Privacy-Preserving Mechanisms for Multi-Label Image Recognition

HONGHUI XU, ZHIPENG CAI, and WEI LI, Georgia State University, Department of Computer Science

Multi-label image recognition has been an indispensable fundamental component for many real computer vision applications. However, a severe threat of privacy leakage in multi-label image recognition has been overlooked by existing studies. To fill this gap, two privacy-preserving models, Privacy-Preserving Multi-label Graph Convolutional Networks (P2-ML-GCN) and Robust P2-ML-GCN (RP2-ML-GCN), are developed in this article, where differential privacy mechanism is implemented on the model's outputs so as to defend black-box attack and avoid large aggregated noise simultaneously. In particular, a regularization term is exploited in the loss function of RP2-ML-GCN to increase the model prediction accuracy and robustness. After that, a proper differential privacy mechanism is designed with the intention of decreasing the bias of loss function in P2-ML-GCN and increasing prediction accuracy. Besides, we analyze that a bounded global sensitivity can mitigate excessive noise's side effect and obtain a performance improvement for multi-label image recognition in our models. Theoretical proof shows that our two models can guarantee differential privacy for model's outputs, weights and input features while preserving model robustness. Finally, comprehensive experiments are conducted to validate the advantages of our proposed models, including the implementation of differential privacy on model's outputs, the incorporation of regularization term into loss function, and the adoption of bounded global sensitivity for multi-label image recognition.

CCS Concepts: • **Security and privacy → Domain-specific security and privacy architectures**;

Additional Key Words and Phrases: Multi-label image recognition, differential privacy, robustness

ACM Reference format:
Honghui Xu, Zhipeng Cai, and Wei Li. 2022. Privacy-Preserving Mechanisms for Multi-Label Image Recognition. *ACM Trans. Knowl. Discov. Data.* 16, 4, Article 69 (January 2022), 21 pages.
https://doi.org/10.1145/3491231

1 INTRODUCTION

Multi-label image recognition is a fundamental component in computer vision applications [7], such as medical diagnosis recognition [12], human attribute recognition [19], and retail checkout recognition [13, 37]. With the rapid development of deep neural networks, the performance of multi-label image recognition is remarkably improved via deep learning models. However, due to the reliance on massive images uploaded to third-party platforms to accomplish multi-label image

This work was partly supported by the National Science Foundation of U.S. (1704287, 1829674, 1912753, and 2011845).
Authors' address: H. Xu, Z. Cai (corresponding author), and W. Li, Georgia State University, Department of Computer Science, Atlanta, 30302 GA; emails: hxu16@student.gsu.edu, {zcai, wli28}@gsu.edu.

1556-4681/2022/01-ART69 $15.00
https://doi.org/10.1145/3491231

recognition, these deep learning models may face a serious risk of privacy leakage [4]. For example, attackers can infer private information via extracted features and/or victim model's weights, causing substantial economic losses for individuals and institutions. More problematically, they even can launch attack mechanisms in black-box applications (APIs) by only utilizing the distribution of model's outputs [27, 33]. As multi-label image recognition plays a pivotally important role in many real applications, it becomes essential to guarantee privacy protection while maintaining prediction performance for the multi-label image recognition models.

Recently, researchers have realized the importance of privacy protection when designing deep neural networks in real applications. One vein of research is to hide sensitive visual information by integrating noise with images for data publishing to protect privacy [5, 6, 28, 34, 41, 43]. However, due to the lack of theoretical privacy guarantee, the performance of those methods heavily rely on discriminator. On the other hand, differential privacy mechanisms are adopted in many deep learning models [38, 47, 49] to theoretically achieve privacy guarantee for different goals, such as generation and determination. In these deep learning models, noise is usually employed to disturb the models' weights in order to keep the models' parameters secure [22, 39, 40, 42], or integrated into the models' input features so as to generate privacy-preserving data for public publishing [15, 25]. However, large aggregated noise brought by deep structure will result in low performance and poor model usability in real applications. Moreover, black-box attack, which can be easily implemented only using the model's outputs, is not considered in the existing works. The aforementioned observations motivated us to work out a solution to ensure privacy protection, maintain prediction accuracy, alleviate the aggregated noise's side effect, and defend black-box attack simultaneously for the multi-label image recognition models.

In this article, we propose P2-ML-GCN mechanism that satisfies ϵ-differential privacy on the outputs of **Multi-label Graph Convolutional Networks (ML-GCN)** [7] with the intention of preventing black-box attack. To further increase the prediction accuracy of P2-ML-GCN, we develop RP2-ML-GCN, where a regularization term is designed to enhance the model's robustness, and the global sensitivity in differential privacy mechanism is smoothed via a proper bound to mitigate excessive noise's side effect. In other words, we can enhance the prediction accuracy using a regularization term and/or a bounded global sensitivity, which pioneers a new research direction for effectively designing privacy-preserving deep learning algorithms. Moreover, through rigorous theoretical analysis, we prove the guarantee of privacy protection for ML-GCN, the effectiveness of our proposed regularization term for robustness improvement, the advantage of utilizing a bounded global sensitivity to alleviate excessive noise's side effect, and the capability of our proposed models to protect the privacy of model's weights and input features. Finally, we evaluate the performance of our proposed models by conducting intensive real-data experiments and comparing them with the-state-of-the-art models. Our multifold contributions are addressed as follows.

- To the best of our knowledge, this is the first work to design privacy-preserving multi-label image recognition models based on differential privacy mechanism.
- Our first model P2-ML-GCN applies differential privacy mechanism on the model's outputs, which can defend black-box attack and avoid large aggregated noise even if a neural network has many layers.
- To improve the prediction accuracy of P2-ML-GCN, a regularization term is designed in our second model RP2-ML-GCN to enhance the model's robustness, and a proper bound of global sensitivity in differential privacy mechanism is set to alleviate the side effect of excessive noise.

— Through rigorous theoretical analysis, we prove that our two proposed models are able to protect the privacy of the model's outputs, weights and input features with the guarantee of ϵ-differential privacy, which provides a guidance for the design of privacy-preserving deep learning algorithms.
— Comprehensive experiments are well-conducted to validate the advantages of P2-ML-GCN and RP2-ML-GCN.

The rest of this article is organized as follows. Related works are briefly summarized in Section 2. After introducing preliminaries in Section 3, we detail our models in Section 4. In Section 5, we conduct real-data experiments and analyze all results. Finally, we end up with a conclusion in Section 6.

2 RELATED WORKS

The state-of-the-art about multi-label image recognition and differential privacy-based machine learning algorithms is summarized in the following.

2.1 Multi-label Image Recognition

A straightforward idea of multi-label recognition is to train independent binary classifiers for each object label based on state-of-the-art deep **Convolutional Neural Networks** (**CNNs**) [16, 30, 32], which, however, ignores the relationship among labels. To improve the efficiency of multi-label image recognition models, the label correlation is taken into account in some works [7, 35, 36, 48]. Wang et al. considered the correlation of labels through employing **Recurrent Neural Networks** (**RNNs**) in embedded label vectors [35]. Zhu et al. studied both semantic and spatial relations of multiple labels to design a spatial regularization network based on weighted attention maps [48]. Wang et al. proposed a spatial transformer layer and **Long-Short Term Memory** (**LSTM**) units to capture label correlation [36]. Recently, Chen et al. proposed a GCN-based ML-GCN model, which applies the directed graph of multiple object labels built by labels' co-occurrence pattern in dataset [7]. So far, the method of [7] outperforms other existing methods. However, the study about how to design a privacy-preserving model for such multi-label image recognition has been overlooked by the existing works.

2.2 Differential Privacy in Deep Learning

Differential privacy mechanism was proposed by Dwork et al. for privacy guarantee on adjacent databases [8]. The incorporation of differential privacy mechanisms and deep learning algorithms in most of the existing works can be briefly divided into two categories. One is to update weights in **stochastic gradient descent** (**SGD**) algorithms with additional noise calculated by the gradient bound [1, 22, 39, 40, 42], or to update weights in regression models with additional noise calculated by the polynomial coefficient of the regression models' parameters [24], which mainly focuses on the parameters of learning models to satisfy differential privacy requirements. The other is to obtain a privacy-preserving generative model by employing a proper noise, which keeps an eye on input features [3, 15, 25, 45]. But, when the number of input features and the number of shared parameters are large, these existing works sacrifice a high privacy budget to maintain models' accuracy. In addition, since differential privacy mechanisms are implemented on either weights or features in every layer of deep learning models, these existing works may suffer from a large aggregated noise when a neural network contains too many layers. Moreover, even if these works can obtain secure weights and features, they cannot resist black-box attack that can be accomplished based on the distribution of models' outputs [27, 29].

In this article, in order to defend black-box attack and protect privacy for multi-label image recognition, we propose two novel models, including P2-ML-GCN and RP2-ML-GCN, by implementing differential privacy mechanisms on ML-GCN's outputs. Compared with the state-of-the-art, our models have three major advantages: (i) the noise added into outputs can be bounded even if the neural network has many layers, which can significantly reduce the aggregated noise of an entire model and thus provide a higher degree of privacy guarantee; (ii) the two proposed models can prevent the aforementioned black-box attack because the noise disturbs the distribution of outputs; and (iii) in RP2-ML-GCN, a regularization item based on the Frobenius norm of weights of classifiers is added to the loss function for the performance improvement, and a bound of global sensitivity in differential privacy mechanisms is set appropriately to mitigate the excessive noise's side effect in P2-ML-GCN. Finally, we rigorously prove that our proposed mechanisms can provide a helpful guidance for the design of privacy-preserving deep learning algorithms.

3 PRELIMINARIES

In this section, we introduce **graph convolutional network** (**GCN**), ML-GCN model for multi-label image recognition [7], and the basics of differential privacy [8].

