

面向车路协同推断的差分隐私保护研究

吴茂强,黄旭民,康嘉文,余 荣

(广东工业大学 自动化学院,广州 510006)

摘 要:车路协同推断通过联合车载终端与路侧边缘服务器进行深度卷积网络推断运算,提高了网络架构推断效率,但是存在用户隐私泄露问题。攻击者在未知车载终端网络结构和参数的前提下,通过训练反卷积网络的方式,可复原车载终端上传的计算结果对应的图像数据,从而发起图像还原攻击。基于差分隐私理论,针对图像还原攻击设计模型扰动、输入扰动、输出扰动3种防御算法,分别在车载终端深度卷积网络的模型参数、输入原始图像、输出计算结果中加入随机拉普拉斯噪声,干扰攻击者的图像还原。通过理论分析得出3种算法均满足差分隐私保护,攻击者难以从计算结果中挖掘出原始数据的隐私信息。实验结果表明,3种算法在有效防御黑盒图像还原攻击的同时能保持推断精确度在90%以上,其中模型扰动算法在均衡隐私保护和推断精确度方面的性能表现优于输入扰动和输出扰动算法。

关键词:车路协同推断;差分隐私;车联网;边缘计算;深度卷积网络

开放科学(资源服务)标志码(OSID):



中文引用格式:吴茂强,黄旭民,康嘉文,等.面向车路协同推断的差分隐私保护研究[J].计算机工程,2022,48(7):29-35.

英文引用格式:WU M Q,HUANG X M,KANG J W,et al.Research on differential privacy protection for collaborative vehicle-road inference[J].Computer Engineering,2022,48(7):29-35.

Research on Differential Privacy Protection for Collaborative Vehicle-Road Inference

WU Maoqiang,HUANG Xumin,KANG Jiawen,YU Rong

(School of Automation,Guangdong University of Technology,Guangzhou 510006,China)

[Abstract] Collaborative vehicle-road inference performs deep convolutional network inference operations by combining vehicular terminals and roadside edge servers.Although this process improves inference efficiency,it creates the issue of user privacy leaks.Attackers can use an image reconstruction attack to recover original image data from intermediate computation results uploaded by the vehicular terminal.This terminal's network structure and parameters are unknown by the attacker through training a deconvolutional network.This study proposes three different differential privacy-based defense algorithms: model perturbation, input perturbation, and output perturbation. These algorithms inject random Laplace noise into the model parameters, inputted original images, and outputted computation results, which inhibits the attacker's image recovery.A theoretical analysis is presented to verify that these defense algorithms satisfy differential privacy, making it difficult for attackers to extract sensitive information. Experimental results demonstrate that all three algorithms can effectively defend against black-box image reconstruction attacks while maintaining inference accuracy above 90%.Furthermore,the model perturbation algorithm yields higher performance on the balance between privacy protection and inference accuracy than the other two algorithms.

[Key words] collaborative vehicle-road inference; differential privacy; Internet of Vehicle(IoV); edge computing; deep convolutional network

DOI:10.19678/j.issn.1000-3428.0062665

基金项目:国家自然科学基金(61971148);广西自然科学基金(2018GXNSFDA281013);桂林市科学研究与技术开发计划项目(20190214-3)。

