深層学習に基づく
カバー画像の生成と利用方法

会津大学コンピュータ理工学部
コンピュータ理工学科
教授　趙　強福

2019年12月17日

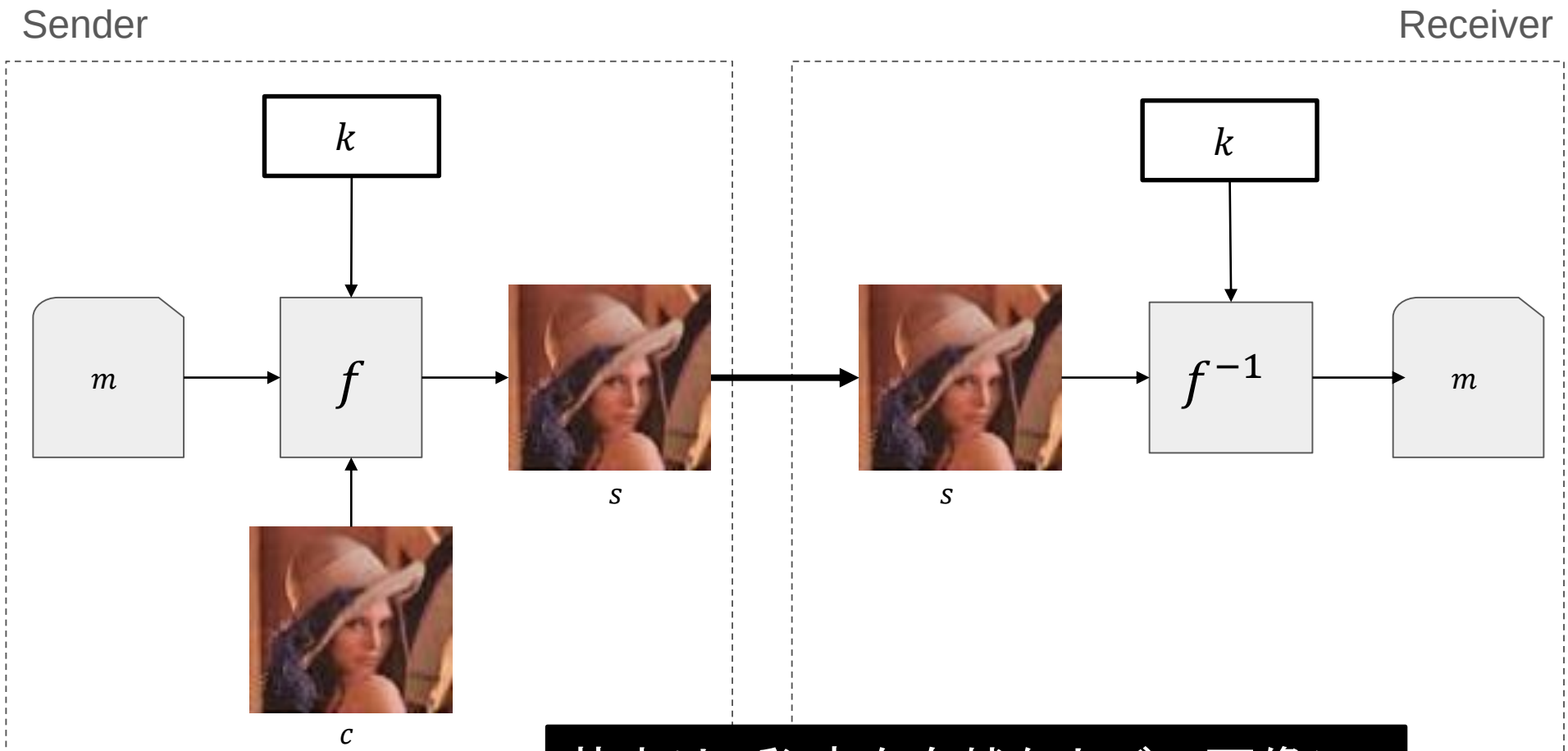
従来技術とその問題点

- ステガノグラフィーシステムは以下の「要素」がある:
  - Message m: 隠したい情報
  - Cover data c: 情報を隠すための「場所」
  - Embedding function f: 情報を隠すメソッド
  - Stego-key k (optional): メソッドfが使用するキー
  - Stego-object s: 情報を埋め込んだデータ
  - Extraction function f^{-1}: 埋め込んだ情報を抽出するメソッド