3.1 Graph Convolutional Network

GCN was introduced in [18] to perform semi-supervised graph classification aiming to update the node representations of a graph by convolutional operations. The two inputs of GCN include the node feature matrix in the lth layer $H^l \in \mathbb{R}^{n\times d}$ and the node correlation matrix $A \in \mathbb{R}^{n\times n}$, where n denotes the number of nodes in a graph and d is the dimension of node features in lth layer. After employing the convolutional operations of [18], the node feature matrix $H^{l+1} \in \mathbb{R}^{n\times d'}$ in the $(l+1)$th layer can be represented as $H^{l+1} = h(\hat{A}H^lW^l)$, where $h(\cdot)$ denotes a non-linear operation, $\hat{A} \in \mathbb{R}^{n\times n}$ is the normalized version of correlation matrix A, and $W^l \in \mathbb{R}^{d\times d'}$ is a transformation matrix to be learned.

3.2 ML-GCN

By taking the label correlation into account, ML-GCN outperforms other existing approaches in multi-label image recognition [7] and thus is adopted as our baseline. In [7], a directed graph is built on all images of a dataset, where the vertices represent object labels, and the weight of a directed edge is the occurrence probability of a head vertex when its corresponding tail vertex occurs. The directed graph is used to mine co-occurrence patterns of object labels within the dataset through GCN. The image features can be extracted by Resnet-101 [14]. Then, the co-occurrence pattern can be combined with features to improve the performance of multi-labels recognition.

Let C be the number of labels' categories, D be the dimension of features, and $\hat{y} \in \mathbb{R}^C$ be the output prediction labels. We can obtain $\hat{y}$ via Equation (1).

$$\hat{y} = Wx, \tag{1}$$

where $W \in \mathbb{R}^{C\times D}$ is the final parameter matrix after GCN has been trained, and $x \in \mathbb{R}^D$ is the feature vector extracted by Resnet-101.

Finally, ML-GCN is trained with the following multi-label classification loss function.

$$L = \sum_{i=1}^{C} y_i \log(\sigma(\hat{y}_i)) + (1 - y_i) \log(1 - \sigma(\hat{y}_i)), \tag{2}$$

where $y_i \in \{0, 1\}$ is the real label of ith category, $\hat{y}_i \in [0, 1]$ is the confidence score of ith category, and $\sigma(\cdot)$ is the sigmoid function [44].

3.3 Differential Privacy

Differential privacy defines a mathematical measurement of data privacy protection for a dataset [8].

Definition 1. A randomized mechanism, $\mathcal{M}$ $(U \rightarrow \mathbb{R})$, satisfies ϵ-differential privacy, if for any two adjacent inputs $u, u' \in U$ and any $S \subset \mathbb{R}$, there is

$$\Pr[\mathcal{M}(u) \in S] \leq e^{\epsilon} \Pr\left[\mathcal{M}\left(u'\right) \in S\right], \tag{3}$$

where ϵ is a positive real number and quantifies information leakage.

To achieve ϵ-differential privacy, $\mathcal{M}$ can be constructed by a Laplace mechanism based on any real-value function f.

With respect to f, the global sensitivity S_f is defined as the maximum absolute distance between any two adjacent inputs in U [17, 21, 31], i.e.,

$$S_f = \sup_{u, u' \in U} |f(u) - f(u')|_1. \tag{4}$$

The randomized mechanism, $\mathcal{M}$, which satisfies ϵ-differential privacy for function f, can be obtained via additive Laplace noise as follows.

$$\mathcal{M}(u) = f(u) + Lap(0, S_f/\epsilon), \tag{5}$$

in which $Lap(0, S_f/\epsilon)$ is the Laplace distribution.

4 PROPOSED APPROACHES

In this section, we elaborate on the details of our proposed models, including P2-ML-GCN and RP2-ML-GCN. In P2-ML-GCN, to achieve privacy-preserving multi-label image recognition, we apply differential privacy mechanism to ML-GCN's prediction outputs based on additive Laplace noise. Notice that the prediction accuracy of P2-ML-GCN may be reduced due to the added additive Laplace noise. Hence, to further improve the image recognition performance, we propose RP2-ML-GCN that enhances the model's robustness with the help of a regularization term. Moreover, we analyze the relationship of privacy guarantee between our proposed models that implement differential privacy mechanisms on the prediction outputs and the approaches that adopt differential privacy mechanisms on input features or parameters, which confirms the effectiveness of our proposed models. Finally, we extend our findings to a more general case to offer a guidance for the design of privacy-preserving deep learning approaches. Since it is hard to show all analysis of multi-layer neural network with limited page length, in this article, we mainly focus on analyzing the bias of loss function and the performance of differential privacy for model weights and features in a single layer perceptron.

4.1 Privacy-Preserving ML-GCN

In P2-ML-GCN, we implement differential privacy mechanism on ML-GCN's prediction output vector, $\hat{y}$, in order to make the model's outputs satisfy ϵ-differential privacy, in which Laplace noise is utilized to disturb ML-GCN's outputs instead of its input features or parameters to resist black-box attack. According to Laplace mechanism, there are two steps to establish a randomized mechanism satisfying ϵ-differential privacy. First, we denote the global sensitivity of $\hat{y}$ as $S_{\hat{y}}$. Second, from Equation (5), we can obtain a randomized mechanism $\hat{y}'$ that satisfies ϵ-differential privacy by adding the Laplace noise $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$ to the output vector $\hat{y}$ as shown in Equation (6), where $\hat{y} \in \mathbb{R}^C$ and α generated from $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$ are C-dimension vectors.

$$\hat{y}' = \hat{y} + \alpha. \tag{6}$$

Theorem 1. *Given the Laplace noise* $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$ *added into the output vector* $\hat{y}$*, each element* $\hat{y_i}'$ *in the disturbed output vector* $\hat{y}'$ *satisfies* ϵ*-differential privacy.*

Proof. Let $\Pr[\cdot]$ be a commonly designed Laplace distribution [9]. Accordingly, we have,

$$\ln \frac{\Pr[\hat{y_i}]}{\Pr[\hat{y_i}']} = \ln \frac{\frac{\epsilon}{2S_{\hat{y_i}}} e^{-\frac{\epsilon}{S_{\hat{y_i}}}|\hat{y_i}|}}{\frac{\epsilon}{2S_{\hat{y_i}}} e^{-\frac{\epsilon}{S_{\hat{y_i}}}|\hat{y_i}'|}} = \frac{\epsilon}{S_{\hat{y_i}}}(|\hat{y_i}'| - |\hat{y_i}|) \leq \epsilon. \tag{7}$$

Equation (7) shows that each element $\hat{y_i}'$ in the disturbed output vector $\hat{y}'$ satisfies ϵ-differential privacy. □

Theorem 1 demonstrates that our proposed model P2-ML-GCN can provide the multi-label image recognition with differential privacy guarantee. Correspondingly, the loss function of multi-label image recognition in P2-ML-GCN can be expressed by the disturbed output vector $\hat{y}'$ in Equation (8).

$$\begin{aligned} L_{P2} &= \sum_{i=1}^{C} y_i \log(\sigma(\hat{y}')) + (1 - y_i) \log(1 - \sigma(\hat{y}')) \\ &= \sum_{i=1}^{C} y_i \log\left(\sigma\left(\hat{y_i} + Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right)\right) \\ &\quad + (1 - y_i) \log\left(1 - \sigma\left(\hat{y_i} + Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right)\right). \end{aligned} \tag{8}$$

During the training process of P2-ML-GCN, we intend to minimize L_{P2} to improve the prediction accuracy of ML-GCN while ensuring ϵ-differential privacy. In addition, we can control the privacy protection degree by adjusting the value of ϵ. Particularly, a smaller ϵ indicates a higher privacy protection degree.

4.2 Robust Privacy-Preserving ML-GCN

The noise added in P2-ML-GCN indeed offers differential privacy guarantee, but may also reduce the prediction accuracy of ML-GCN. Therefore, we design a more robust model, RP2-ML-GCN, to alleviate the influence on the prediction accuracy of multi-label image recognition while gaining the same degree of differential privacy guarantee. Specifically, in RP2-ML-GCN, we integrate the loss function of P2-ML-GCN with a regularization term to increase the prediction accuracy of ML-GCN.

There are three phases in RP2-ML-GCN. (i) In the first phase, we simplify the traditional multi-label loss function for better theoretical analysis. (ii) In the second phase, we calculate the bias of loss function to analyze the influence of the additive Laplace noise on the prediction accuracy of multi-label image recognition model. (iii) In the third phase, we theoretically prove that the regularization term can improve the model's robustness from the viewpoint of linear regression.

4.2.1 Function Simplification. Since the sigmoid function is differentiable at the point 0, we can obtain an approximate quadratic polynomial in Equation (9) through Taylor Theorem [26] at the point 0.

$$log(1 + e^{-\hat{y_i}}) \approx \log 2 - \frac{1}{2}\hat{y_i} + \frac{1}{8}(\hat{y_i})^2. \tag{9}$$

Then, we simplify the traditional multi-label loss function via the sigmoid function and its approximate quadratic polynomial function. The simplification process of traditional multi-label loss

function is presented as follows:

$$\begin{aligned} L &= \sum_{i=1}^{C} y_i \log(\sigma(\hat{y_i})) + (1-y_i)\log(1-\sigma(\hat{y_i})) \\ &= \sum_{i=1}^{C} -y_i \log(1+e^{-\hat{y_i}}) + (1-y_i)(\log(e^{-\hat{y_i}})) \\ &\quad + (y_i-1)\log(1+e^{-\hat{y_i}}) \\ &= \sum_{i=1}^{C} y_i\hat{y_i} - \hat{y_i} - \log(1+e^{-\hat{y_i}}) \\ &\approx \sum_{i=1}^{C} y_i\hat{y_i} - \hat{y_i} - \left(\log 2 - \frac{1}{2}\hat{y_i} + \frac{1}{8}(\hat{y_i})^2\right) \\ &= \sum_{i=1}^{C} -\frac{1}{8}(\hat{y_i})^2 - \frac{1}{2}\hat{y_i} + y_i\hat{y_i} - \log 2. \end{aligned} \tag{10}$$

By substituting Equation (1) into Equation (10), we obtain Equation (11).

$$\begin{aligned} L &= \sum_{i=1}^{C} -\frac{1}{8}(\hat{y_i})^2 - \frac{1}{2}\hat{y_i} + y_i\hat{y_i} - \log 2 \\ &= -\frac{1}{8}(Wx)^T(Wx) - \left(\frac{1}{2} - y_i\right)(Wx) - C\log 2, \end{aligned} \tag{11}$$

where W is the parameter matrix learned by GCN, x is feature vector extracted by Resnet-101, $y_i \in \{0,1\}$ is the groundtruth label of ith category, and C is the number of categories.