作者简介:吴茂强(1989—),男,博士,主研方向为车载边缘计算;黄旭民,副教授、博士;康嘉文,教授、博士;余 荣(通信作者),教授、博士、博士生导师。

收稿日期:2021-09-12 **修回日期:**2021-11-03 **E-mail:**maoqiang.wu@vip.163.com

0 概述

随着人工智能技术的快速发展,各种智能驾驶应用通过深度卷积网络的实时推断提高车辆驾驶的安全性^[1-2]和效率^[3-4],但是车载终端计算资源有限,难以承担深度卷积网络推断的计算开销^[5]。为了减少车辆计算负荷,车联网(Internet of Vehicle, IoV)边缘计算通过在路侧部署边缘服务器,就近为车辆提供丰富的计算资源^[6],然而车辆采集的数据量庞大,传输到边缘服务器进行推断处理需要消耗巨大的带宽资源,导致服务应用的时延过高^[7]。为此,研究人员提出车路协同推断架构,旨在联合车载终端与边缘服务器进行推断运算,从而提高推断效率^[8]。具体地,深度卷积网络被切分成两部分,依次由车载终端和边缘服务器运行^[9]。车载终端上传前半部分网络的计算结果(即中间数据)至边缘服务器接力处理,边缘服务器得到最终的结果并返回至车辆^[10]。车路协同推断主要有3个优势:前半部分网络一般为特征提取网络,计算开销较小,适合车载终端执行^[11];中间数据远小于原始图像,传输过程的通信量明显减少^[12];车载终端保留原始数据,在一定程度上保护了数据隐私^[13]。

然而,针对现有车路协同推断架构,攻击者可能根据上传的中间数据复原车载终端的原始图像,从而泄露用户隐私^[14]。为了防御图像还原攻击,文献[15-16]选择更深的深度卷积网络层作为切割点,减少中间数据的信息量,降低图像复原效果。文献[17]在输出的中间数据中添加随机噪声,从而干扰图像还原。这些防御方法需要通过总结大量实验的经验,在实际应用中可操作性不强,同时主要针对白盒攻击,即假设已知车载终端的深度卷积网络结构和参数,无法很好地适配实际情况。

本文研究车路协同推断的黑盒图像还原攻击并提出3种防御算法,分别在车载终端深度卷积网络的模型参数、输入图像、输出结果中添加随机噪声,干扰黑盒攻击者对图像的复原,同时对于3种防御算法进行差分隐私理论分析及性能评估。

1 车路协同推断的隐私泄露问题

1.1 车路协同推断架构

如图1所示,车路协同推断架构由车载终端、边缘服务器以及深度卷积网络组成。以路标识别应用为例,车载终端设备利用深度卷积网络推断,对车载摄像头拍摄到的路标图像进行类型识别。深度卷积网络被切分成两部分:前半部分 f_{θ_1} 在车载终端处进行处理;后半部分 f_{θ_2} 在边缘服务器处进行处理。车载终端以路标图像 X_0 作为输入,执行前半部分网络,得到中间数据 $f_{\theta_1}(x_0)$ 作为输出。车载终端上传中间数据给边缘服务器,边缘服务器以中间数据 $f_{\theta_1}(x_0)$ 作为输入,执行后半部分网络得到最终识别结果 $f_{\theta_2}(f_{\theta_1}(x_0))$,将结果返回车载终端,至此完成协同推断任务。

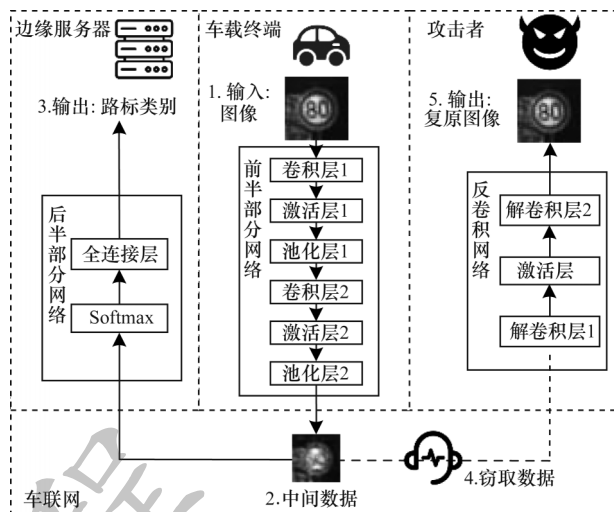


图1 车路协同推断架构

Fig.1 Collaborative vehicle-road inference architecture

1.2 黑盒攻击

在车路协同推断中,攻击者可能是开放环境下的窃听器,从车联网通信中窃取车载终端上传的中间数据以复原原始图像,泄露用户隐私。假设攻击者不知道车载终端设备存储的前半部分网络 f_{θ_1} 的结构,但是攻击者可以输入图像集合 $X = \{x_1, x_2, \dots\}$ 到网络去查询计算结果 $V = f_{\theta_1}(X)$ 。该假设通常应用于车载终端开放自己的API给其他车辆使用时,通过提供查询服务获取利益。