$$s = f(m, c, [k])$$
$$m = f^{-1}(s, [c], [k])$$

cとsは、同じ種類のデータ（画像、音声など）である。

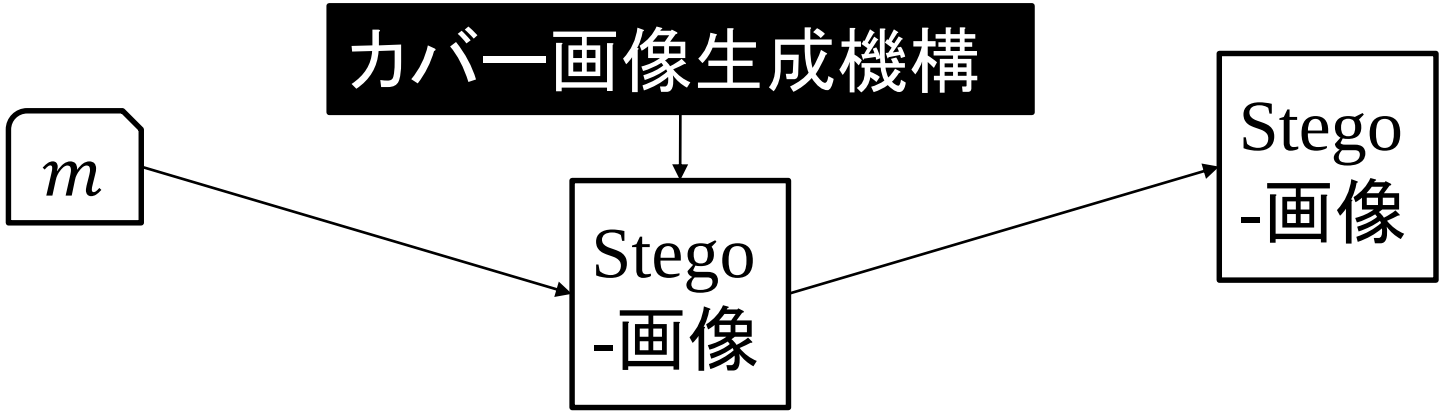
従来技術とその問題点



基本は、秘密を自然なカバー画像に隠し、その存在が公開しないこと

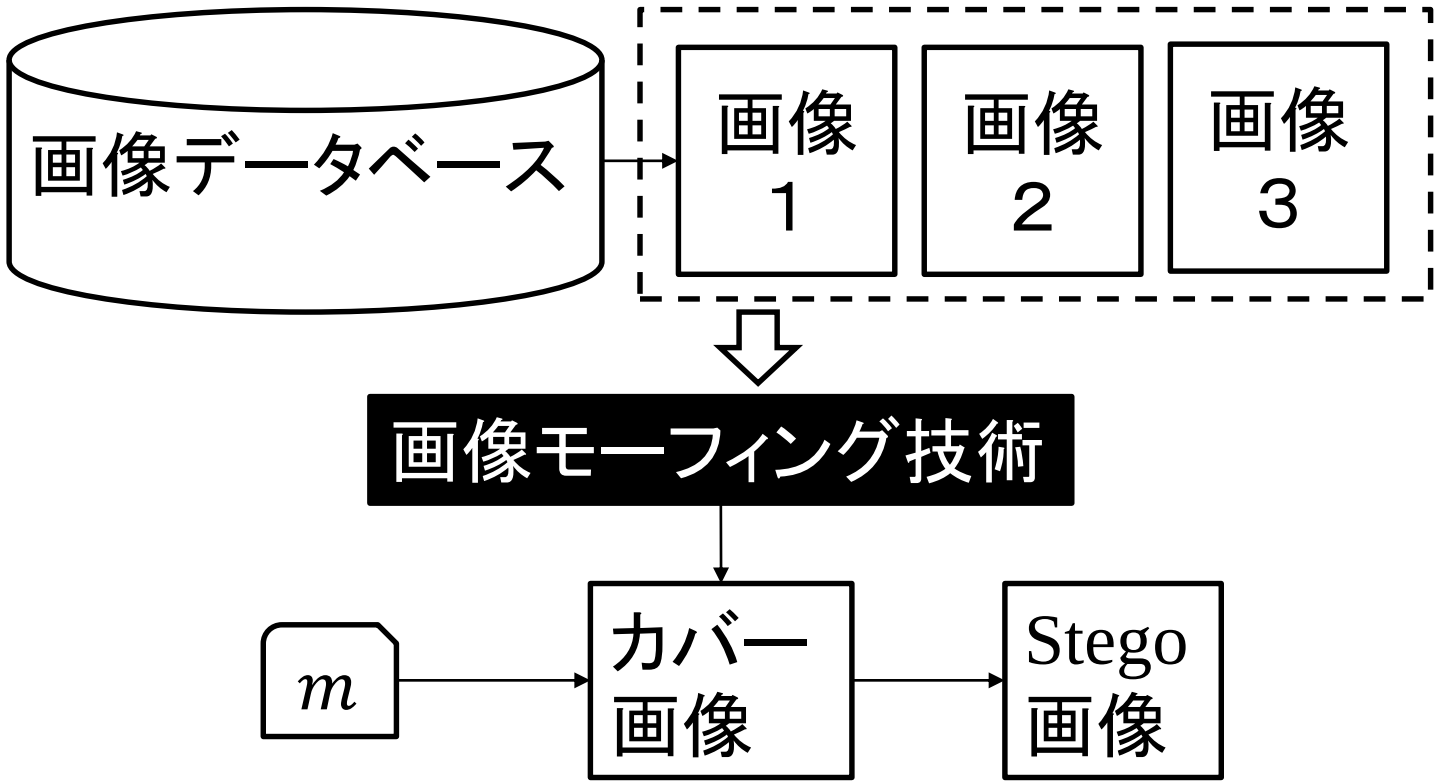
従来技術とその問題点

- カバー画像を複数回使用することができない。
  - 同じカバー画像から生成された異なるバージョンのステゴ画像を比較することによって、秘密の存在がばれてしまう。
