，画像や映像などの視覚メディアの生成に関しても大きな発展が起きている．例えば，テキストから画像の生成，画像から画像への変換，超解像技術，画像修復，非写実的画像の生成など，多くの成功例を挙げることができる．そして本稿の話題である人物を含む画像・映像の生成に関しても大きな進展が起きている．

この深層学習による生成画像とコンピュータ生成画像（Computer-generated Imagery: CGI）はどちらもコンピュータを利用して視覚メディアを作成するという点では共通する一方，その生成の際に必要な人間による作業量は大きく異なる．前者の場合，DNNによる生成モデルの学習を一旦行ってしまえば，人による作業をほぼ必要とせずに画像の生成が可能であり，それゆえ画像や映像をすばやく，そして，安価に自動生成できる（図1）．一方，後者のCGIは，2D/3Dアニメーション，キャラクタ，映画，テレビ番組，ビデオゲーム，特殊効果などの視覚メディアを作成するため利用されているが，その多くのプロセスは専門的な技能を持つ人に依存し，作業コストや所要時間も大きく異なる．

図に示すように，深層学習による生成画像は極めて写実的であり，教育，映画のプロダクション，芸術作品など幅広く利用できる可能性があることから，コンピュータビジョンやコンピュータグラフィックス分野でも注目されている．

一方で，深層学習による顔画像生成がDeepfakeという別名で呼ばれていることが暗示しているように，使い方によっては画像・映像の改ざんと偽造が可能になり，なりすまし，嫌がらせ，おどし，詐欺，フェイクニュースなどへの悪用が起こり得る．実際，すでに，ビデオ中の任意の人物の顔を別の人物の顔に入れ替えるDeepfakeソフトウェアも幾つか公開されている．そして，その端的な悪用例が任意の人物をポルノビデオの出演者にしてしまうDeepNudeなる名称のソフトウェアであり，対策が必要な喫緊の懸念事項である[4]．また，映像のみならず，音声の生成技術「音声合成」に対しても同じ問題が生じている．ウォールストリート・ジャーナル紙によると，英国エネルギー関連会社の社員が，AI音声技術を利用して生成された音声の指示に従い詐欺被害にあったと報道されている[31]．合成された音声は親会社の上司の特徴を模倣しており，上司と思い込んだ社員は，220,000ユーロをハンガリーにある発注先の口座に振り込んでしまったという．また，アメリカの歴代大統領および著名人の合成音声もYoutube上に公開されている*1．深層学習による文章の自動生成ももちろん可能であり，文献[30]では本物のような国連総会のスピーチ文章を容易に生成できることに触れている．

前述の通り，ビデオ中の任意の人物の顔を別の人物の顔に入れ替えるDeepfakeソフトウェアはスマートフォンのアプリとしても利用可能になるなどその利用障壁は下がっており，使い方によってエンタテインメントにも，他人を



図1　左画像：StyleGAN2[10]によって生成された画像
右画像：著者らにより構築されたOpenForensicsデータセット[12]に含まれる匿名化処理を行った顔画像（上：元の顔画像，下：匿名化顔画像）．

†国立情報学研究所
"Deepfake Generation and Detection in Practice" by Trung-Nghia Le, Huy Hong Nguyen, JunichiYamagishi, and Isao Echizen (National Institute of Informatics, Tokyo)

*1　Vocal Synthesis, https://www.youtube.com/channel/UCRtfquxnij9wDnFJnpPS2Q/feautured

攻撃するツールにもなり得るこの技術を理解し，その性質を把握し，そして対策を検討することが現在重要な課題となっている．本稿ではDeepfakeの生成と検出の現状について概説した後，種々の現実的な課題について言及する．

❷ Deepfake生成

これ以降，深層学習による種々の画像生成技術において顔に着目した生成手法を「Deepfake生成」と呼び，その手法について解説する．なお深層学習の知識を保有するがメディアフォレンジック（Media Forensics）の知識までは持たない技術者・研究者を読者と想定した．