在黑盒攻击的假设下,本文采用反卷积网络算法^[14],通过训练一个反卷积网络,学习中间结果和原始图像之间的关系。

算法1 反卷积网络算法

输入 图像集合 X ,目标图像的中间数据 v_0 ,批量大小 B ,训练迭代次数 T ,学习率 η

输出 复原图像 x'_0

1. 输入 X 查询模型 f_{θ_1} 得到中间数据集合 $V = f_{\theta_1}(X)$

//查询阶段

2. 初始化反卷积网络 g_w 参数 w_0 //训练阶段

3. for $t \in T$ do

4. 切分 V 为批量集合 β ,每个批量样本数为 B

5. for $b \in \beta$ do

6. 更新反卷积网络参数 $w_{t+1} \leftarrow w_t - \eta \nabla l(w_t; b)$

7. end for

8. end for

9. 输入 v_0 到 g_w 得到复原数据 $x'_0 = g_w(v_0)$ //复原阶段

10. return x'_0

算法1包含3个阶段:1)查询阶段,攻击者使用图像集合 $X = \{x_1, x_2, \dots\}$ 作为输入查询车载终端设备网络 f_{θ_1} ,得到中间数据集合 $V = \{f_{\theta_1}(X_1), f_{\theta_1}(X_2), \dots\}$;2)训练阶段,攻击者用中间数据集合 V 作为输入,图像集合 X 作为目标对象,样本数量为 n ,训练反卷积网络 g_w ,采用图像像素空间的 l_2 范数作为损失函数(如式(1)所示),其中反卷积网络 g_w 的结构不需要关联车载终端设备网络 f_{θ_1} 的结构,本文实验中所用的

反卷积网络结构和车载终端设备网络结构完全不同;3)复原阶段,攻击者对训练好的反卷积网络输入获得的中间数据 $v_0 = f_{\theta_1}(x_0)$, 得到复原数据 $x'_0 = g_{\omega}(v_0)$ 。

$$l(\omega; X) = \frac{1}{n} \sum_{i=1}^n \|g_{\omega}(f_{\theta_1}(x_i)) - x_i\|_2^2 \quad (1)$$

2 差分隐私防御

2.1 差分隐私理论

差分隐私是数据分析和机器学习中定义隐私保护程度的一种数学范式^[18-19]。

定义 1 当给定的两个相邻数据集 D 和 D' 中至少有一条数据不同时,如果算法 M 符合条件式(2),则算法 M 满足 ϵ 差分隐私^[20]。

$$\Pr(M(D) \in S) < e^{\epsilon} \Pr(M(D') \in S) \quad (2)$$

其中: S 为算法输出集合。隐私预算 ϵ 控制输入 D 和 D' 的算法输出分布的接近程度,体现了隐私保护程度。 ϵ 越小,算法输出分布越接近,攻击者越难区分,隐私保护程度越高。

常见的隐私保护方法是在算法输出分布中加入随机噪声进行扰动^[21],如拉普拉斯噪声^[22]。如果要保证算法满足 ϵ 差分隐私,需要加入的噪声随机采样自均值为 0、尺度为 $\sigma \geq \Delta/\epsilon$ 的拉普拉斯分布^[23],其中 $\Delta = \max_{D,D'} \|M(D) - M(D')\|$ 为全局敏感度,即任意相邻数据集 D 和 D' 的算法输出的最大差异。

2.2 防御算法设计

本文提出 3 种差分隐私防御机制,即模型扰动、输入扰动、输出扰动。如图 2 所示,分别在车载终端网络的模型参数、输入图像和输出数据中加入拉普拉斯噪声扰动,影响攻击者复原的图像质量。