- ステガノグラフィー技術を実用させるためには、異なる秘密情報を異なるカバー画像に埋め込む必要がある。
- そのためには、任意の秘密情報に対して、ユニークなカバー画像を生成することが重要である。



従来技術とその問題点

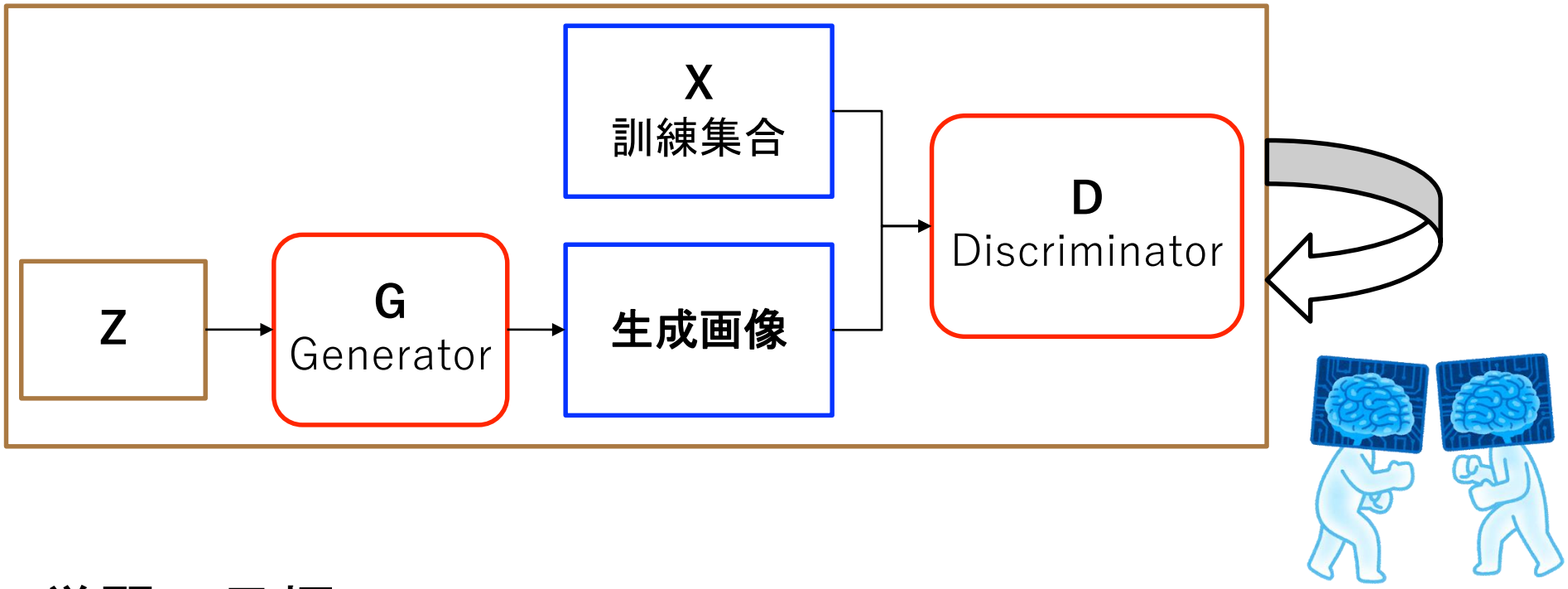
- われわれが以前に「一般化画像モーフィング技術」を提案し、それに基づくカバー画像生成法を提案した。
- 手動で「参考点」を求める必要があり、カバー画像を大量に生成することが難しい。