2.1　概　要

Deepfake生成手法は，顔の入れ替え（Identity Swapping），表情の入れ替え（Expression Reenactment），顔全体の合成（Face Synthesis），顔の属性操作（Attribute Manipulation）など，その用途によって分類することができる．例えば，顔の入れ替えは画像や映像中の対象人物の顔を別の人物の顔に入れ替え，表情の入れ替えでは顔の表情を他の人物の表情となるように操作する．顔全体の合成は，ノイズ（擬似乱数など）から実世界に存在しない顔画像を生成する．顔の属性操作は，顔画像の髪の色や肌の色といった属性を変更した顔画像を生成する．この用途による分類は学術論文などで広く利用されているが，当然ながら手法の中には複数の機能を持つものがあり，上記以外にもさまざまな分類がある．

そこで本稿では，Deepfake生成モデルを学習する際の学習データの観点から，Deepfake生成手法を対データ学習（Paired Training），少数ショット学習（Few-shot Learning）とゼロショット学習（Zero-shot Learning）に大別した上で，上記の用途による分類を可能な限り概説する．また，使いやすさ，処理時間，出力の本物らしさの観点からも，それぞれの長所・短所について議論する．

対データ学習によるDeepfake生成手法は特定の人物を対象とする枠組みで，入れ替え元の人物と入れ替え先の人物の双方について大量の学習データが必要となり，データ収集およびモデル学習に時間・コストを要する．それに対し，少数ショット学習によるDeepfake生成手法は，不特定人物の大量データに基づいた事前学習済みモデルを利用することで少量の学習データから対象の人物のDeepfake画像生成モデルを構築する枠組みである．事前学習モデルの学習には時間とコストを要するものの，ターゲットとなる人物用の生成モデルの構築は少量の学習データで済む点が特徴であり，対データ学習の制約が緩和されている．ゼロショット学習は少数ショット学習の発展系であり，ターゲットとなる人物の画像を一切学習に利用せずDeepfake生成モデルを構築するという枠組みである．

2.2　対データ学習による生成手法

2.2.1　概　要

対データ学習によるDeepfake生成手法は，特定の人物の顔画像を対象としたものであり，早くは2017年頃から利用されている[1]．

顔の入れ替え（Identity Swapping）を行う初期のDeepfake生成では，Autoencoder（AE）という深層学習ネットワークがその用途に利用されていた．AEはエンコーダ部とデコーダ部から構成され，エンコーダ部は入力顔画像から顔特徴量を抽出し，デコーダ部はその特徴量から元の顔画像を復元する．入れ替え先の顔画像データと入れ替え元人物の顔画像データをそれぞれ適量用意し，AEネットワークを各々の人物に対して個別に学習する．そしてデコーダ部を入れ替えることで，入れ替え元の顔画像の特徴量から入れ替え先の顔画像を生成するネットワークを構築する．これにより，入れ替え元の顔の表情に基づき入れ替え先の顔画像を生成することができる．この手法は入れ替え元，入れ替え先の人物双方について大量の学習データが必要となるものの，現在でもDeepfake生成手法として最も広く利用されている手法である．

その後，敵対的生成ネットワーク（Generative Adversarial Network: GAN）を利用した手法も開発され，生成された顔画像がよりリアルになっていった．同時に，DeepFaceLab[26]やFaceSwap-GAN[2]のような顔の入れ替え用ツールとモデルがGitHub上でチュートリアル付きオープンソースコードとして公開されたことにより，Deepfake画像や映像を作る難易度が下がっていった．

次に，顔の表情の入れ替え（Expression Reenactment）について触れる．この手法では，入れ替え元の人物の表情と頭部の動きを入れ替え先の人物に適用し，入れ替え先の人物のIDは維持する．言うならば表情のコピーである．ただ，コピーされる表情には各画像の唇の形状も含まれていることから，結果として異なる発話内容に唇の動きを同期させるという用途にも応用可能であることに注意が必要である．研究分野においては，一般的なウェブカメラで撮影された入れ替え元の人物の顔表情を，入れ替え先の人物の顔の3次元モデルに適用してリアルタイムでレンダリングする手法がThiesらによって提案されている[32]．またKimらは，入れ替え元の人物の情報を時空間的にエンコーディングし，比較することで，生成された映像の時間的整合性を担保する手法を提案している[11]．これにより，頭部の動きや顔表情，目の動きをより正確に入れ替えることを可能にしている．