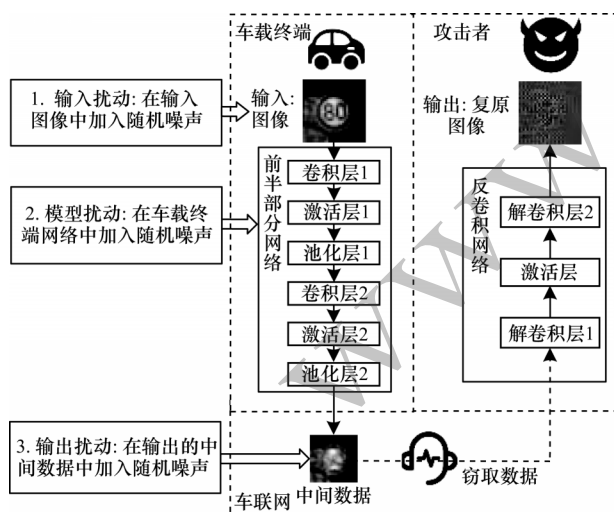


图 2 车路协同推断的差分隐私保护过程

Fig.2 Process of differential privacy protection for collaborative vehicle-road inference

模型扰动机制的具体流程如算法 2 所示。首先,将所有模型参数的大小限制在阈值 C_M 以内。根

据文献[18],阈值可以设定为模型参数的无穷范数的中值。然后,对模型参数添加随机生成的均值为 0、尺度为 $\sigma_M = 2C_M/\epsilon$ 的拉普拉斯噪声。最后,用噪声扰动后的深度卷积网络进行推断,得到中间数据 \bar{V} 。算法的复杂度取决于模型参数的维度,为 $O\left(\sum_{l=1}^L K_l^2 H_{l-1} H_l\right)$,其中, K_l 为第 l 个卷积层的卷积核边长, H_{l-1} 为该卷积层的输入通道数,即上一层的输出通道, H_l 为该卷积层的输出通道数。

算法 2 模型扰动算法

输入 图像集合 X , 车载终端深度卷积网络 f_{θ_1} , 模型参数阈值 C_M , 隐私预算 ϵ

输出 模型扰动后的输出结果 \bar{V}

步骤 1 限制模型参数大小 $\bar{\theta}_1 = \theta_1 / \max\left(1, \frac{\|\theta_1\|_{\infty}}{C_M}\right)$ 。

步骤 2 在模型参数中注入噪声 $\bar{\theta}_1 = \bar{\theta}_1 + \text{Lap}\left(\frac{2C_M}{\epsilon}\right)$ 。

步骤 3 计算经模型扰动后的输出结果 $\bar{V} = f_{\bar{\theta}_1}(X)$ 。

输入扰动机制的具体流程如算法 3 所示。首先,限制输入图像像素值在阈值 C_1 以内。根据文献[18],阈值可以设定为模型训练期间一组输入样本的无穷范数的中值。然后,随机生成均值为 0、尺度为 $\sigma_1 = 2C_1/\epsilon$ 的拉普拉斯噪声并注入输入图像的灰度值中。最后,用扰动后的图像进行推断,得到中间数据 \tilde{V} 。算法的复杂度取决于输入图像的维度,为 $O(n_1^2)$,其中 n_1 为输入图像的边长。

算法 3 输入扰动算法

输入 图像集合 X , 车载终端深度卷积网络 f_{θ_1} , 输入图像像素阈值 C_1 , 隐私预算 ϵ

输出 输入扰动后的输出结果 \tilde{V}

步骤 1 限制输入图像大小 $\tilde{X} = X / \max\left(1, \frac{\|X\|_{\infty}}{C_1}\right)$ 。

步骤 2 在输入图像中注入噪声 $\tilde{X} = \tilde{X} + \text{Lap}\left(\frac{2C_1}{\epsilon}\right)$ 。

