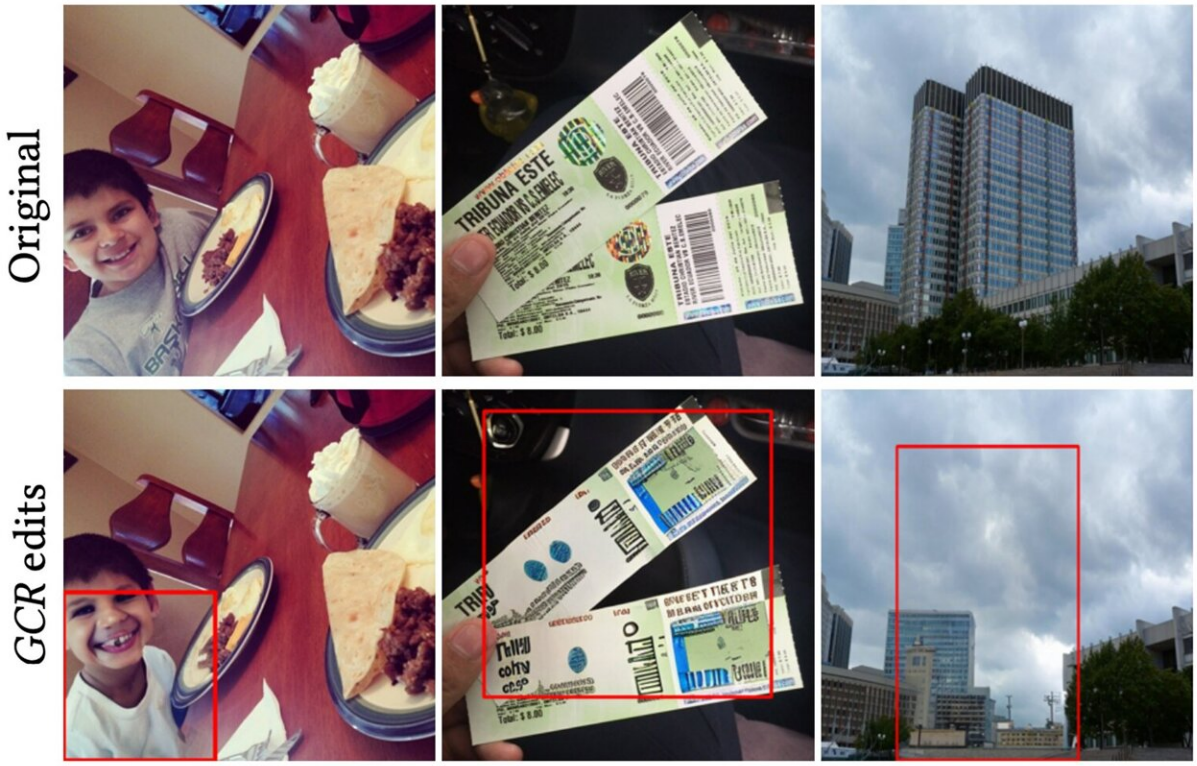
TechXplore

Hiding in plain sight: AI may help to replace confidential information in images with similar visuals

May 29 2024



The sections of these images outlined with a red box were annotated as privacy threatening with the use of an open source dataset called DIPA. GCR then used the annotated text prompts to replace the sections with visually similar or well-integrated substitutes. Credit: 2024 A. Xu, S. Fang, H. Yang et al./ Association for Computing Machinery

Image privacy could be protected with the use of generative artificial intelligence. Researchers from Japan, China and Finland created a system which replaces parts of images that might threaten confidentiality with visually similar but AI-generated alternatives.

Named "generative content replacement," in tests, 60% of viewers couldn't tell which images had been altered. The researchers intend for this system to provide a more visually cohesive option for image censoring, which helps to preserve the narrative of the image while protecting privacy.