4.2.2 Bias Analysis. According to Equation (11), we can rewrite the loss function of P2-ML-GCN in Equation (12).

$$\begin{aligned} L_\alpha = &-\frac{1}{8}((Wx+\alpha)^T(Wx+\alpha)) \\ &-\left(\frac{1}{2} - y_i\right)(Wx+\alpha) - C\log 2. \end{aligned} \tag{12}$$

In the analysis of machine learning algorithms, the bias of loss function is typically used to investigate the influence of the additive noise on the prediction accuracy.

Lemma 1. *The expectation of Laplace noise* $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$ *is*

$$\mathbb{E}\left(Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right) = 0.$$

Lemma 2. *The expectation of square Laplace noise* $\mathbb{E}(Lap(0, \frac{S_{\hat{y}}}{\epsilon})^2)$ *is equal to*

$$\mathbb{E}\left(Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right) + \mathrm{Var}\left(Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right) = \frac{2S_{\hat{y}}}{\epsilon^2}.$$

According to Lemma 1, Lemma 2, Equation (11), and Equation (12), the bias of loss function, denoted by $\mathbb{E}(\Delta L)$, can be calculated via Equation (13).

$$\begin{aligned}
\mathbb{E}(\Delta L) &= \mathbb{E}(|L_\alpha - L|) \\
&= \mathbb{E}\left(\left|-\frac{1}{8}\alpha^T W x - \frac{1}{8}x^T W^T \alpha - \frac{1}{8}\alpha^T \alpha - \left(\frac{1}{2} - y_i\right)\alpha\right|\right) \\
&= \left|-\frac{1}{8}\mathbb{E}(\alpha^T)\mathbb{E}(Wx) - \frac{1}{8}\mathbb{E}(x^T W^T)\mathbb{E}(\alpha) - \frac{1}{8}\mathbb{E}(\alpha^T \alpha) - \left(\frac{1}{2} - y_i\right)\mathbb{E}(\alpha)\right| \\
&= \left|-\frac{1}{8}\mathbb{E}\left(Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)^2\right)\right| \\
&= \left|-\frac{1}{8} \times \frac{2S_{\hat{y}}}{\epsilon^2}\right| \\
&= \left|-\frac{S_{\hat{y}}}{4\epsilon^2}\right| \\
&= \frac{S_{\hat{y}}}{4\epsilon^2}.
\end{aligned} \tag{13}$$

From the expression of $\mathbb{E}(\Delta L)$, we can see that there exists an inverse proportion between $\mathbb{E}(\Delta L)$ and ϵ; that is, the smaller ϵ is, the greater $\mathbb{E}(\Delta L)$ is. In other words, a higher privacy protection degree reduces the prediction accuracy of the multi-label recognition model.

In order to alleviate the side-effect of additive Laplace noise, weight decay mechanism [46] inspires us to increase the prediction accuracy by reducing $W^T W$ for the purse of improving P2-ML-GCN's robustness. Accordingly, we propose our model RP2-ML-GCN by integrating P2-ML-GCN's loss function with a regularization term as shown in Equation (14).

$$\begin{aligned}
L_{RP2} = \sum_{i=1}^{C} & y_i \log\left(\sigma\left(\hat{y}_i + Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right)\right) \\
& + (1 - y_i)\log\left(1 - \sigma\left(\hat{y}_i + Lap\left(0, \frac{S_{\hat{y}}}{\epsilon}\right)\right)\right) + \lambda ||W||_2^F,
\end{aligned} \tag{14}$$

where λ is a hyperparameter to control the weight of the regularization term, and the Forbenius norm $||W||_2^F$ is equal to the value of $W^T W$.

During the training process in P2-ML-GCN, we accomplish image recognition with privacy guarantee by minimizing L_{RP2} and improve the model's robustness by minimizing $W^T W$. Notably, in fact, the regularization term can improve the robustness of the traditional ML-GCN model even without additional noise.

4.2.3 Robustness Analysis. In the following, we theoretically investigate how the regularization term can improve P2-ML-GCN's robustness from two aspects. On the one hand, the regularization term helps shrink the space of weights so as to avoid overfitting. On the other hand, the utilization of the regularization term can reduce the variance of weights.

The training process of P2-ML-GCN can be treated as a generalized linear regression without regularization, while the training process of RP2-ML-GCN can be treated as a generalized ridge regression with regularization. Let W_{LR} and W_{Ridge} denote the weight matrixes trained in P2-ML-GCN and RP2-ML-GCN, respectively, which can be computed in Equation (15) and Equation (16), respectively.

$$W_{LR} = \underset{\boldsymbol{W}}{\operatorname{argmin}} ||\hat{y}' - Wx||^2. \tag{15}$$

$$W_{Ridge} = \underset{\boldsymbol{W}}{\operatorname{argmin}} ||\hat{y}' - Wx||^2 + \lambda ||W||^2. \tag{16}$$

Assume that x is centralized and standardized, and xx^T is reversible. We can obtain two estimators, *i.e.*, $\hat{W}_{LR}$ for W_{LR} and $\hat{W}_{Ridge}$ for W_{Ridge}.

$$\hat{W}_{LR} = \hat{y}' x^T (xx^T)^{-1}. \tag{17}$$

$$\begin{aligned}
\hat{W}_{Ridge} &= \hat{y}' x^T (xx^T + \lambda \mathbf{I})^{-1} \\
&= \hat{y}' x^T (xx^T)^{-1} (xx^T)(xx^T + \lambda \mathbf{I})^{-1} \\
&= \hat{W}_{LR} (xx^T)(xx^T + \lambda \mathbf{I})^{-1} \\
&= \hat{W}_{LR} (xx^T + \lambda \mathbf{I} - \lambda \mathbf{I})(xx^T + \lambda \mathbf{I})^{-1} \\
&= \hat{W}_{LR} (\mathbf{I} - \lambda (xx^T + \lambda \mathbf{I})^{-1}) \\
&\leq \hat{W}_{LR}.
\end{aligned} \tag{18}$$

Remark: From Equation (17) and Equation (18), $\hat{W}_{Ridge}$ can be considered as the shrinkage of $\hat{W}_{LR}$, achieving weight decay to avoid overfitting.

Lemma 3. *If $\hat{V}$ is the unbiased estimator of any one random variable V, $\mathbb{E}(\hat{V}) = V$.*

Lemma 4. *Three numerical characteristics in matrix theory are shown as follows:*

$$\begin{aligned}
&\mathbb{E}(o^T G o) = (\mathbb{E}(o))^T G \mathbb{E}(o) + tr(G \operatorname{Var}(o)), \\
&tr(EFG) = tr(FEG) = tr(GEF), \\
&tr(G^T) = tr(G),
\end{aligned}$$

where o is white noise, and E, F, and G represent any matrix.

Furthermore, to demonstrate that the regularization term indeed improves the model's robustness, we need to prove that the variance of $\hat{W}_{Ridge}$ is lower than the variance of $\hat{W}_{LR}$. Let $\hat{y}' = Wx + o$, where o is the white noise following $\mathcal{N}(0, \sigma^2)$. We rewrite $\hat{W}_{LR}$ in Equation (19).

$$\begin{aligned}
\hat{W}_{LR} &= \hat{y}' x^T (xx^T)^{-1} \\
&= (Wx + o) x^T (xx^T)^{-1} \\
&= W_{LR} + o x^T (xx^T)^{-1}.
\end{aligned} \tag{19}$$

Since $\hat{W}_{LR}$ is an unbiased estimator, the variance of $\hat{W}_{LR}$ can be calculated using Lemma 3 and Lemma 4 as follows:

$$\begin{aligned}
\operatorname{Var}(\hat{W}_{LR}) &= \mathbb{E}(\hat{W}_{LR} - \mathbb{E}(\hat{W}_{LR}))^2 \\
&= \mathbb{E}(\hat{W}_{LR} - W_{LR})^2 \\
&= \mathbb{E}[(\hat{W}_{LR} - W_{LR})^T (\hat{W}_{LR} - W_{LR})] \\
&= \mathbb{E}[(o x^T (xx^T)^{-1})^T o x^T (xx^T)^{-1}] \\
&= \mathbb{E}[((xx^T)^{-1})^T x o^T o x^T (xx^T)^{-1}] \\
&= \sigma^2 tr(((xx^T)^{-1})^T x x^T (xx^T)^{-1}) \\
&= \sigma^2 tr[((xx^T)^{-1})]^T \\
&= \sigma^2 tr((xx^T)^{-1}) \\
&= \sigma^2.
\end{aligned} \tag{20}$$

Similarly, the variance of $\hat{W}_{Ridge}$ can be calculated by:

$$\begin{aligned} \mathrm{Var}(\hat{W}_{Ridge}) &= \sigma^2 \left[\sum_{i=1}^{\mathcal{K}} \frac{k_i}{(k_i+\lambda)^2} \right] \\ &= \left[\sum_{i=1}^{\mathcal{K}} \frac{k_i}{(k_i+\lambda)^2} \right] \mathrm{Var}(\hat{W}_{LR}) \\ &= z\,\mathrm{Var}(\hat{W}_{LR}), \end{aligned} \tag{21}$$

where $\mathcal{K}$ is the rank of xx^T, $(k_1, k_2, \ldots, k_{\mathcal{K}})$ is the set of eigenvalues of xx^T, and $z = [\sum_{i=1}^{\mathcal{K}} \frac{k_i}{(k_i+\lambda)^2}]$ denotes variance expansion factor.