新技術の特徴･従来技術との比較

- 新しい方法は、GAN (**G**enerative **A**dversarial **N**etwork)を利用して自然らしいカバー画像を生成する。
- GANは、深層学習の一種で、与えられた画像の集合をもとに、元の画像と似たような性質を持つ画像を「無限に」生成することができる。

新技術の特徴・従来技術との比較

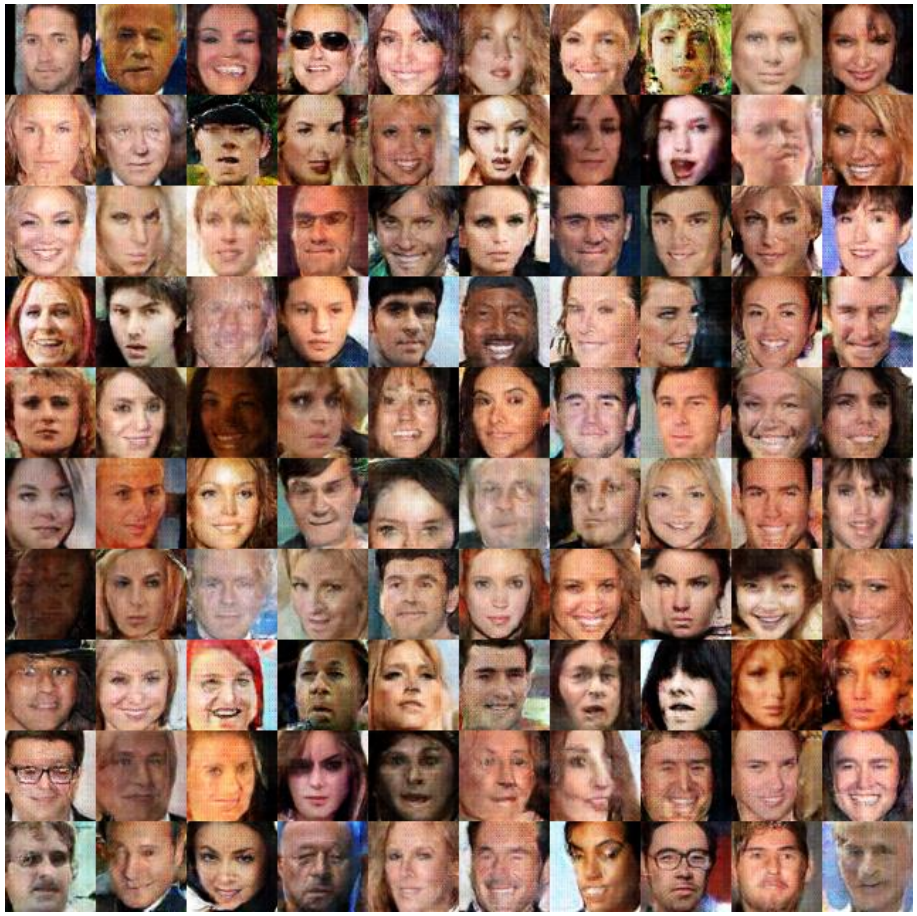


- 学習の目標
  - Generator: Discriminatorを騙すように画像を生成する。
  - Discriminator: 元の画像の分布に従っているか生成画像を評価する。