2.2.2　注意点

対データ学習によるDeepfake生成手法はDNN学習データに含まれる特定の人物を対象としており，他の人物に対して適用することができない．他の人物に適用する際には，入れ替え元，入れ替え先の両方の人物の画像を用意し再学習を行う必要がある．また前述したGANを利用しリアルさを向上させる手法は，一般により大量の学習データが必要になる傾向にある．したがって学習にも長時間を要する．

一方で，対データ学習による手法は，入れ替え元，入れ

替え先の両方の人物について大量のデータを利用して学習することから，頭部の姿勢やその動きが大きく異なる場合でも高品質な生成が可能である．

2.3　少数ショット学習による生成手法

対データ学習のアプローチと比較して必要になる画像データ量を減らした，少数ショット学習（Few-shot Learning）と呼ばれる手法も存在する．この手法では，必要となる対象人物の画像データ量を削減するため，DNNの学習を二段階に分けて行うことが多い．それぞれ事前学習とファインチューニングと呼ばれ，利用するデータ量が大きく異なる．一段階目の事前学習では多数の人物の画像に基づく大量データを利用し学習を行い，二段階目のファインチューニングでは対象人物の数枚の画像（または1枚の画像）のみを学習に利用する．一段階目の事前学習ではDNNのパラメータを大まかに学習し，二段階目の学習において学習済みパラメータを対象人物に合わせて微調整するという流れである．

この少数ショット学習手法は，表情の入れ替えに対して利用されることが多い様である．ここでは研究分野において発表された論文をいくつか紹介する．細かい技術的な内容は各文献を参考のこと．Zakharovらは，入れ替え先の人物の1枚の画像中の顔表情を，入れ替え元の人物の顔表情に入れ替える手法を提案した[37]．Pumarolaらは，顔画像から抽出されるAction Unitを画像生成の条件とし，顔表情が変化する動画を生成するモデルを構築した[27]．Soumyaらの手法は，1枚の画像に対してGANを利用することで画像内の顔の姿勢と表情を制御する[33]．この手法では，姿勢をある画像から，表情を別の画像からというように，入れ替えのために複数の異なる画像を利用する．Haらは，データセットに含まれない個人に対する利用であっても生成画像の品質を維持するために，顔の部位の位置合わせを行う手法を提案した[6]．風景生成のために提案された画像生成手法を顔表情の入れ替えに利用する試みもある[38)8)]．

2.4　ゼロショット学習による手法

2.4.1　概　要

最後にゼロショット学習について言及する．これは少数ショット学習の発展系であり，ターゲットとなる人物の画像を一切学習に利用せず，事前学習モデルのみで任意の人物のDeepfake生成モデルを構築することを目指す枠組みである．少数ショット学習以上に，大量のデータにより事前学習されたモデルを積極的に活用することを基本的な方針とするが，現時点で確立された手法はまだ存在しない．

研究分野において発表された顔の入れ替え手法の論文をいくつか紹介する．Natsumeらは，入れ替え先の人物を含む画像の顔以外の領域と，入れ替え元の人物の特徴量を入力として，偽の人物を生成するVAE-GANを提案した[23]．文献[14]において提案されたFaceShifterは入れ替え元に忠実で，遮蔽された領域が存在する顔画像にも適用可能な2段階処理方法を実現した．文献[35]において提案されたHi_Faceでは，本物のように見える生成結果を得るために，入れ替え先の人物の顔の形を残しつつ，入れ替え元の人物の顔に入れ替える枠組みを提案している．

顔の入れ替え手法に加え，顔画像全体の合成（Face Synthesis）と顔画像の属性操作（Attribute Manipulation）も，ゼロショット学習の一種と見做すことができる．例えば，GANによる顔画像の合成は，ターゲットとなる人物の情報ではなく，擬似乱数から顔画像を生成する．段階的に画像の解像度や品質，写実性を向上させることで，StyleGAN[9]は画像解像度を128×128から1024×1024まで向上させることに成功している．そして，StyleGAN2[10]では，不自然な部分を除去することで画像の品質をさらに向上させている．