Remark: The variance of $\hat{W}_{Ridge}$ can be lower than $\hat{W}_{LR}$ by adjusting λ. On the other hand, variance expansion factor z becomes smaller when the value of λ is increased, which further reduces the variance of $\hat{W}_{Ridge}$. Therefore, a conclusion can be drawn that RP2-ML-GCN indeed improves the robustness of P2-ML-GCN by adding the regularization term from the viewpoint of the linear regression.

4.3 Bound of Global Sensitivity

To improve the prediction accuracy of P2-ML-GCN, there are two methods: one is to enhance the model's robustness, and the other is to decrease excessive noise added into the prediction outputs. A regularization term in RP2-ML-GCN can improve the model's robustness. In this subsection, we show that an appropriate bound of global sensitivity in differential privacy mechanisms can alleviate excessive noise's side effect. Before introducing our method, we present a critical observation as follows.

Observation 1. *Most existing analyses on differential privacy mechanisms assume that the maximum contribution (*i.e., *the global sensitivity of query function) is fixed in advance. However, we may end up adding excessive noise for privacy protection due to some outliers in database, resulting in the reduction of prediction accuracy of learning models. Therefore, a bound of global sensitivity of query function can be set to mitigate the side effect of excessive noise, which can improve the model performance [2].*

According to Observation 1, the calculation of global sensitivity, $S_{\hat{y}}$, in P2-ML-GCN is affected by the imbalanced distribution of outputs, causing excessive noise. Inspired by the idea of [2], we set a bound factor, denoted by $S_b \in (0, 1)$, to mitigate excessive noise's side effect for the improvement of P2-ML-GCN's accuracy.

In the following, we reimplement differential privacy mechanism with a bounded global sensitivity to see how it works to improve P2-ML-GCN's accuracy. First, we substitute $S_{\hat{y}}$ with $S_b S_{\hat{y}}$. According to Theorem 1, the disturbed output function satisfies $\frac{\epsilon}{S_b}$-differential privacy that is called relaxed-differential privacy in this article because $S_b \in (0, 1)$. Second, we rewrite the bias of loss function by replacing $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$ with $Lap(0, \frac{S_b S_{\hat{y}}}{\epsilon})$ in Equation (13), which is shown in Equation (22).

$$\mathbb{E}(\Delta L) = \left| -\frac{1}{8} \times \frac{2S_b}{\epsilon^2} \right| = \frac{S_b S_{\hat{y}}}{4\epsilon^2}. \tag{22}$$

Equation (22) implies that we can indeed decrease the bias of loss function in P2-ML-GCN by reducing the value of S_b and thus improve the prediction accuracy of P2-ML-GCN.

Remark: To guarantee relaxed-differential privacy and improve prediction accuracy simultaneously, we can select an appropriate bound for the global sensitivity in P2-ML-GCN and RP2-ML-GCN based on the specific distribution of outputs to alleviate excessive noise's side effect.

4.4 Model Effectiveness

As aforementioned in Section 2, prior differential privacy-based privacy-preserving deep learning approaches either protect the model's weights or input features. Different from the state-of-the-art, in our proposed models, privacy-preserving mechanisms are applied to protect the model's outputs, which can prevent black-box attack. In this subsection, we theoretically prove that our proposed models are also able to ensure ϵ-differential privacy for the model's weights and input features.

4.4.1 Effectiveness for Weights' Differential Privacy. Since the outputs of ML-GCN are calculated by both the weights and the input features, the noise added into outputs will reflect on the weights of classifiers and features through a backward propagation training process. In order to find out how the disturbed output vector $\hat{y}'$ influences the parameter matrix, the feature vector x is supposed to be fixed. We can rewrite the disturbed output vector with the disturbed parameter matrix, denoted by W_α, in Equation (23).

$$\hat{y}' = \hat{y} + \alpha = W_\alpha x, \tag{23}$$

where $\hat{y}$ is the original output vector, and α is the additional Laplace noise used in differential privacy mechanism.

Let $\gamma_1 = \max\{|x_i^{-1}|\}$ with x_i^{-1} being the ith element in vector x^{-1}, where each element in x^{-1} is the reciprocal of the corresponding element in x. Then, we can obtain the inequality in Equation (24).

$$\begin{aligned} W_\alpha &= (\hat{y} + \alpha)x^{-1} = \hat{y}x^{-1} + \alpha x^{-1} \\ &= W + \alpha x^{-1} \leq W + \gamma_1 \alpha. \end{aligned} \tag{24}$$

Let $\overline{W}$ be the maximum value of elements in W and $\overline{W_\alpha}$ be the maximum value of elements in W_α, and $\max\{\alpha\}$ be the maximum value of elements in α. According to Equation (24), we have $\overline{W_\alpha} = \overline{W} + \gamma_1 \max\{\alpha\}$.

Theorem 2. *If the disturbed output vector $\hat{y}'$ satisfies ϵ-differential privacy, the disturbed parameter matrix W_α satisfies $\frac{\epsilon(\overline{W}+\max\{|x_i^{-1}|\}\max\{\alpha\})}{\max\{|x_i^{-1}|\}^2}$-differential privacy.*

Proof. $\Pr[\cdot]$ is commonly designed as Laplace distribution. Since α follows $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$, the additional Laplace noise can be designed as $\gamma_1\alpha$, which follows $Lap(0, \frac{\gamma_1^2 S_{\hat{y}}}{\epsilon})$, for the disturbed weight matrix W_α according to Equation (24). Thus, we have

$$\begin{aligned} \ln \frac{\Pr[W]}{\Pr[W_\alpha]} &= \ln \frac{\frac{\epsilon}{2\gamma_1^2 S_{\hat{y}}} e^{-\frac{\epsilon}{\gamma_1^2 S_{\hat{y}}}|W|}}{\frac{\epsilon}{2\gamma_1^2 S_{\hat{y}}} e^{-\frac{\epsilon}{\gamma_1^2 S_{\hat{y}}}|W_\alpha|}} \\ &= \frac{\epsilon}{\gamma_1^2 S_{\hat{y}}}(|W_\alpha| - |W|) \leq \frac{\epsilon(\overline{W} + \gamma_1 \max\{\alpha\})}{\gamma_1^2} \\ &= \frac{\epsilon(\overline{W} + \max\{|x_i^{-1}|\}\max\{\alpha\})}{\max\{|x_i^{-1}|\}^2}. \end{aligned} \tag{25}$$

That is, we can prove that the disturbed weight matrix W_α satisfies $\frac{\epsilon(\overline{W}+\max\{|x_i^{-1}|\}\max\{\alpha\})}{\max\{|x_i^{-1}|\}^2}$-differential privacy. □

4.4.2 Effectiveness for Features' Differential Privacy. Similarly, in order to find out how the disturbed output vector $\hat{y}'$ influences the feature vector, we assume that the parameter matrix W is fixed. The disturbed output vector is rewritten with the disturbed feature vector, denoted by x_α, in Equation (26).

$$\hat{y}' = \hat{y} + \alpha = W x_\alpha. \tag{26}$$

Let $\gamma_2 = \max\{|W_{ij}^{-1}|\}$ where W_{ij}^{-1} is the element in ith row and jth column in matrix W^{-1}. We can obtain the inequality in Equation (27).

$$\begin{aligned} x_\alpha &= W^{-1}(\hat{y} + \alpha) = W^{-1}\hat{y} + W^{-1}\alpha \\ &= x + W^{-1}\alpha \le x + \gamma_2 \alpha. \end{aligned} \tag{27}$$

Let $\overline{x}$ be the maximum value of elements in x and $\overline{x}_\alpha$ be the maximum value of elements in x_α. From Equation (27), there is $\overline{x}_\alpha = \overline{x} + \gamma_2 \max\{\alpha\}$.

Theorem 3. *If the disturbed output vector $\hat{y}'$ satisfies ϵ-differential privacy, the disturbed feature vector x_α satisfies $\frac{\epsilon(\overline{x}+\max\{|W_{ij}^{-1}|\}\max\{\alpha\}))}{\max\{|W_{ij}^{-1}|\}^2}$-differential privacy.*

Proof. $\Pr[\cdot]$ is commonly designed as Laplace distribution. Since α follows $Lap(0, \frac{S_{\hat{y}}}{\epsilon})$, the additional Laplace noise can be designed as $\gamma_2\alpha$, which follows $Lap(0, \frac{\gamma_2^2 S_{\hat{y}}}{\epsilon})$, for the disturbed feature vector x_α according to Equation (27). Then we can prove that the disturbed feature vector x_α satisfies $\frac{\epsilon(\overline{x}+\max\{|W_{ij}^{-1}|\}\max\{\alpha\}))}{\max\{|W_{ij}^{-1}|\}^2}$-differential privacy as follows:

$$\begin{aligned} \ln \frac{\Pr[x]}{\Pr[x_\alpha]} &= \ln \frac{\frac{\epsilon}{2\gamma_2^2 S_{\hat{y}}} e^{-\frac{\epsilon}{\gamma_2^2 S_{\hat{y}}}|x|}}{\frac{\epsilon}{2\gamma_2^2 S_{\hat{y}}} e^{-\frac{\epsilon}{\gamma_2^2 S_{\hat{y}}}|x_\alpha|}} \\ &= \frac{\epsilon}{\gamma_2^2}(|x_\alpha| - |x|) \le \frac{\epsilon(\overline{x} + \gamma_2 \max\{\alpha\})}{\gamma_2^2} \\ &= \frac{\epsilon(\overline{x} + \max\{|W_{ij}^{-1}|\}\max\{\alpha\}))}{\max\{|W_{ij}^{-1}|\}^2}. \end{aligned} \tag{28}$$

□

Remark: As analyzed in priors works [1, 42], $\overline{W}$ and $\overline{x}$ are finite values. Although in our proposed models, we implement differential privacy mechanism on the model's outputs, Theorem 2 and Theorem 3 show the effectiveness of our models to achieve ϵ-differential privacy for model's weights and input features, which provides a new direction to perform differential privacy in deep learning algorithms.