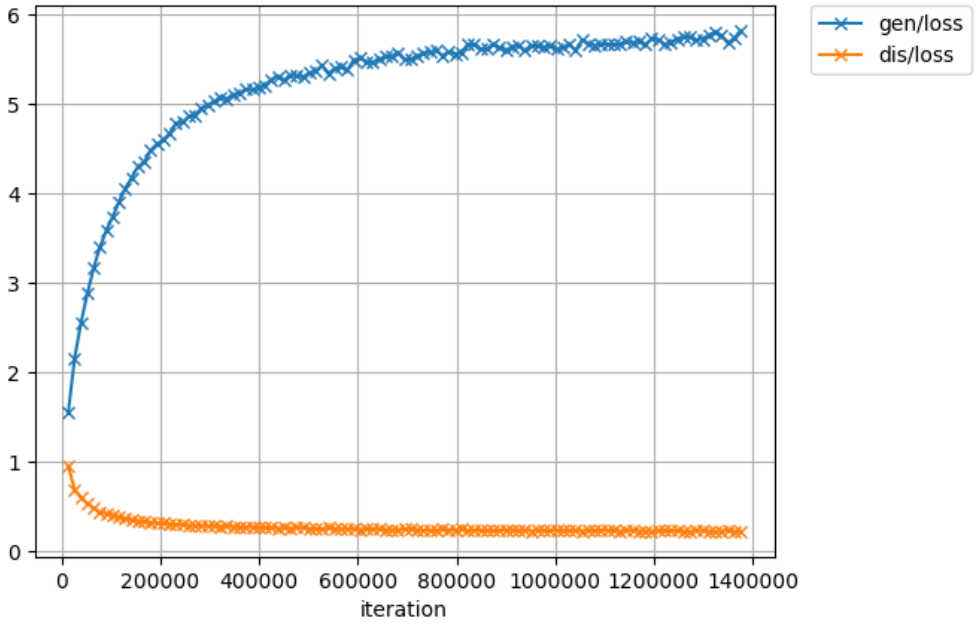
新技術の特徴・従来技術との比較

- 実施例：
  - The number of training images: 208,251.
  - Epochs for training: 500
  - Batch size: 81
  - Experimental environment:
  - OS: Ubuntu 18.04.1 LTS
  - Memory: 31.1 GiB
  - Processor: Intel® Core™ i7-7800X CPU @ 3.50GHz × 12
  - Graphic: GeForce GTX 1080/PCIe/SSE2
  - Programming Language: python (anaconda3-4.3.0)
  - GAN was trained using 1 GPU.