顔画像の属性操作もGANを応用して行われる．操作可能な画像の属性には，髪の色，髪型，肌の色，性別，年齢，表情，メガネ，化粧などがある．StarGAN[5]では，複数のデータセットからモデルを同時に学習することによって，さまざまな顔属性や表情を同時に操作することを可能にした．

文献[29]では，特徴量の解きほぐし（Disentanglement）によって特徴空間の解釈性を向上させ，より正確に属性を操作することを試みている．

2.4.2　注意点

ゼロショット学習によるDeepfake生成は，特定の対象の人物に依存しない汎用性の高い技術である．しかしそれが故に，特定の人物の顔画像を生成するという用途に関しては問題も多く，任意の条件で実行可能とは言い難いのが現状である．これは，ゼロショット学習による生成モデルは過学習しやすい傾向にあるためであり，顔画像が学習データセットに含まれている人物に対する生成結果と比較して，学習データセットに含まれていない人物に対する生成結果は実際とはまったく異なる人物の顔画像を生成することが多い．また，顔の向きが正面から大きく離れる場合や，遮蔽領域が存在する場合には画像生成に失敗するなどその性能に限界があることも知られている．

またゼロショット学習による属性操作の多くの手法では，単一の属性を変更したい場合でも複数の属性が変更された画像が生成され，出力の制御が容易ではない．前述した通り特徴量の解きほぐし手法が提案されてはいるものの，データセットに含まれない属性についても同様に適用できるかについては疑問が残る状態である．

さらにゼロショット学習では，対象となる人物の画像を直接考慮する必要はないものの，対象となり得る人に基本的に共通する特徴や情報をあらかじめすべて学習しておく必要があり，非常に膨大な数の画像を利用して事前学習モデルを構築する必要がある．当然ながら対データ学習と少数ショット学習よりも一段とコストと学習時間が要求される．

このようにゼロショット学習技術は難易度が高く，画像処理に関する研究分野ではなく，実際のDeepfake生成現場においてどの程度利用されているかは現時点では不明である．

2.5　Deepfake生成画像のアーティファクト

前述した通り，Deepfake生成モデルの汎化性能に問題があり，結果として，低解像度，画質の低さ，不自然さなどさまざまなアーティファクトが発生する．

まずDeepfake生成モデルの学習データ量もしくは学習時間が不足している場合，高解像度・高品質の画像を生成することは困難であり，結果として，生成された画像解像度が低く，画質も悪くなる．

また生成された画像には不自然さも残る．まず元画像と生成画像の境界に不自然さが存在することが多い．これはDeepfake生成モデルが引き起こしたアーティファクトではなく，境界をなめらかにつなぐブレンディングアルゴリズムの限界によるものである．

そして画像中の顔の向きがモデル学習条件と異なる際にも生成画像は不自然になることが多い．これは事前学習の際には正面向きの画像を多く多用しているのに対し，日常の中で実際に撮影された画像中の顔の向きは大きく異なることがあり，対データ学習以外の手法では正面から大きく異なる向きの顔を対処することができないのが現状だからである．最新のブレンディングアルゴリズム[14]によってある程度不自然さを緩和することはできるものの，完全に自然と言える画像を生成することはできていない．

遮蔽が存在する場合も，生成された顔画像に不自然さが残る．遮蔽とは例えば髪の毛やメガネ，他のものによって顔の一部が隠された状況のことである．ごく最近の一部の手法[14]を除いて，既存のDeepfake生成手法では遮蔽に対して適切な処理を行うことが不可能であり，結果として不自然な画像が生成される．

2.6　Deepfake生成の正しい活用例

最後に，まえがきで言及した悪用例とは反対に，Deepfake生成を正しく活用した事例を紹介する．

その一例が，顔画像の属性操作モデルをメークアップに応用したバーチャルメイクである．この応用例では，利用者は各化粧品を使用したイメージを，Deepfake生成技術によって生成された自分の顔画像を見ることで確認することができる[19]．

