UTokyo

PRESS RELEASE

東京大学

拡散モデルを用いた写真内の情報秘匿技術

発表のポイント

- ◆ 生成 AI を用いた画像の生成的コンテンツ置換（GCR）法を開発し、秘匿しつつ画像の見た目と内容の両方を維持する加工手法を開発しました。
- ◆ 新たに開発された手法では、画像の全体的な内容とプライバシーに関係しうる部分を特定した上で、拡散モデルを用いて代替画像を生成し、元の画像に適応させることで、プライバシーを守りつつ画像の視覚的魅力を保持する革新的なアプローチを提供します。
- ◆ 本手法は、SNS での画像共有、プレゼンテーション、ビジュアルデザインなど写真の視覚的美しさが重要な場面への応用が見込まれています。

本研究が提案する画像秘匿手法 GCR による秘匿加工例

概要

東京大学大学院工学系研究科電気系工学専攻の矢谷浩司准教授らのグループは、生成 AI を用いた画像の生成的コンテンツ置換（GCR）法を開発し、秘匿しつつ画像の見た目と内容の両方を維持する加工手法を開発しました。画像の秘匿化は、SNS の普及により重要性が増しています。従来の秘匿化手法にはモザイクやぼかしがありますが、これらの方法は手間がかかり、しばしば秘匿が不十分であるだけでなく、画像の見た目や統一感を損なう問題がありました。開発された手法は、画像全体とプライバシーに関連しうる部分の内容を表現するテキストを生成し、それらから拡散モデルにより代替画像を生成し、元の画像に配置することで、プライバシー保護と視覚的美しさのバランスを実現する画期的な方法となっています。SNS での画像共有やプレゼンテーション、ビジュアルデザインへの応用が期待されます。また、将来的には動画への応用や、より使用しやすいインタフェースの開発も進められています。

発表内容

画像の秘匿化は、画像の中に含まれているプライバシーに関係する情報を保護するために、大変重要な編集方法です。近年では画像のごく一部の情報から、撮影者の場所や属性が漏洩することが度々発生していますが、SNS などの急速な普及により、そのような危険性を十分に理解しないまま、画像が一般に公開されていることも数多くあります。秘匿化を実現する既存の編集方法としては、モザイクやぼかし、あるいは絵文字などを重ねる、などがありますが、多くの場合ではユーザが直接編集を施す必要があるため、多くの手間を要したり、秘匿が十分で

なかったりすることがあります。またこのような秘匿加工を施すと、元の画像から比べて見た目の美しさや統一感を損なってしまうことがあり、SNS などでの共有を目的とする場合には好ましくないこともあります。

この研究では、生成 AI 技術を用いてプライバシーに関係する情報を現実的な類似の代替物でシームレスに置き換える生成的コンテンツ置換（Generative Content Replacement, GCR）法を構築しました（図 1）。この方法では、ユーザが加工を行いたい画像をシステムにアップロードします。システムはアップロードされた画像に対して BLIP-2 モデル（注 1）を用い、画像全体の内容を表現するようなテキストを生成します。さらに、DIPA（注 2）と呼ばれる矢谷研究室が構築したデータセットにより提供されているマスク情報を用いて、画像内のプライバシーに関連しうる部分を抽出し、その部分の内容を表現するようなテキストを生成します。この 2 つのテキストをもとに、Stable diffusion（現在はバージョン 2.1 を使用）（注 3）して、画像内のプライバシーに関連しうる部分に類似した代替画像を生成し、元画像上に配置することで、コンテンツの置換を行います。これにより、もと画像にあったプライバシーに関連しうる情報は秘匿化されながらも、画像の見た目や内容を維持することが可能となります。

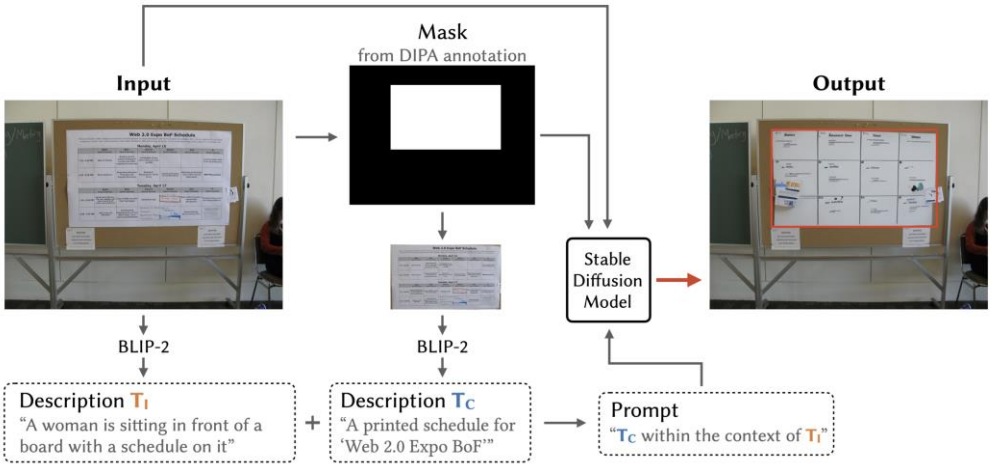

図 1：GCR の処理フロー

ユーザがアップロードした画像に対して、画像全体とプライバシーに関連しうる部分のテキストを生成し、それを元に画像内のプライバシーに関連しうる部分に類似した代替画像を生成し、元画像上に配置することで、コンテンツの置換を行います。

