Privacy Challenges and Solutions for Image Data Sharing

Liyue Fan
Department of Computer Science
University of North Carolina at Charlotte
Email: liyue.fan@uncc.edu

***Abstract*—Sharing image data benefits a wide range of applications, including social media, medical imaging, and intelligent systems. Image data often contain sensitive information, the sharing of which may inflict individual privacy concerns. Traditional image privacy techniques, such as pixelization and blurring, do not provide effective protection. In this paper, we discuss privacy challenges and solutions for image data sharing. Specifically, we review existing solutions based on cryptography and federated learning, and discuss recent results on differential privacy in image domain. While differential privacy provides provable guarantees, we identify specific privacy challenges for image data and point out several considerations for future research.**

1. Introduction

Large amounts of image data are captured nowadays by smartphones, medical imaging equipment, and body-worn and surveillance cameras. Sharing image data would benefit various research communities and applications, advancing machine learning, training and education, as well as intelligent and sustainable systems. While sharing the image data is beneficial, individual *privacy concerns* would explicitly arise.In fact, traffic cameras capture the vehicle make, model, and license plate. Surveillance and body-worn cameras may archive persons and activities. Research research has shown that medical imaging data may leak sensitive information about patients.

The protection of private content in image data has been studied in computer vision and image processing. Traditional privacy enhancing techniques obscure the image by detecting and obfuscating the region of interest (ROI). However, such approaches cannot guarantee privacy. Recent studies [1], [2] have shown that deterministic obfuscation, such as pixelization and blurring, does not provide sufficient protection against inference attacks. Therefore, it is imperative to provide a rigorous guarantee of privacy when sharing image data with untrusted parties.

In this paper, we will review existing privacy solutions for image data sharing, discuss recent results in differential privacy, and identify challenges and future work directions.

2. Background

2.1. Private Image Computation

Computational solutions based on cryptography and federated learning have been developed to involve untrusted parties for image storage, sharing, and analysis. Those solutions have enabled private image sharing with intended recipients [3], private image retrieval [4], and training machine learning models [5], [6]. The key advantage of those solutions is that they do not directly share input image data with untrusted parties. However, they often result in computation and communication overheads. Furthermore, the computation results, such as intermediate updates and model predictions, can be used by adversaries to launch inference attacks [7], [8], [9].

2.2. Differential Privacy for Databases

Differential privacy has become the state-of-the-art privacy paradigm for quantifying privacy leakage in statistical databases. It assumes a trusted data curator is responsible for data aggregation and guarantees that an adversary is not able to decide whether a particular record is included or not in the input database, regardless of the amount of additional information available to the adversary. More formally, given any neighboring databases $\mathcal{D}$ and $\mathcal{D}'$ that differ by at most one record, a randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-differential privacy if for any $Z \subset range(\mathcal{M})$,

$$\Pr[\mathcal{M}(\mathrm{D}) \in \mathrm{Z}] \leq \mathrm{e}^{\epsilon} \cdot \Pr[\mathcal{M}(\mathrm{D}') \in \mathrm{Z}] + \delta. \qquad (1)$$

The parameters ϵ and δ specify the degree of privacy provided by the mechanism. Smaller ϵ and δ values indicate stronger privacy protection, and vice versa. A plethora of studies have applied differential privacy to data sharing, most notably in databases and data mining applications.

3. Differential Privacy for Image Data Sharing

To provide rigorous privacy in image data sharing, it is important to quantify the information leakage. In this section, we review recent results in differential privacy for image data and discuss their strengths and weaknesses.

3.1. Recent Results in Differential Privacy

Machine Learning. Deep learning methods have shown to achieve state-of-the-art results on computer vision problems. To achieve privacy in this setting, [10] proposed DP-SGD, which provides differential privacy guarantees to image samples used to train deep learning models. Specifically, the authors proposed sanitizing the gradients during neural network optimization, which ultimately limits the overall influence of any training example on the model. Privacy accountants [10], [11] have been proposed to account for differential privacy across training epochs, which provides stronger estimates of privacy loss compared to other composition theorems [12]. DP-SGD has been applied to image classification and medical segmentation tasks.

Individual-level Data Sharing. Another school of privacy solutions aim to enable individual-level data sharing, *without a trusted data curator*. Recently proposed solutions in this setting work to enhance image obfuscation, providing differential privacy guarantees to content within an input image. To provide a few examples, [13] was the first approach to adapt the notion of neighboring databases to the image domain and developed a differentially private pixelization method to protect m pixels simultaneously in the input image. [14] provides weaker differential privacy guarantees and protects individual pixels. [15] provides indistinguishability for SVD features. Such solutions have been validated with face and eye image data [16], [17] as well as in cloud-based machine learning tasks [18].

3.2. Strengths and Weaknesses of Differentially Private Image Data Sharing

From a computational perspective, the majority of differential privacy-based image data sharing solutions are relatively lightweight. This holds both for training machine learning models and for obfuscating individual-level images. From the privacy perspective, the differential privacy guarantee is resistant to post-processing, which means differentially private models and obfuscated images can be utilized in other analyses without inflicting additional differential privacy leakage. Empirically, recent results [18] show that face images obfuscated by DP-Pix [13] are difficult to re-identify even after denoising.