プライバシー保護のため，顔画像の合成技術を匿名化手法として活用することも検討されている．例えば，ドライビングレコーダを通して記録された映像には，多くの人が映り込む．この運転映像には活用可能性があるが，プライバシーの保護のためには，映り込んだ人物を匿名化する処理が必要になる．そこで，MaximovらはGANによる顔画像合成技術を応用し，映像や画像に映り込んだ人物の顔に対し，その人物の識別に必要な特徴を除去した顔を合成し入れ替えることで，映像中の人物を自動で匿名化する手法を提案した[22]．また，本原稿の著者らも，画像中の複数の顔を匿名化するため，擬似乱数からGANにより生成された世の中に存在しない人の顔画像を生成し元顔画像と入れ替える枠組みを提案している[12]．

❸ Deepfake検出

3.1　概　要

このようにDeepfake生成は使い方によっては悪用も活用も可能な技術である．本稿のこれ以降のセクションではDeepfake生成が悪用された場合に，それに対する防御策として，Deepfake生成画像を自動検出する枠組みについて着目する．

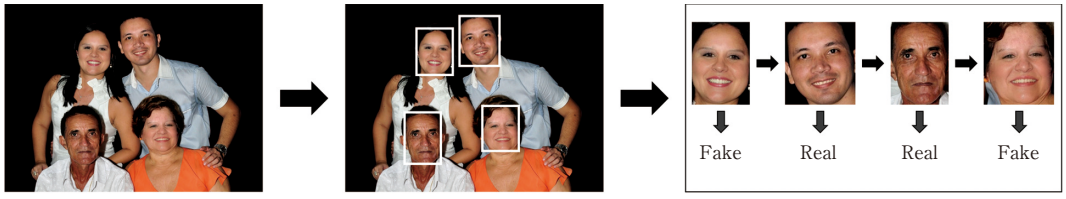
前述の通り，Deepfake生成手法によって生成された画像にはアーティファクトが依然として見られることがある．2.5項で言及したアーティファクトはほんの一例であり，人間に判別可能なアーティファクトと人間には判別不可能だが機械には判別可能なアーティファクトの両方を含めると，さまざまな違いがあると考えられる．言い換えると，本物の画像と何らかの操作を加えられた画像との違いをDNN等の機械学習手法によって学習することによって，Deepfakeかどうかを自動で判別するシステムを，Deepfake画像データベースから構築できることを示唆しており，いろいろな検出システムが提案されている．

初期のDeepfake検出は，図2に示すように切り抜かれた顔領域に対して本物か偽物かの判定を行うことが多い[3)25)]．一方，最近の検出手法では，顔領域の真正性の判定だけではなく，操作された領域をバウンディングボックス，またはピクセルレベルで特定する[24)12)]．以下では，Deepfake検出システムを利用されている概念に基づいて分類し，概説する．

3.2　Deepfake検出手法の分類

3.2.1　データ駆動型のアプローチ

データ駆動型手法では，本物の画像とDeepfake画像とが



(a) 入力画像　(b) 顔検出　(c) フェイク顔識別

図2　初期のDeepfake検出システムのフロー
二つのステップから構成されており，第1ステップでは画像（または映像）に含まれる顔領域を検出する．第2ステップでは，切り抜かれた顔画像それぞれに対し，本物か偽物かの識別を行う．

対になったデータベースから畳み込みニューラルネットワーク（Convolutional Neural Network: CNN）等を直接学習した識別器に基づく枠組みで，Deepfakeに関する特定の解析方法や手法自体に関する知識に依存しない．MesoNet[3]やCapsuleNet[25]，XceptionNet[28]などのように，よく知られたCNNアーキテクチャが利用されている．

3.2.2　フォレンジック的解析に基づくアプローチ

フォレンジック的解析に基づくアプローチによるDeepfake検出は，Deepfake画像に見られる特有のアーティファクトの知見や分析に基づき特徴量をデザインし，識別器を構築する手法である．例えば，瞬きの有無[16]や頭部の姿勢[36]，肌の質感[20]，虹彩や歯の色[21]などを特徴量とした報告がされており，その有効性が確認されている．他にも，顔の変形[17]や口の不自然な動き[7]を利用するものも提案されている．