步骤 3 计算经输入扰动后的输出结果 $\tilde{V} = f_{\theta_1}(\tilde{X})$ 。

输出扰动机制的具体流程如算法 4 所示。首先,推断得到中间数据 V ,并限制输出结果元素值在 C_0 内。根据文献[18],阈值可以设定为模型训练期间一组输出结果的无穷范数的中值。其次,对每个元素添加随机生成的均值为 0、尺度为 $\sigma_0 = 2C_0/\epsilon$ 的拉普拉斯噪声,得到中间数据 \hat{V} 。算法的复杂度取决于输出结果的维度,为 $O(n_0^2)$,其中 n_0 为输出结果的边长。

算法 4 输出扰动算法

输入 图像集合 X , 车载终端深度卷积网络 f_{θ_1} ,

输出结果元素阈值 C_0 , 隐私预算 ϵ

输出 模型扰动后的输出结果 \hat{V}

步骤1 计算输出结果 $V=f_{\theta_1}(X)$ 。

步骤2 限制输出结果大小 $\hat{V}=V/\max\left(1, \frac{\|V\|_{\infty}}{C_o}\right)$ 。

步骤3 对输出结果注入噪声 $\hat{V}=\hat{V}+\text{Lap}\left(\frac{2C_o}{\varepsilon}\right)$ 。

上述3种算法是在车载终端执行模型推断的过程中,分别对模型参数、输入图像、输出结果执行添加随机拉普拉斯噪声的操作。因此,算法的复杂度与深度模型本身的计算无关,而与添加噪声的对象维度有关。显然,输出结果维度最小,其次输入图像,模型参数维度最大。因此,输出扰动算法复杂度最低,其次输入扰动算法,模型扰动算法复杂度最高。

2.3 隐私保护分析

定理1 给定输入图像集合 X 和车载终端深度卷积网络 f_{θ_1} ,当注入模型参数的拉普拉斯噪声尺度为 $2C_M/\varepsilon$ 时,算法2满足 ε 差分隐私保护。

证明 给定任意相邻输入 X 和 X' ,有模型参数 θ_1 和 $\bar{\theta}_1$,使得 $f_{\theta_1}(X) \in S, f_{\bar{\theta}_1}(X') \in S$ 。因为模型参数大小阈值为 C_M ,所以全局敏感度表示如下:

$$\Delta_M = \max_{X, X'} \|\theta_1 - \bar{\theta}_1\|_1 \leq 2C_M \quad (3)$$

计算得到:

$$\frac{\Pr\left[f_{\theta_1 + \text{Lap}\left(0, \frac{2C_M}{\varepsilon}\right)}(X) = S\right]}{\Pr\left[f_{\bar{\theta}_1 + \text{Lap}\left(0, \frac{2C_M}{\varepsilon}\right)}(X') = S\right]} = \frac{e^{-\frac{\varepsilon}{2C_M}|\theta_1|}}{e^{-\frac{\varepsilon}{2C_M}|\bar{\theta}_1|}} = e^{\frac{\varepsilon}{2C_M}(|\theta_1| - |\bar{\theta}_1|)} \leq e^{\frac{\varepsilon}{2C_M}|\theta_1 - \bar{\theta}_1|} \leq e^{\varepsilon} \quad (4)$$

因此,根据定义1,当噪声尺度 $\sigma_M \geq 2C_M/\varepsilon$ 时,算法2满足 ε 差分隐私。

定理2 给定输入图像集合 X 和车载终端深度卷积网络 f_{θ_1} ,当注入输入图像的拉普拉斯噪声尺度为 $2C_I/\varepsilon$ 时,算法3满足 ε 差分隐私保护。

证明 因为输入图像像素值在范围 C_I 内,所以对于任意相邻输入 X 和 X' ,全局敏感度表示如下:

$$\Delta_I = \max_{X, X'} \|X - X'\|_1 \leq 2C_I \quad (5)$$

计算得到:

$$\frac{\Pr\left[f_{\theta_1}\left(X + \text{Lap}\left(0, \frac{2C_I}{\varepsilon}\right)\right) = S\right]}{\Pr\left[f_{\theta_1}\left(X' + \text{Lap}\left(0, \frac{2C_I}{\varepsilon}\right)\right) = S\right]} = \frac{e^{-\frac{\varepsilon}{2C_I}|X|}}{e^{-\frac{\varepsilon}{2C_I}|X'|}} = e^{\frac{\varepsilon}{2C_I}(|X| - |X'|)} \leq e^{\frac{\varepsilon}{2C_I}|X - X'|} \leq e^{\varepsilon} \quad (6)$$