4.5 Model Generalization

To further illustrate that our proposed models can achieve any degree of differential privacy for model's weights or input features, we extend our theoretical analysis to a more general scenario, in which two corollaries can be directly derived from Theorem 2 and Theorem 3.

Let $\mathcal{P}$ and $\gamma_\mathcal{P}$ be two finite values representing two scale parameters for the design of differential privacy on the model's weights. In the training process of the multi-label image recognition model, we set $\gamma_\mathcal{P} \ge \max\{|x_i^{-1}|\}$ by controlling the feature extractor first. Then, we set $\overline{W_\alpha} \le \gamma_\mathcal{P}^2 \mathcal{P}$ when updating parameters. Accordingly, we can obtain Corollary 1.

Corollary 1. *If the disturbed output vector $\hat{y}'$ satisfies ϵ-differential privacy with $\gamma_{\mathcal{P}} \geq \max\{|x_i^{-1}|\}$ and $\overline{W_\alpha} \leq \gamma_{\mathcal{P}}^2 \mathcal{P}$, the disturbed weight matrix W_α satisfies $\epsilon\mathcal{P}$-differential privacy.*

Proof. If $\gamma_{\mathcal{P}} \geq \max\{|x_i^{-1}|\}$, $Lap(0, \frac{\gamma_{\mathcal{P}}^2}{\epsilon})$ can be considered as the additional Laplace noise to disturb weight matrix W according to Equation (24). If $\overline{W_\alpha} \leq \gamma_{\mathcal{P}}^2 \mathcal{P}$, we can prove that the disturbed weight matrix W_α satisfies $\epsilon\mathcal{P}$-differential privacy as below:

$$\begin{aligned} \ln \frac{\Pr[W]}{\Pr[W_\alpha]} &= \ln \frac{\frac{\epsilon}{2\gamma_{\mathcal{P}}^2} e^{-\frac{\epsilon}{\gamma_{\mathcal{P}}^2}|W|}}{\frac{\epsilon}{2\gamma_{\mathcal{P}}^2} e^{-\frac{\epsilon}{\gamma_{\mathcal{P}}^2}|W_\alpha|}} \\ &= \frac{\epsilon}{\gamma_{\mathcal{P}}^2}(|W_\alpha| - |W|) \leq \frac{\epsilon\gamma_{\mathcal{P}}^2\mathcal{P}}{\gamma_{\mathcal{P}}^2} \\ &= \epsilon\mathcal{P}. \end{aligned} \tag{29}$$

□

Similarly, we use two finite values, $\mathcal{Q}$ and $\gamma_{\mathcal{Q}}$, to denote two scale parameters for the design of differential privacy on model's input features. In the training process, we set $\gamma_{\mathcal{Q}} \geq \max\{|W_{ij}^{-1}|\}$ when updating parameters and set $\overline{x_\alpha} \leq \gamma_{\mathcal{Q}}^2 \mathcal{Q}$ when extracting features. Then, we can obtain Corollary 2.

Corollary 2. *If the disturbed output vector $\hat{y}'$ satisfies ϵ-differential privacy with $\gamma_{\mathcal{Q}} \geq \max\{|W_{ij}^{-1}|\}$ and $\overline{x_\alpha} \leq \gamma_{\mathcal{Q}}^2 \mathcal{Q}$, the disturbed feature vector x_α satisfies $\epsilon\mathcal{Q}$-differential privacy.*

Proof. If $\gamma_{\mathcal{Q}} \geq \max\{|W_{ij}^{-1}|\}$, $Lap(0, \frac{\gamma_{\mathcal{Q}}^2}{\epsilon})$ can be treated as the additional Laplace noise to disturb feature vector x according to Equation (27). If $\overline{x_\alpha} \leq \gamma_{\mathcal{Q}}^2 \mathcal{Q}$, the disturbed feature vector x_α satisfies $\epsilon\mathcal{Q}$-differential privacy, which is proved below.

$$\begin{aligned} \ln \frac{\Pr[x]}{\Pr[x_\alpha]} &= \ln \frac{\frac{\epsilon}{2\gamma_{\mathcal{Q}}^2} e^{-\frac{\epsilon}{\gamma_{\mathcal{Q}}^2}|x|}}{\frac{\epsilon}{2\gamma_{\mathcal{Q}}^2} e^{-\frac{\epsilon}{\gamma_{\mathcal{Q}}^2}|x_\alpha|}} \\ &= \frac{\epsilon}{\gamma_{\mathcal{Q}}^2}(|x_\alpha| - |x|) \leq \frac{\epsilon\gamma_{\mathcal{Q}}^2\mathcal{Q}}{\gamma_{\mathcal{Q}}^2} \\ &= \epsilon\mathcal{Q}. \end{aligned} \tag{30}$$

□

Remark: Since $\mathcal{P}$ and $\mathcal{Q}$ can be any finite real number, we can successfully protect the model's weights and input features with any degree of differential privacy by implementing ϵ-differential privacy mechanisms on the model's outputs in our proposed models. Moreover, Corollary 1 and Corollary 2 provide a guidance for the design of privacy-preserving deep learning algorithms in a general scenario.

5 EXPERIMENTS

In this section, comprehensive experiments are conducted to validate that our two proposed models, P2-ML-GCN and RP2-ML-GCN, can effectively accomplish multi-label image recognition while guaranteeing ϵ-differential privacy; especially, compared with P2-ML-GCN, RP2-ML-GCN can increase prediction accuracy. Besides, experiments are set up to confirm that our proposed regularization term indeed improves the performance of ML-GCN model even without Laplace noise.

Moreover, we investigate the advantage of setting a proper bound of global sensitivity to increase the accuracy of P2-ML-GCN by fine-tuning different values of S_b. Finally, the effectiveness of our proposed models is further evaluated through a comparison with the state-of-the-art.

5.1 Experiment Settings

The datasets, performance metrics, and mechanism implementation in our experiments are described below. Our implementation codes can be found in https://github.com/ahahnut/-R-P2-ML-GCN.

5.1.1 Datasets. We report the experimental results on two benchmark multi-label image recognition datasets, including Voc2007 [10] and MS-COCO [20]. Notice that there are 20 categories of images in Voc2007 (*i.e.,* $C = 20$) and 80 categories of images in MS-COCO (*i.e.,* $C = 80$). According to the definition of global sensitivity in Equation (4), the global sensitivity $S_{\hat{y}}$ is set as 20 when we use Voc2007 in our experiments and is set as 80 when we use MS-COCO in our experiments.

In machine learning training, the number of proper epochs for different datasets are different. According to the state-of-the-art, we train Voc2007 dataset with 40 epochs and defaultly train MS-COCO dataset with 20 epochs. Similarly, we need to set different values of ϵ to make sure ϵ-differential privacy guarantee when training different datasets. The value of ϵ, which is so-called "privacy budget", indicates the degree of privacy protection. More specifically, a smaller ϵ implies a higher privacy protection degree.

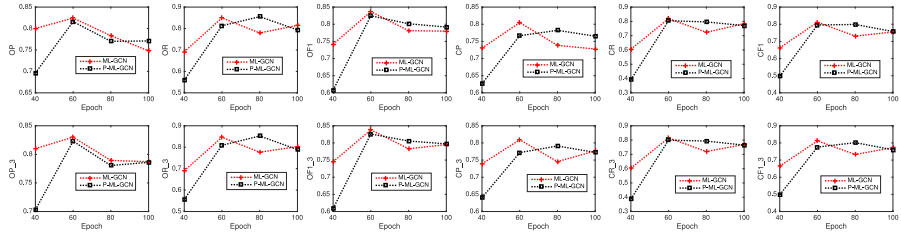
5.1.2 Performance Metrics. Typically, the average per-class precision (CP), recall (CR), F1 (CF1), the average overall precision (OP), recall (OR), F1 (OF1), and mean average precision (mAP) are adopted to quantify prediction performance [11, 35, 48]. For a fair comparison, the prediction performance of top-3 labels is also evaluated using the above performance metrics [11, 48] represented by (∗_3), where ∗ could be OP, OR, OF1, CP, CR, or CF1.

5.1.3 Mechanism Implementation. For clear performance evaluation, ML-GCN is adopted as the baseline model in our experiments. Our proposed models, including P2-ML-GCN and RP2-ML-GCN, are implemented according to the instructions of ML-GCN [7]. There are four main steps in our experiments:

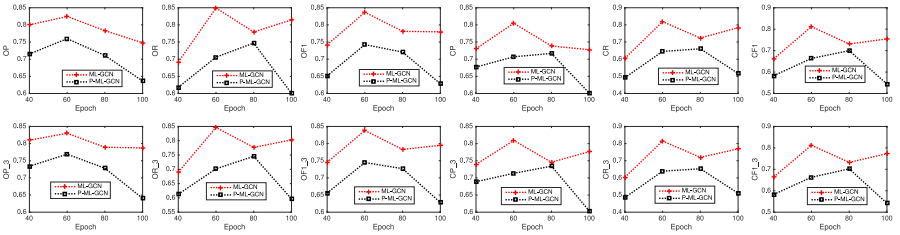
(1) The dimensions of output features in two GCN layers are 1,024 and 2,048, respectively.
(2) Label representations in GCN are adopted for training on Wikipedia dataset [23].
(3) Resnet-101 [14] is utilized to extract features of images resized into 448×448.
(4) The parameter ϵ in Laplace noise is used to adjust the degree of privacy protection for privacy-preserving training.