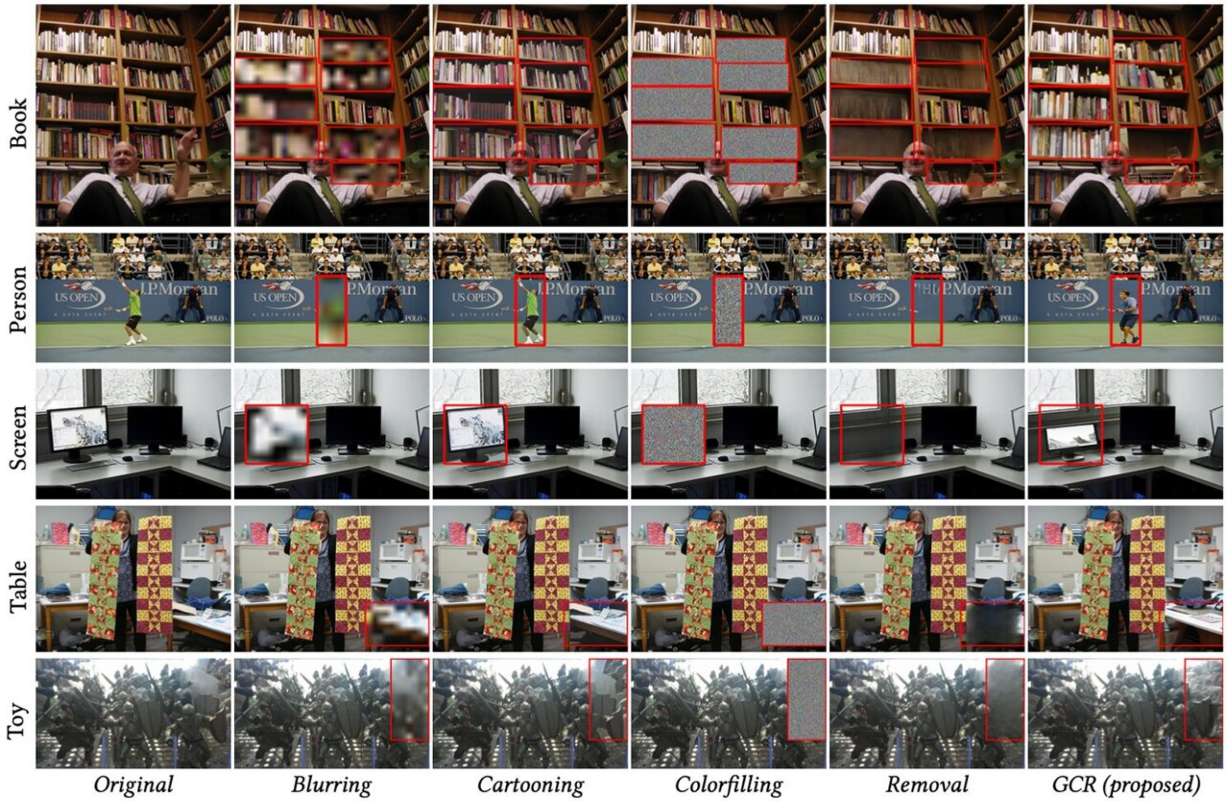
This research was presented at the Association for Computing Machinery's CHI Conference on Human Factors in Computing Systems, held in Honolulu, Hawaii, in the U.S., in May 2024.

With just a few text prompts, generative AI can offer a quick fix for a tricky school essay, a new business strategy or endless meme fodder. The advent of generative AI into daily life has been swift, and the potential scale of its role and influence are still being grappled with. Fears over its impact on future job security, online safety and creative originality have led to strikes from Hollywood writers, court cases over faked photos and heated discussions about authenticity.