図 2 に示す通り、ぼかし、カートゥーニング（画像の一部を非現実的な程度に強調する方法）、色塗り、除去（画像内の物体等を消し去り、背景で置き換える）、GCR の 5 つを比較したユーザ実験の結果、GCR による秘匿加工では、画像内で加工が行われた場所を見つけ出すことが最も難しかったことが確認されました。また、他の秘匿加工手法と比較して、加工後の視覚的な調和が最も保たれていることも確認されました。元画像が持つストーリー性の維持に関しては、GCR はカートゥーニングよりも劣ったものの、プライバシー保護の強さにおいては GCR が秀でており、GCR による秘匿加工が、プライバシー保護と画像の視覚的美しさを両立しうる手法であることが確認されました。そのほか、GCR による秘匿加工の一例を図 3 に示します。

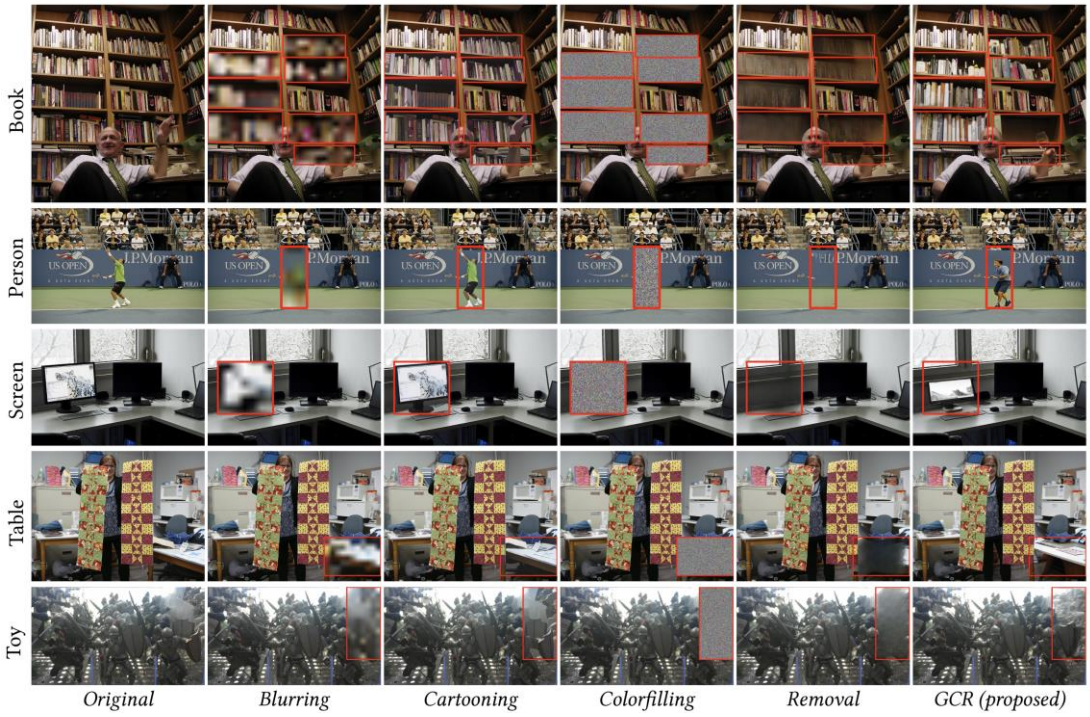

図 2：秘匿加工方法の比較

左から、元画像、ぼかし、カートゥーニング、色塗り、除去、GCR。

図 3：GCR による秘匿加工の一例

（上）後ろにいる男性を置換している。左が元画像、右が加工後の画像。（下）前面にある車は維持しつつ、背景にある車やナンバープレートを置換している。左が元画像、右が加工後の画像。

この研究成果は、画像のプライバシー保護と有用性の両方が求められる応用において、実用的な生成 AI の応用例を示すものです。SNS での画像共有のほか、プレゼンテーションやビジュアルデザインへの応用も期待されます。将来への展望として、研究室では、一般的なユーザがより簡単に GCR を使用できるインタフェースを構築しているほか、動画への応用を検討しています。

本研究は Microsoft Research Asia D-CORE Program、および株式会社メルカリ R4D とインクルーシブ工学連携研究機構との共同研究である価値交換工学の成果の一部です。

発表者・研究者等情報

東京大学大学院工学系研究科電気系工学専攻
矢谷　浩司　准教授

論文情報

雑誌名：Proccedings of the ACM Conference on Human Factors in Computing Systems (CHI 2024)
題　名：Examining Human Perception of Generative Content Replacement in Image Privacy Protection
著者名：Anran Xu*, Shitao Fang, Huan Yang, Simo Hosio, and Koji Yatani*

用語解説

（注 1）BLIP-2
与えられた画像から情報を抽出し、画像を説明するテキストを生成するマルチモーダル学習技術をベースに構築された人工知能技術。

（注 2）DIPA
矢谷研究室で構築した画像内においてプライバシーに関連しうる物体にアノテーションを施したデータセット。

（注 3）Stable Diffusion
拡散モデルと呼ばれる確率的プロセスを用い、テキストの記述に基づいて画像を生成する人工知能技術。

問合せ先

（研究内容については発表者にお問合せください）

東京大学大学院工学系研究科
准教授　矢谷　浩司（やたに　こうじ）

東京大学大学院工学系研究科 広報室