5.2 Evaluation of Privacy Preservation

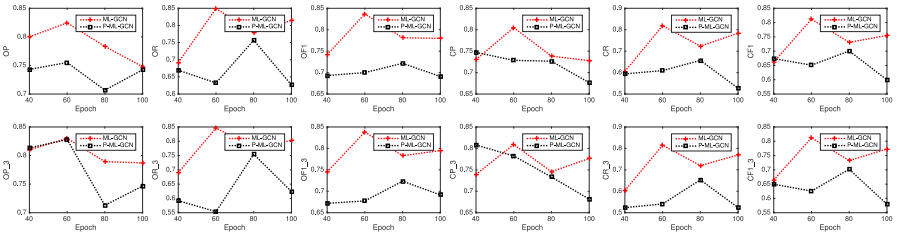
We implement our model P2-ML-GCN with $\epsilon = 8, 10, 30$ in the experiments, which is reasonable and applicable in real applications according to the scenario of our studied problem and the setting of ϵ in previous works [1, 3, 40, 45]. To illustrate the feasibility of P2-ML-GCN, the results on Voc2007 dataset with 40, 60, 80, and 100 epochs are presented in Figure 1, and the results on MS-COCO dataset with 10, 15, 20, and 25 epochs are presented in Figure 2. In Figures 1 and 2, obviously, P2-ML-GCN can achieve different degrees of ϵ-differential privacy guarantee by adjusting the values of ϵ; especially, a lower ϵ indicates a higher degree of privacy protection. In specific, take the OP value of P2-ML-GCN in Figure 1 as an example. For P2-ML-GCN on Voc2007 dataset with 40 epochs, $OP = 0.6963$ in P2-ML-GCN with $\epsilon = 8$, $OP = 0.7152$ in P2-ML-GCN with $\epsilon = 10$, and $OP = 0.7431$ in P2-ML-GCN with $\epsilon = 30$. By comparing these OP values, we can find



(a) $\epsilon = 8$



(b) $\epsilon = 10$



(c) $\epsilon = 30$

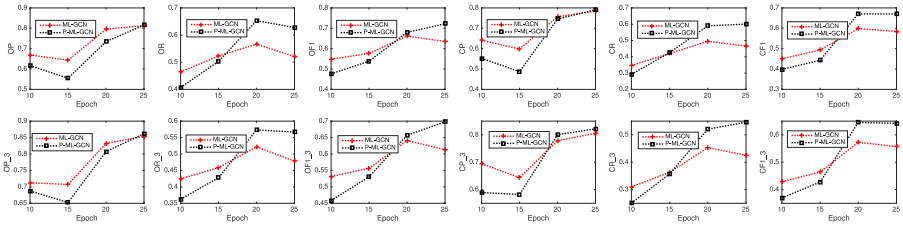
Fig. 1. P2-ML-GCN v.s. ML-GCN on Voc2007 dataset with different ϵ.

that the increase of the added Laplace noise does not cause too much decrease of prediction performance of P2-ML-GCN. The same conclusion can be obtained by comparing other performance metrics on Voc2007 dataset in Figure 1. Besides, we can also get the same conclusion by comparing all performance metrics on MS-COCO dataset in Figure 2. To sum up, compared with ML-GCN, the prediction performance of P2-ML-GCN does not suffer a lot with the increase of the added Laplace noise, which indicates that P2-ML-GCN can maintain the performance of multi-label image recognition while providing ϵ-differential privacy guarantee. These results demonstrate the effectiveness of P2-ML-GCN for privacy protection as analyzed in Section 4.1.

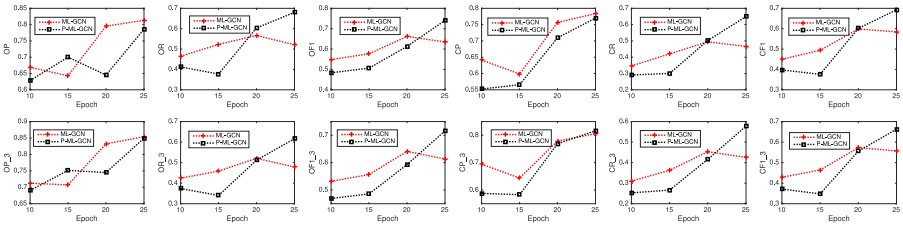
5.3 Ablation Study

We analyze that the scale parameter λ can be used to reduce the variance of loss function in Section 4.2.3, and the bounded global sensitivity $S_b S_{\hat{y}}$ can be used to decrease the bias of loss function in Section 4.3. Thus, in order to validate the analysis of the regularization term and the bounded global sensitivity, we present the following experiment results.

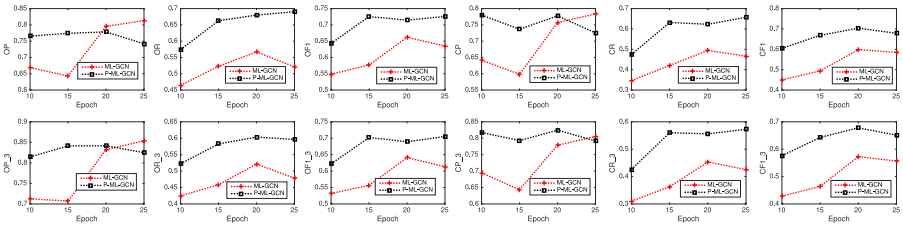
In Figure 3(a), RP2-ML-GCN is trained on Voc2007 dataset with 40 epochs by fixing $\epsilon = 10$ and varying λ from 0.1 to 0.9 with 0.2 step size. We change the value of λ that represents the weight of the regularization term to observe its impact on the prediction performance in RP2-ML-GCN. Specifically, we use the OP value as an example to analyze this impact. In Figure 3(a), $OP = 0.3835$ when $\lambda = 0.1$, $OP = 0.6022$ when $\lambda = 0.3$, $OP = 0.7793$ when $\lambda = 0.5$, $OP = 0.7702$ when $\lambda = 0.7$,



(a) $\epsilon = 8$

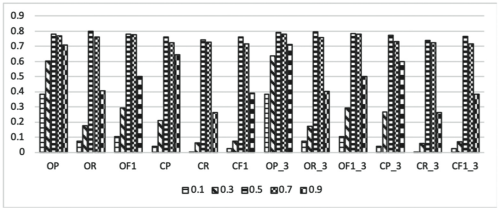


(b) $\epsilon = 10$

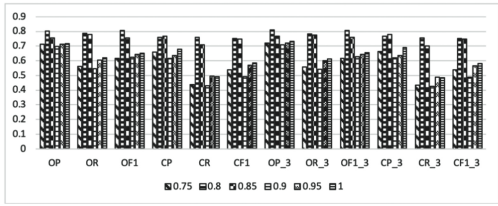


(c) $\epsilon = 30$

Fig. 2. P2-ML-GCN v.s. ML-GCN on MS-COCO dataset with different ϵ.



(a) RP2-ML-GCN on Voc2007 Dataset with Different λ (Epoch:40; $\epsilon = 10$)

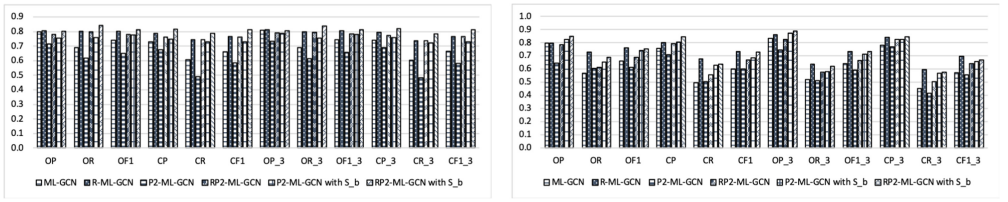


(b) P2-ML-GCN with S_b on Voc2007 Dataset (Epoch:40; $\epsilon = 10$; Different S_b)

Fig. 3. Ablation study.

and $OP = 0.7083$ when $\lambda = 0.9$. From these OP values, one can find that the OP value of RP2-ML-GCN can be highly improved when λ is increased from 0.1 to 0.5 but gradually decreases when λ is increased from 0.5 to 0.9. The same trend can also be observed by comparing other performance metrics. These phenomenons illustrate that the regularization term can be used to reduce the variance of loss function by adjusting the scale parameter λ as mentioned in Section 4.2.3. We will use $\lambda = 0.5$ when implementing RP2-ML-GCN model in the following experiments.