3.2.3　周波数領域での信号処理によるアプローチ

近年提案されたいくつかのDeepfake検出では，周波数領域での手掛かりを利用している．例えば，Liらは適応的周波数特徴生成モジュール（Adaptive Frequency Feature Generation Module）を提案し，周波数領域の情報と距離学習を組み合わせることで，特徴空間での分離可能性を向上させた[13]．また，Liuらは画像とその位相スペクトラムを組み合わせることで，アップサンプリングの際に生じる歪みを捉え，Deepfake検出の性能を向上している[18]．人間による判別には必ずしも有益な情報ではないが機械によるDeepfake検出には役に立つという一例である．

3.2.4　説明可能なAIによるアプローチ

データ駆動型のアプローチに基づくDeepfake検出はブラックボックスなCNN等のモデルに完全に依存しており，モデルが顔画像内のどういう情報に基づいて判別をしているのかは明らかにならない．これらのモデルによる判断が常に正しいとは言えないこと，さらにモデルに対する敵対的攻撃に脆弱である可能性を考慮すると，判断根拠を明確に示すことができないモデルの出力を盲目的に受け入れることには危険が伴う．そこでDeepfake検出結果をより信頼できるものにするために，アテンション機構やセグメンテーションを利用し，エビデンスとなる領域を可視化する手法が提案されている．アテンション機構は，識別精度向上につながる領域を重みづけするために利用され，その重みを可視化することにより識別に貢献した領域を示すことができる[39][34]．同様にセグメンテーションは，顔画像中の領域の操作がどのDeepfake生成によるものかを提示するために利用されている[24][12]．この他，元画像と生成画像の境界領域のみを検出する枠組みもある．この境界領域検出は特定のDeepfake生成手法に依らずに利用できるというメリットがあるが，その一方で種々のブレンディング手法を利用した学習データベースを用意する必要がある[15]．

3.3　Deepfake検出のその他の課題

Deepfake検出技術にもいろいろな課題が存在する．

まず，通常のDeepfake検出は，与えられた顔画像1枚1枚に対して，本物か偽物かを分類するため，複数の顔を含む画像に対しては，処理に時間を要するという問題がある．また，前処理となる顔検出技術の検出精度にも大きく左右される．

また，Deepfake検出モデルはDeepfake生成手法により生成された画像や映像からなる大量のデータを訓練データとすることから，利用する学習データベースに含まれるDeepfake生成手法とその内容に強く依存する傾向がある．

したがって画像や映像のバリエーションもしくはDeepfake生成手法が少ない場合は，過学習が生じる可能性がある．

さらに，Deepfake検出の精度は，圧縮やノイズの付与，色差や輝度の変化を含む画像処理や映像処理に対して充分に頑健でない可能性もある．モデル学習に利用した際と同一の条件による画像に対しては高い識別性能が得られるものの，学習データに含まれない条件下の画像に対しては正しく検出することができない可能性がある．

これらのことから言えることは，Deepfake検出技術を一般のユーザに広くそして実用的に利用してもらうためには，まだ種々の改善が必要であり，更なる技術開発が必要であるということであろう．

❹ むすび

本稿では，顔に着目した深層生成モデルであるDeepfakeおよびその検出技術の両方について現状と限界を概観した．どちらも発展途中の技術であり今後更なる展開があるであろう．またDeepfake生成の正しい活用例についても紹介し，技術の2面性についても言及した．

最後に，国立情報学研究所のシンセティックメディア国際研究センターではDeepfake検出に関する技術開発を促進することを目的に，SYNTHETIQ＊2という名称のプログラムを開発し，現在企業パートナーを募集している．こちらも参考になれば幸いである．

謝辞　本稿は，JSPS科研費JP16H06302，JP18H04120，JP21H04907，JP20K23355，JP21K18023，およびJST CREST JPMJCR18A6，JPMJCR20D3の支援による．