One important challenge in differentially private image data sharing has to do with a basic assumption of differential privacy: removing a record from the input database would be sufficient to protect its privacy. This assumption may hold in many database applications, e.g., where records are *i.i.d.* samples from a population. However, we may observe exceptions in many image applications. For example, many face images of the same individual may be included in the training set of a face recognition system. The differential privacy guarantee provided for each training sample may not equate to that of the participating individual. Furthermore, pixels within each image may not be *i.i.d.* either, e.g., adjacent pixels having similar values. See Figure 1 for an illustration. As a result, obfuscation approaches designed with weaker privacy guarantees, such as [14], may be less effective than intended.

(a) Input Image (b) Pixel Samples (c) Interpolation

Figure 1: Removing pixels from an input image may not protect privacy: (a) input image where the pixel samples are taken from; (b) the quantity of pixel samples is small and barely visible; (c) post-processing the pixel samples reconstructs the input image with high accuracy.

In addition, the choice of privacy parameters remains a big challenge in differentially private image data sharing. On one hand, ϵ and δ values indicate the level of privacy guarantees in differential privacy: the lower the better. On the other hand, there is an intrinsic trade-off between privacy protection and utility, where stronger privacy often leads to poorer utility. That has also been observed in differentially private image data sharing.

4. Considerations for Private Image Data Sharing

It is thus natural to ask the following. *How can we improve differentially private image data sharing? Is differential privacy the key to privacy challenges in image data sharing?* Needless to say, those are difficult questions. Below, we outline some first steps toward finding the answers.

4.1. Privacy and Utility

To overcome the challenges in privacy-protecting image data sharing, it is essential to characterize the privacy risks and the utility goals associated with the data and the application. In the machine learning context, membership inference [7] and model inversion [8] are highly relevant and application-specific utility, such as medical imaging and activities of daily living, should be evaluated. In the context of image obfuscation, e.g., with eye data, iris authentication is an appropriate risk measure and utility has been measured for pupil detection and gaze estimation [16]. Future computer vision research may uncover new privacy risks and develop more complex applications. In return, image privacy solution should take into account those results in the target image domain.

4.2. Usability

It is also important to take into account the usability of image privacy solutions to facilitate their adoption and future development. A large amount of image data shared nowadays are mainly for human consumption, e.g., in social

media. However, image privacy solutions may adversely affect viewers' satisfaction, which may cause people to avoid using them. Research studies that evaluate viewer experience regarding image privacy techniques [19], [20] may provide insights on enhancing viewers' satisfaction. In addition, future research should improve the experience of institutional users, e.g., data curators.

4.3. Fairness and Robustness

There are other important considerations that should be incorporated into the design of better image privacy solutions. It is known that differential privacy may amplify the unfairness of non-private models, e.g., in image classification [21], and could provide robustness against adversarial examples [22]. However, the fairness and robustness of other types of image privacy solutions, e.g., image obfuscation, have not been sufficiently studied. Furthermore, future research may consider addressing privacy, fairness, and robustness simultaneously for a target image domain.

References

[1] R. McPherson, R. Shokri, and V. Shmatikov, "Defeating image obfuscation with deep learning," *CoRR*, vol. abs/1609.00408, 2016.

[2] S. Hill, Z. Zhou, L. Saul, and H. Shacham, "On the (in) effectiveness of mosaicing and blurring as tools for document redaction," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 403–417, 2016.

[3] M.-R. Ra, R. Govindan, and A. Ortega, "P3: Toward privacy-preserving photo sharing," in *Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. Lombard, IL: USENIX, 2013, pp. 515–528. [Online]. Available: https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/ra

[4] L. Weng, L. Amsaleg, A. Morton, and S. Marchand-Maillet, "A privacy-preserving framework for large-scale content-based information retrieval," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 152–167, Jan 2015.

[5] R. Xu, J. B. Joshi, and C. Li, "Cryptonn: Training neural networks over encrypted data," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 2019, pp. 1199–1209.

[6] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[7] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE Symposium on Security and Privacy (SP)*, May 2017, pp. 3–18.

[8] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[9] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the gan: information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, 2017, pp. 603–618.

[10] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[11] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 2017, pp. 263–275.

[12] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[13] L. Fan, "Image pixelization with differential privacy," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2018, pp. 148–162.

[14] B. John, A. Liu, L. Xia, S. Koppal, and E. Jain, "Let it snow: Adding pixel noise to protect the user's identity," in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '20 Adjunct. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3379157.3390512

[15] L. Fan, "Practical image obfuscation with provable privacy," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 784–789.

[16] D. Reilly and L. Fan, "A comparative evaluation of differentially private image obfuscation," in *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. IEEE, 2021, pp. 80–89.

[17] M. U. Saleem, D. Reilly, and L. Fan, "Dp-shield: Face obfuscation with differential privacy." in *EDBT*, 2022, pp. 2–578.

[18] A. Parnami, M. Usama, L. Fan, and M. Lee, "Privacy enhancement for cloud-based few-shot learning," *arXiv preprint arXiv:2205.07864*, 2022.

[19] Y. Li, N. Vishwamitra, B. P. Knijnenburg, H. Hu, and K. Caine, "Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, pp. 1–24, 2017.

[20] R. Hasan, E. Hassan, Y. Li, K. Caine, D. J. Crandall, R. Hoyle, and A. Kapadia, "Viewer experience of obscuring scene elements in photos to enhance privacy," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–13.

[21] E. Bagdasaryan, O. Poursaeed, and V. Shmatikov, "Differential privacy has disparate impact on model accuracy," *Advances in neural information processing systems*, vol. 32, 2019.

[22] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, "Certified robustness to adversarial examples with differential privacy," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 656–672.