In Figure 3(b), P2-ML-GCN is trained on Voc2007 dataset with 40 epochs by fixing $\epsilon = 10$ and changing S_b from 0.75 to 1. The results show that with the same ϵ, different values of S_b can



(a) Evaluation Results on Voc2007 Dataset (Epoch: 40; $\epsilon = 10$; $\lambda = 0.5$; $S_b = 0.8$)

(b) Evaluation Results on MS-COCO Dataset (Epoch: 20; $\epsilon = 10$; $\lambda = 0.5$; $S_b = 0.8$)

Fig. 4. Comparison results of six models.

indeed affect the performance of P2-ML-GCN. From Figure 3(b), we observe that $OP = 0.7134$ when $S_b = 0.75$, $OP = 0.8032$ when $S_b = 0.8$, $OP = 0.7555$ when $S_b = 0.85$, $OP = 0.6988$ when $S_b = 0.9$, $OP = 0.7131$ when $S_b = 0.95$, and $OP = 0.7152$ when $S_b = 1$. By comparing these OP values, a proper bound factor (*i.e.,* $S_b = 0.8$ in our experiments) can be used to improve the OP value of the original P2-ML-GCN. For the other performance metrics in Figure 3(b), the utilization of the proper bound factor $S_b = 0.8$ can also improve other performance metrics of the original P2-ML-GCN. In other words, a proper bound can indeed help enhance P2-ML-GCN's prediction performance while ensuring relaxed-differential privacy, which confirms the advantage of using a proper bounded global sensitivity to increase P2-ML-GCN's accuracy. More concretely, Theorem 1 tells that the disturbed output vector satisfies $\frac{\epsilon}{S_b}$-differential privacy when a bound factor, S_b, is set to the global sensitivity. Thus, according to the observation in Figure 3(b), we set the proper bound factor as $S_b = 0.8$ to design Laplace noise with ϵ in the following experiments.

5.4 Evaluation of Our Proposed Approaches

The comparison results for ML-GCN, R-ML-GCN, P2-ML-GCN, RP2-ML-GCN, P2-ML-GCN with S_b, and RP2-ML-GCN with S_b are shown in Figure 4. These six models are implemented on Voc2007 dataset with 40 epochs by setting $\epsilon = 10$, $\lambda = 0.5$ and $S_b = 0.8$, whose results are shown in Figure 4(a). And they are also trained on MS-COCO dataset with 20 epochs by fixing $\epsilon = 10$, $\lambda = 0.5$ and $S_b = 0.8$, whose results are shown in Figure 4(b).

The OP value is used as an example for analysis. In Figures 4(a) and 4(b), $OP = 0.7152$ in P2-ML-GCN on Voc2007, $OP = 0.7793$ in RP2-ML-GCN on Voc2007, $OP = 0.6452$ in P2-ML-GCN on MS-COCO, and $OP = 0.7842$ in RP2-ML-GCN on MS-COCO. Obviously, the OP value of RP2-ML-GCN is higher than that of P2-ML-GCN on both two datasets. Also, from Figures 4(a) and 4(b), we observe that RP2-ML-GCN's other performance metrics are higher than P2-ML-GCN's on both two datasets through simple comparison. All comparison results for P2-ML-GCN and RP2-ML-GCN demonstrate that RP2-ML-GCN can improve P2-ML-GCN's prediction performance by reducing the bias of loss function with the help of an additional regularization term, which is consistent with our theoretical analysis in Section 4.2.2. In order to clearly illustrate the effectiveness of our proposed regularization term, we incorporated the regularization term into ML-GCN without adding Laplace noise, which is named R-ML-GCN. Concretely, we have $OP = 0.8001$ in ML-GCN on Voc2007, $OP = 0.8055$ in R-ML-GCN on Voc2007, $OP = 0.7954$ in ML-GCN on MS-COCO, and $OP = 0.7966$ in R-ML-GCN on MS-COCO, showning that R-ML-GCN's OP is higher than ML-GCN's on both datasets. Additionally, notice that R-ML-GCN's other performance metrics are higher than ML-GCN's on both two datasets. All these comparison results for ML-GCN and R-ML-GCN confirm that the regularization term can improve the prediction performance of ML-GCN even without Laplace noise, which has been analyzed in Section 4.2.2.

Similarly, by observing Figures 4(a) and 4(b), we obtain $OP = 0.7555$ in P2-ML-GCN with S_b on Voc2007, $OP = 0.7152$ in P2-ML-GCN on Voc2007, $OP = 0.8231$ in P2-ML-GCN with S_b on MS-COCO, and $OP = 0.6452$ in P2-ML-GCN on MS-COCO, which indicates that the OP value of P2-ML-GCN with S_b is better than that of P2-ML-GCN on both two datasets. In addition, P2-ML-GCN with S_b is better than P2-ML-GCN in terms of other performance metrics on both two datasets. Thus, we can improve the prediction performance of P2-ML-GCN by setting a proper bound to avoid excessive noise as analyzed in Section 4.3. Moreover, we train RP2-ML-GCN with S_b by integrating a regularization term and a proper bounded global sensitivity. Particularly, in Figure 4(a), $OP = 0.8011$ in RP2-ML-GCN with S_b on Voc2007, $OP = 0.7152$ in P2-ML-GCN on Voc2007, $OP = 0.7793$ in RP2-ML-GCN on Voc2007, and $OP = 0.7555$ in P2-ML-GCN with S_b on Voc2007; in Figure 4(b), we have $OP = 0.8491$ in RP2-ML-GCN with S_b on MS-COCO, $OP = 0.6452$ in P2-ML-GCN on MS-COCO, $OP = 0.7842$ in RP2-ML-GCN on MS-COCO, and $OP = 0.8231$ in P2-ML-GCN with S_b on MS-COCO. From these OP values, we can conclude that RP2-ML-GCN with S_b outperforms P2-ML-GCN, RP2-ML-GCN, and P2-ML-GCN with S_b on both two datasets. Besides, we can obtain the same conclusion by comparing other performance metrics. That is, RP2-ML-GCN with S_b is the best privacy-preserving deep learning model for multi-label image recognition among our proposed models.

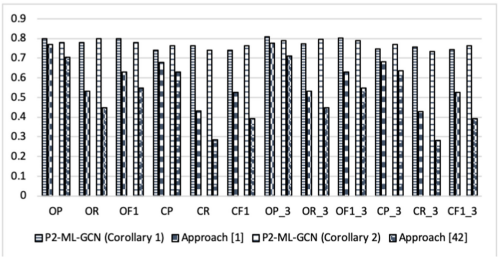
5.5 Our Proposed Approaches *v.s.* the State-of-the-Art

According to Corollary 1, we set $\epsilon = 10$ in P2-ML-GCN and $\mathcal{P} = 1/10$, making the model's weights satisfy 1-differential privacy. Meanwhile, as indicated by Corollary 2, we set $\epsilon = 10$ in P2-ML-GCN and $Q = 1/10$, making the model's input features achieve 1-differential privacy.

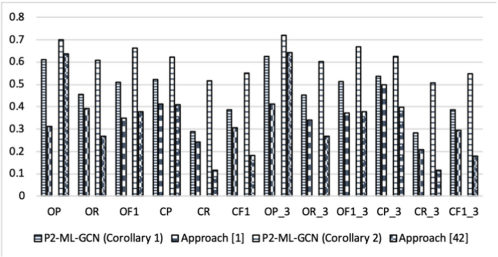
For a fair comparison, two existing schemes are adopted: (i) the scheme of [1] that adds the Laplace noise to make the model's weights meet ($\epsilon_w = 1$)-differential privacy; and (ii) the scheme of [45] that adds the Laplace noise to make the model's input features reach ($\epsilon_f = 1$)-differential privacy. The comparison results for 1-differential privacy on Voc2007 dataset and MS-COCO dataset are shown in Figures 5(a) and 5(c), respectively. In a similar way, we conduct comparative experiments to make weights or input features satisfy 0.1-differential privacy, whose results are presented in Figures 5(b) and 5(d).

For a clear illustration, we compare the OP values in Figure 5(a). As shown in Figure 5(a), we have $OP = 0.7995$ in P2-ML-GCN using Corollary 1 on Voc2007, $OP = 0.7714$ in the scheme of [1] on Voc2007, $OP = 0.7802$ in P2-ML-GCN using Corollary 2 on Voc2007, and $OP = 0.7063$ in the scheme of [45] on Voc2007. It can be seen that with the same degree of ϵ-differential privacy, the OP value of P2-ML-GCN is better than that of the two existing schemes. And via the same simple comparison, in Figure 5(a), other performance metrics of P2-ML-GCN are also better than those of the two existing schemes. That is, with the same degree of ϵ-differential privacy, the prediction performance of P2-ML-GCN is better than that of the two existing schemes, indicating that our P2-ML-GCN model outperforms the two existing schemes. Additionally, we can also obtain the same conclusion by comparing the results of Figures 5(b), 5(c), and 5(d). Furthermore, compared with P2-ML-GCN, RP2-ML-GCN can achieve the same degree of ϵ-differential privacy and enhanced prediction performance. Thus, we can conclude that RP2-ML-GCN also outperforms the two existing schemes, for which the main reason is that the noise added to the model's outputs in both P2-ML-GCN and RP2-ML-GCN can be bounded in deep learning training even if a neural network contains many layers.

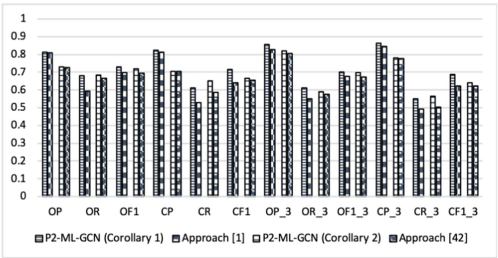
Evaluation Summary: All of the above experiments clearly demonstrate the superiority of our two proposed models, P2-ML-GCN and RP2-ML-GCN, in ensuring privacy protection, mitigating noise's side effect, and maintaining the model's accuracy, which is consistent with our theoretical analysis in Section 4.