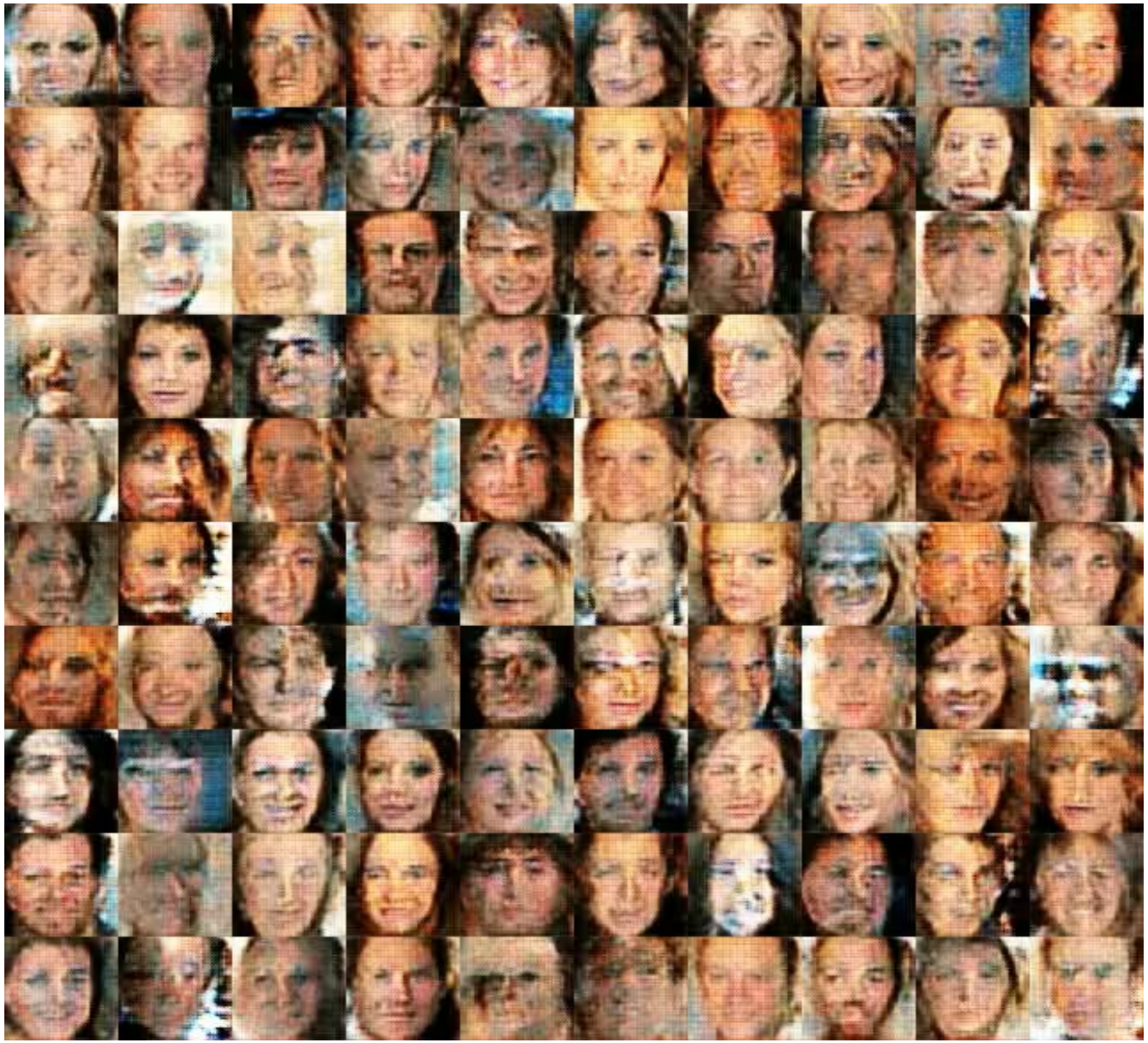
新技術の特徴・従来技術との比較



500エポックの結果



新技術の特徴・従来技術との比較



新技術の特徴・従来技術との比較

- GANに基づくカバー画像生成法がすでに提案されている。
- 文献1の問題
  - 送信側：不自然なカバー画像が生成された場合、そのままステゴ画像を作ると、秘密データの「存在」がばれてしまう。
  - 受信側：第三者からの「偽」ステゴ画像を受信した場合、そのままデータを抽出すると、「操られる」可能性がある。