（2022年4月22日受付）

〔文　献〕

1）Deepfakes software for all, https://github.com/deepfakes/faceswap（2021年2月18日閲覧）
2）A denoising auto encoder, adversarial losses and attention mechanisms for face swapping, https://github.com/shaoanlu/faceswap-GAN（2021年2月18日閲覧）
3）D. Afchar, V. Nozick, J. Yamagishi and I. Echizen: "Mesonet: A compactfacial video forgery detection network", In International Workshop on Information Forensics and Security, pp.1-7 (2018)
4）J. Brandon: "Terrifying high-tech porn: Creepy 'deepfake' videos are on the rise", https://www.foxnews.com/tech/terrifying-high-tech-porn-creepy-deepfake-videos（2021年2月18日閲覧）

＊2　https://research.nii.ac.jp/~iechizen/synmediacenter/synthetiq

5) Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim and J. Choo: "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation", In CVPR (2018)
6) S. Ha, M. Kersner, B. Kim, S. Seo and D. Kim: "Marionette: Few-shotface reenactment preserving identity of unseen targets", In AAAI Conference on Artificial Intelligence (2020)
7) A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic: "Lips don't lie: A generalisable and robust approach to face forgery detection", In CVPR, pp.5039-5049 (2021)
8) H. Hao, S. Baireddy, A.R. Reibman and E.J. Delp: "Far-gan for one-shot face reenactment", In CVPR Workshops (2020)
9) T. Karras, S. Laine, and T. Aila: "A style-based generator architecture for generative adversarial networks", In CVPR, pp.4401-4410 (2019)
10) T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila: "Analyzing and improving the image quality of StyleGAN", In CVPR (2020)
11) H. Kim, P. Garrido, A. Tewari, W. Xu, J. Thies, M. Niessner, P. Pérez, C. Richardt, M. Zollhöfer and C. Theobalt: "Deep video portraits", ACM Transactions on Graphics, 37, 4 (2018)
12) T.-N. Le, H.H. Nguyen, J. Yamagishi and I. Echizen: "Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild", In ICCV (2021)
13) J. Li, H. Xie, J. Li, Z. Wang and Y. Zhang: "Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection", In CVPR, pp.6458-6467 (June 2021)
14) L. Li, J. Bao, H. Yang, D. Chen and F. Wen: "Faceshifter: Towards high fidelity and occlusion aware face swapping", InCVPR (2019)
15) L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen and B. Guo: "Face x-ray for more generalface forgery detection", In CVPR (June2020
16) Y. Li, M. Chang, and S. Lyu "In ictu oculi: Exposingai created fake videos by detecting eye blinking", In International Workshop on Information Forensics and Security, pp.1-7 (2018)
17) Y. Li and S. Lyu: "Exposing deepfake videos by detecting face warping artifacts", In CVPR Workshops (2019)
18) H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang and N. Yu: "Spatial-phase shallow learning: rethinkingfaceforgery detectionin frequency domain", In CVPR, pp.772-781 (2021)
19) S. Liu, W. Jiang, C. Gao, R. He, J. Feng, B. Li and S. Yan: "Psgan++: Robust detail-preserving makeup transfer and removal", IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
20) Z. Liu, X. Qi, and P.H.S. Torr: "Global texture enhancement for fake face detection in the wild", In CVPR (June2020
21) F. Matern, C. Riess and M. Stamminger: "Exploiting visual artifacts to expose deepfakes and face manipulations", In Winter Applications of Computer Vision Workshops, pp.83-92 (2019)
22) M. Maximov, I. Elezi and L.L.-Taixe: "Ciagan: Conditional identity anonymization generative adversarial networks", In CVPR (June 2020)
23) R. Natsume, T. Yatagawa and S. Morishima: "Fsnet: An identity-aware generative model for image-basedface swapping", In ACCV (2018)
24) H.H. Nguyen, F. Fang, J. Yamagishi and I. Echizen: "Multi-task learning for detecting and segmenting manipulatedfacial images and videos", In International Conference on Biometrics: Theory, Applications and Systems (2019)
25) H.H. Nguyen, J. Yamagishi, and I. Echizen: "Capsule-forensics: Using capsule networks to detect forged images and videos", In ICASSP (2019)
26) I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, Mr. Dpfks, C. ShiftFacenheim, L. RP, J. Jiang, S. Zhang, P. Wu, B. Zhou and W. Zhang: "Deepfacelab: A simple, flexible and extensible face swapping framework", Arxiv Pre-print Arxiv:2005.05535 (2020)
27) A. Pumarola, A. Agudo, A.M. Martinez, A. Sanfeliu and F.M.-Noguer: "Ganimation: One-shot anatomically consistentfacial animation", International Journal of Computer Vision (2019)
28) A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner: "Faceforensics++: Learning to detect manipulated facial images". In ICCV, pp.1-11 (Oct 2019)
29) Y. Shen, J. Gu, X. Tang, and B. Zhou: "Interpreting the latent space of gans for semantic face editing", In CVPR (2020)
30) J. Shieber: "For less than $10, anyone can make an ai write a fake un speech", https://techcrunch.com/2019/06/07/for-less-than-10-anyone-can-make-an-ai-write-a-fake（2021年2月18日閲覧）
31) C. Stupp: "Fraudsters used ai to mimic ceo's voice in unusual cybercrime case", https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual（2021年2月18日閲覧）
32) J. Thies, M. Zollhöfer and M. Nießner: "Deferred neural rendering: Image synthesis using neural textures", ACM Transactions on Graphics, 38, 4 (2019)
33) S. Tripathy, J. Kannala and E. Rahtu: "Icface: Interpretable and controllable face reenactment using gans", In WACV (2020)
34) C. Wang and W. Deng: "Representative forgery mining for fake face detection", In CVPR, pp.14923-14932 (June 2021)
35) Y. Wang, X. Chen, J. Zhu, W. Chu, Y. Tai, C. Wang, J. Li, Y. Wu, F. Huang, and R. Ji: "Hififace: 3d shape and semantic prior guided high fidelity face swapping", In IJCAI, pp.1136-1142 (2021)
36) X. Yang, Y. Li and S. Lyu: "Exposing deepfakes using inconsistent head poses", In ICASSP, pp.8261-8265 (2019)
37) E. Zakharov, A. Shysheya, E. Burkov, and V. Lempitsky: "Few-shot adversarial learning of realistic neural talking head models", In ICCV (2019)
38) Y. Zhang, S. Zhang, Y. He, C. Li, C.C. Loy and Z. Liu: "One-shot face reenactment", In BMVC (2019)
39) B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang: "Wild deepfake: A challenging real-world dataset for deepfake detection", In ACM MM, pp.2382-2390 (2020)