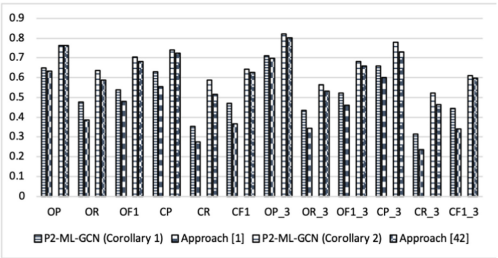
(a) P2-ML-GCN on Voc2007 Dataset Satisfying 1-differential privacy (Epoch:40)



(b) P2-ML-GCN on Voc2007 Dataset Satisfying 0.1-differential privacy (Epoch:40)



(c) P2-ML-GCN on MS-COCO Dataset Satisfying 1-differential privacy (Epoch:20)



(d) P2-ML-GCN on MS-COCO Dataset Satisfying 0.1-differential privacy (Epoch:20)

Fig. 5. P2-ML-GCN v.s. baselines.

6 CONCLUSION

In this article, we firstly propose P2-ML-GCN model to achieve privacy guarantee while accomplishing multi-label image recognition. Then, the Forbenius norm of weights in GCN is designed as a regularization term in RP2-ML-GCN to improve the prediction accuracy and robustness of P2-ML-GCN. Additionally, the idea of bounded global sensitivity is exploited to enhance the prediction accuracy. In both P2-ML-GCN and RP2-ML-GCN, our privacy-preserving mechanism implemented on the model's outputs not only can defend black-box attack but also can provide privacy protection for the model's weights and input features. Moreover, the effectiveness of privacy protection, regularization term, and bounded global sensitivity in our proposed models has been rigorously proved. The results of comprehensive real-data experiments, especially the comparison with the state-of-the-art, can validate the advantages of our proposed models.

REFERENCES

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 308–318.

[2] Kareem Amin, Alex Kulesza, Andres Munoz, and Sergei Vassilvtiskii. 2019. Bounding user contributions: A bias-variance trade-off in differential privacy. In *Proceedings of the International Conference on Machine Learning*. 263–271.

[3] Brett K. Beaulieu-Jones, Zhiwei Steven Wu, Chris Williams, Ran Lee, Sanjeev P. Bhavnani, James Brian Byrd, and Casey S. Greene. 2019. Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes* 12, 7 (2019), e005122.

[4] Karla Brkic, Ivan Sikiric, Tomislav Hrkac, and Zoran Kalafatic. 2017. I know that person: Generative full body and face de-identification of people in images. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 1319–1328.

[5] Zhipeng Cai, Zuobin Xiong, Honghui Xu, Peng Wang, Wei Li, and Yi Pan. 2021. Generative adversarial networks: A survey towards private and secure applications. *ACM Computing Surveys (CSUR)* 54, 6 (132), 1–38.

[6] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. 2018. Cartoongan: Generative adversarial networks for photo cartoonization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 9465–9474.
[7] Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. 2019. Multi-label image recognition with graph convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5177–5186.
[8] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our data, ourselves: Privacy via distributed noise generation. In *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 486–503.
[9] Torbjørn Eltoft, Taesu Kim, and Te-Won Lee. 2006. On the multivariate Laplace distribution. *IEEE Signal Processing Letters* 13, 5 (2006), 300–303.
[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision* 88, 2 (2010), 303–338.
[11] Weifeng Ge, Sibei Yang, and Yizhou Yu. 2018. Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1277–1286.
[12] Zongyuan Ge, Dwarikanath Mahapatra, Suman Sedai, Rahil Garnavi, and Rajib Chakravorty. 2018. Chest X-rays classification: A multi-label and fine-grained problem. arXiv:1807.07247. Retrieved from https://arxiv.org/abs/1807.07247.
[13] Marian George and Christian Floerkemeier. 2014. Recognizing products: A per-exemplar multi-label image classification approach. In *Proceedings of the European Conference on Computer Vision*. Springer, 440–455.
[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
[15] Briland Hitaj, Giuseppe Ateniese, and Fernando Perez-Cruz. 2017. Deep models under the GAN: Information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 603–618.
[16] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708.
[17] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM Journal on Computing* 40, 3 (2011), 793–826.
[18] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*.
[19] Yining Li, Chen Huang, Chen Change Loy, and Xiaoou Tang. 2016. Human attribute recognition by deep hierarchical contexts. In *Proceedings of the European Conference on Computer Vision*. Springer, 684–700.
[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*. Springer, 740–755.
[21] Magnus Lundmark and Carl-Johan Dahlman. 2017. Differential Privacy and Machine Learning: Calculating Sensitivity with Generated Data Sets. Master's thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.
[22] H. Brendan McMahan, Galen Andrew, Ulfar Erlingsson, Steve Chien, Ilya Mironov, Nicolas Papernot, and Peter Kairouz. 2018. A general approach to adding differential privacy to iterative training procedures. arXiv:1812.06210. Retrieved from https://arxiv.org/abs/1812.06210.
[23] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. 1532–1543.
[24] NhatHai Phan, Yue Wang, Xintao Wu, and Dejing Dou. 2016. Differential privacy preservation for deep auto-encoders: An application of human behavior prediction. In *Proceedings of the AAAI*, Vol. 16. AAAI, 1309–1316.
[25] NhatHai Phan, Xintao Wu, Han Hu, and Dejing Dou. 2017. Adaptive laplace mechanism: Differential privacy preservation in deep learning. In *Proceedings of the 2017 IEEE International Conference on Data Mining*. IEEE, 385–394.
[26] Abedallah Rababah. 1993. Taylor theorem for planar curves. *Proceedings of the American Mathematical Society* 119, 3 (1993), 803–810.
[27] Md Atiqur Rahman, Tanzila Rahman, Robert Laganière, Noman Mohammed, and Yang Wang. 2018. Membership inference attack against differentially private deep learning model. *Transactions on Data Privacy* 11, 1 (2018), 61–79.
[28] Nisarg Raval, Ashwin Machanavajjhala, and Landon P. Cox. 2017. Protecting visual secrets using adversarial nets. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 1329–1332.
[29] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In *Proceedings of the 2017 IEEE Symposium on Security and Privacy*. IEEE, 3–18.
[30] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR'15)*.

[31] Jordi Soria-Comas, Josep Domingo-Ferrer, David Sánchez, and David Megías. 2017. Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security* 12, 6 (2017), 1418–1429.

[32] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2818–2826.

[33] Stacey Truex, Ling Liu, Mehmet Emre Gursoy, Lei Yu, and Wenqi Wei. 2018. Towards demystifying membership inference attacks. arXiv:1807.09173. Retrieved from https://arxiv.org/abs/1807.09173.

[34] Ries Uittenbogaard, Clint Sebastian, Julien Vijverberg, Bas Boom, Dariu M. Gavrila, et al. 2019. Privacy protection in street-view panoramas using depth and multi-view imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 10581–10590.

[35] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. 2016. Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2285–2294.

[36] Zhouxia Wang, Tianshui Chen, Guanbin Li, Ruijia Xu, and Liang Lin. 2017. Multi-label image recognition by recurrently discovering attentional regions. In *Proceedings of the IEEE International Conference on Computer Vision*. 464–472.

[37] Xiu-Shen Wei, Quan Cui, Lei Yang, Peng Wang, and Lingqiao Liu. 2019. RPC: A large-scale retail product checkout dataset. arXiv:1901.07249. Retrieved from https://arxiv.org/abs/1901.07249.

[38] Nan Wu, Farhad Farokhi, David Smith, and Mohamed Ali Kaafar. 2020. The value of collaboration in convex machine learning with differential privacy. In *Proceedings of the 2020 IEEE Symposium on Security and Privacy*. IEEE, 304–317.

[39] Xi Wu, Fengan Li, Arun Kumar, Kamalika Chaudhuri, Somesh Jha, and Jeffrey Naughton. 2017. Bolt-on differential privacy for scalable stochastic gradient descent-based analytics. In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1307–1322.

[40] Jing Xia, Weihua Huang, Zhong Ma, Xinfa Dai, and Li He. 2019. Gradient-based differential privacy optimizer for deep learning model using collaborative training mode. In *Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology*. IEEE, 208–215.

[41] Zuobin Xiong, Honghui Xu, Wei Li, and Zhipeng Cai. 2021. Multi-source adversarial sample attack on autonomous vehicles. *IEEE Transactions on Vehicular Technology* 70, 3 (2021), 2822–2835.

[42] Chugui Xu, Ju Ren, Deyu Zhang, Yaoxue Zhang, Zhan Qin, and Kui Ren. 2019. GANobfuscator: Mitigating information leakage under GAN via differential privacy. *IEEE Transactions on Information Forensics and Security* 14, 9 (2019), 2358–2371.

[43] Honghui Xu, Zhipeng Cai, Daniel Takabi, and Wei Li. 2021. Audio-visual autoencoding for privacy-preserving video streaming. *IEEE Internet of Things Journal* (2021).

[44] Xinyou Yin, JAN Goudriaan, Egbert A. Lantinga, JAN Vos, and Huub J. Spiertz. 2003. A flexible sigmoid function of determinate growth. *Annals of Botany* 91, 3 (2003), 361–371.

[45] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2019. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *Proceedings of the International Conference on Learning Representations*.

[46] Guodong Zhang, Chaoqi Wang, Bowen Xu, and Roger Grosse. 2018. Three mechanisms of weight decay regularization. In *the Processings of the 7th International Conference on Learning Representations (ICLR'19)*.

[47] Huadi Zheng, Haibo Hu, and Ziyang Han. 2020. Preserving user privacy for machine learning: Local differential privacy or federated machine learning? *IEEE Intelligent Systems* 35, 4 (2020), 5–14.

[48] Feng Zhu, Hongsheng Li, Wanli Ouyang, Nenghai Yu, and Xiaogang Wang. 2017. Learning spatial regularization with image-level supervisions for multi-label image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5513–5522.

[49] Tianqing Zhu and S. Yu Philip. 2019. Applying differential privacy mechanism in artificial intelligence. In *Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems*. IEEE, 1601–1609.

Received October 2020; revised August 2021; accepted October 2021