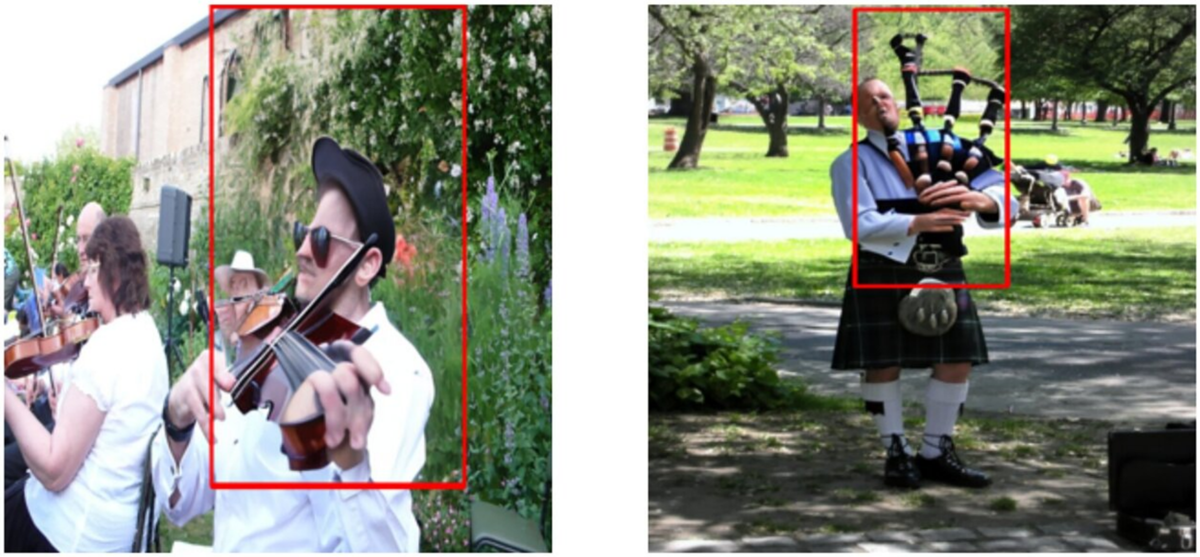
However, a team of researchers has proposed using a sometimes controversial feature of generative AI—its ability to manipulate images—as a way to solve privacy issues.

"We found that the existing image privacy protection techniques are not necessarily able to hide information while maintaining image aesthetics. Resulting images can sometimes appear unnatural or jarring. We considered this a demotivating factor for people who might otherwise consider applying privacy protection," explained Associate Professor Koji Yatani from the Graduate School of Engineering at the University of Tokyo.

"So, we decided to explore how we can achieve both—that is, robust privacy protection and image useability—at the same time by incorporating the latest generative AI technology."



Examples of popular methods for image content replacement and protection (outlined here by red boxes), and how they compare to GCR in the far-right column. Credit: 2024 A. Xu, S. Fang, H. Yang et al./ Association for Computing Machinery



As has been comically noted in many other AI-generated images shared online, for now GCR can also struggle to recreate realistic hands and facial features. Credit: 2024 A. Xu, S. Fang, H. Yang et al. / Association for Computing Machinery

The researchers created a computer system which they named generative content replacement (GCR). This tool identifies what might constitute a privacy threat and automatically replaces it with a realistic but artificially created substitute. For example, personal information on a ticket stub could be replaced with illegible letters, or a private building exchanged for a fake building or other landscape features.

"There are a number of commonly used image protection methods, such as blurring, color filling or just removing the affected part of the image. Compared to these, our results show that generative content replacement can better maintain the story of the original images and higher visual harmony," said Yatani. "We found that participants couldn't detect GCR in 60% of images."

For now, the GCR system requires a lot of computation resources, so it won't be available on any personal devices just yet. The tested system was fully automatic, but the team has since developed a new interface to allow users to customize images, giving more control over the final outcome.

Although some may be concerned about the risks of this type of realistic image alteration, where the lines between original and altered imagery become more ambiguous, the team is positive about its advantages.

"For public users, we believe that the greatest benefit of this research is providing a new option for image privacy protection," said Yatani. "GCR offers a novel method for protecting against privacy threats, while maintaining visual coherence for storytelling purposes and enabling people to more safely share their content."

More information: Anran Xu et al, Examining Human Perception of Generative Content Replacement in Image Privacy Protection, *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). DOI: 10.1145/3613904.3642103

Provided by University of Tokyo

Citation: Hiding in plain sight: AI may help to replace confidential information in images with similar visuals (2024, May 29) retrieved 20 June 2024 from https://techxplore.com/news/2024-05-plain-sight-ai-confidential-images.html