Trung-Nghia Le（チュン ギア レ） 国立情報学研究所シンセティックメディア国際研究センター特任助教．総合研究大学院大学複合科学研究科情報学専攻博士課程修了．専門はコンピュータビジョン，機械学習．研究テーマは顕著物体領域検出，メディアフォレンジクスなど．博士（情報学）．



Huy Hong Nguyen（フイ ホン グエン） 総合研究大学院大学複合科学研究科情報学専攻博士課程修了．現在，国立情報学研究所情報社会相関研究系特任研究員．専門は機械学習，画像・映像処理，生体認証．研究テーマはメディアフォレンジクス，メディアセキュリティなど．博士（情報学）．



山岸　順一（やまぎし じゅんいち） 東京工業大学大学院総合理工学研究科物理情報システム創造専攻博士課程修了．現在，国立情報学研究所コンテンツ科学研究系教授．専門は音声情報処理，機械学習，生体認証．研究テーマは，音声をはじめとする生体特徴のディジタルクローン，ライブネス検知，メディアフォレンジクスなど．博士（工学）．



越前　功（えちぜん いさお） 現在，国立情報学研究所情報社会相関研究系主幹・教授，同研究所シンセティックメディア国際研究センター長，東京大学大学院情報理工学系研究科電子情報学専攻教授（併任）．専門はマルチメディア，情報セキュリティ，プライバシー保護．研究テーマはマルチメディアフォレンジクス，マルチメディアセキュリティ，生体情報保護．博士（工学）．正会員．