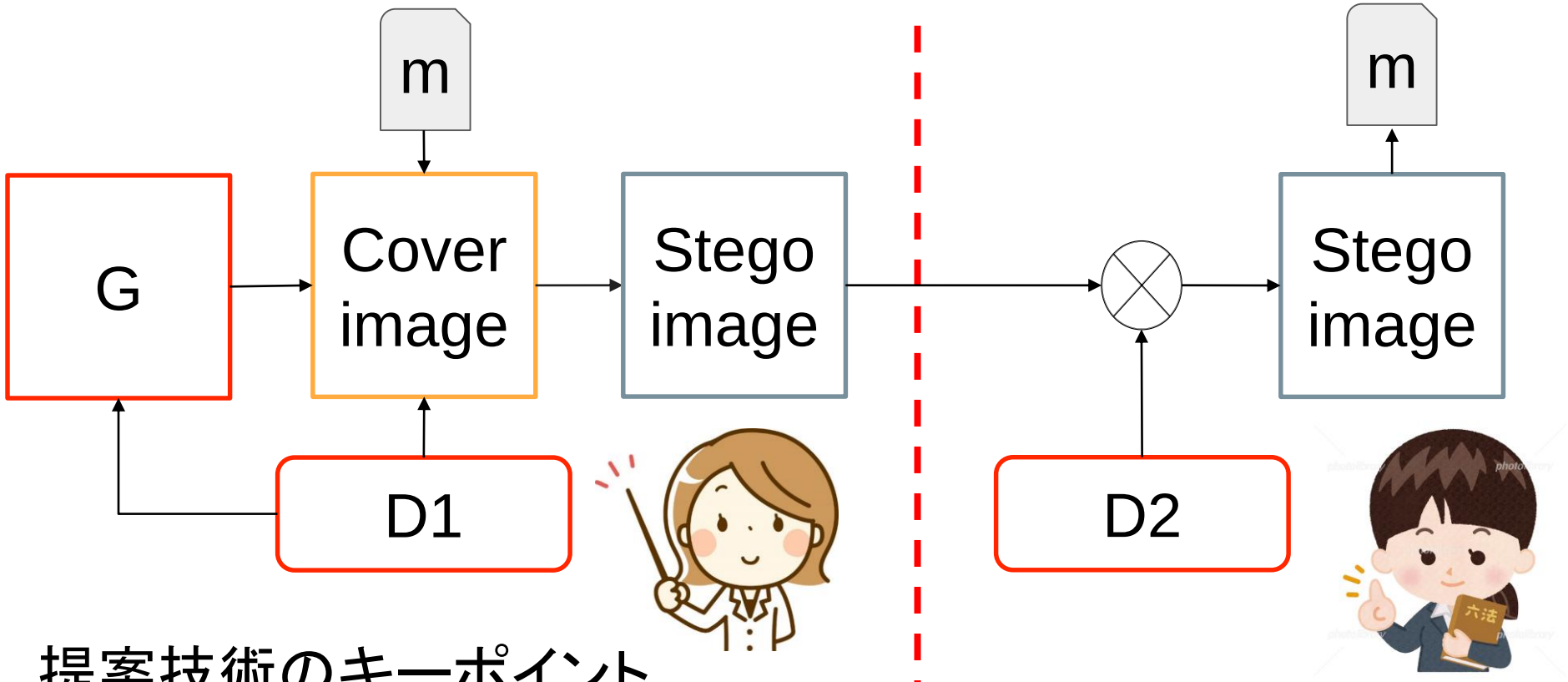
1. A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks, Donghui Hu et al. https://www.researchgate.net/publication/326192682_A_Novel_Image_Steganography_ Method_via_Deep_Convolutional_Generative_Adversarial_Networks/download

新技術の特徴・従来技術との比較

- 文献2の問題（追加問題）：
  - 「無数の秘密データ」の中で、任意のデータを「トリガー」として直接に「ステゴ画像」を生成すると、トリガーを回復できる保証はない。
  - $m \neq m'$

2. GENERATIVE ADVERSARIAL NETWORKS FOR IMAGE STEGANOGRAPHY, https://pdfs.semanticscholar.org/147b/7998526ebbdf64b1662503b378d9f6456ccd.pdf

新技術の特徴・従来技術との比較



- 提案技術のキーポイント
  - 送信側に、D1を利用して、生成画像の「自然さ」を評価する。
  - 受信側に、D2を利用して、ステゴ画像の「合法性」を評価する。
- 従来技術の問題点が解決できる。

想定される用途

- 本技術は秘密情報、個人情報を守るために有用である。
- 本技術と既存の暗号化技術と合わせて使用すると、より強固な情報セキュリティーが保証できる:
  - 暗号化:情報が読めなくなる。
  - ステガノグラフィー:情報の存在を見えなくなる。

実用化に向けた課題

- 現在、カバー画像の自然さやステゴ画像の合法性を評価するネットワーク(D1 and D2)の設計方法について検証中である。
  - GANで得られたDiscriminatorはそのまま使えないことが実験で確認した。
- 顔画像以外のカバー画像の生成についても実験して確認する必要がある。

企業への期待

- 本技術をアプリとして使用する場合、以下の課題を解決する必要がある:
  - User friendly GUI
  - Generator and discriminatorの実装・蒸留
  - 既存ファイルシステム・メールシステムへの組み込み方法

本技術に関する知的財産権

- 発明の名称　秘密データの通信方法、秘密データの通信プログラム及び秘密データの通信システム
- 出願番号　:特願2019-117078
- 出願人　:会津大学
- 発明者　:趙　強福、内藤　寛士

お問い合わせ先（必須）

会津大学
産学官連携コーディネーター　石橋　史朗

ＴＥＬ　０２４２－３７－２７７６
ＦＡＸ　０２４２－３７－２７７８
e-mail　ubic-adm@ubic-u-aizu.jp