因此,根据定义1,当噪声尺度 $\sigma_I \geq 2C_I/\varepsilon$ 时,算法3满足 ε 差分隐私保护。

定理3 给定输入图像 X 和车载终端深度卷积网络 f_{θ_1} ,当注入计算结果的拉普拉斯噪声尺度为

$2C_o/\varepsilon$ 时,算法4满足 ε 差分隐私保护。

证明 因为计算结果元素值限制在范围 C_o 内,所以对于任意相邻输入 X 和 X' ,全局敏感度表示如下:

$$\Delta_o = \max_{X, X'} \|f_{\theta_1}(X) - f_{\theta_1}(X')\|_1 \leq 2C_o \quad (7)$$

计算得到:

$$\frac{\Pr\left[f_{\theta_1}(X) + \text{Lap}\left(0, \frac{2C_o}{\varepsilon}\right) = S\right]}{\Pr\left[f_{\theta_1}(X') + \text{Lap}\left(0, \frac{2C_o}{\varepsilon}\right) = S\right]} = \frac{e^{-\frac{\varepsilon}{2C_o}|f_{\theta_1}(X)|}}{e^{-\frac{\varepsilon}{2C_o}|f_{\theta_1}(X')|}} = e^{\frac{\varepsilon}{2C_o}(|f_{\theta_1}(X)| - |f_{\theta_1}(X')|)} \leq e^{\frac{\varepsilon}{2C_o}|f_{\theta_1}(X) - f_{\theta_1}(X')|} \leq e^{\varepsilon} \quad (8)$$

因此,根据定义1,当噪声尺度 $\sigma_o \geq 2C_o/\varepsilon$ 时,算法4满足 ε 差分隐私保护。

3 实验结果与分析

3.1 实验设置

实验使用GTSRB数据集^[24],该数据集用于路标识别任务,由39 208个训练样本和12 630个测试样本组成。如图3所示,实验所用的深度卷积网络由6个卷积层和2个全连接层组成。每个卷积层有32个通道,核大小为3,并使用ReLU函数作为激活函数。每2个卷积层后面连接1个池化层。考虑车路协同推断选择第2个池化层作为切分点。攻击者采用的反卷积网络由2个解卷积层和1个激活层组成。深度卷积网络和反卷积网络的训练均采用ADAM优化器,学习率为0.001。

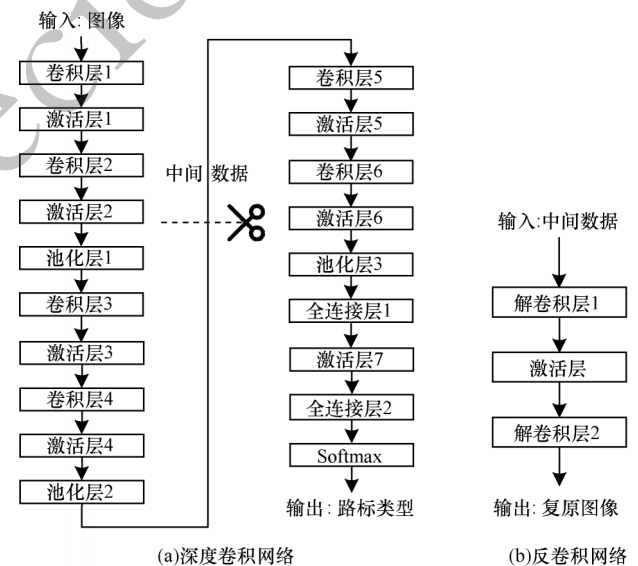


图3 深度卷积网络与反卷积网络结构

Fig.3 Structure of deep convolutional network and deconvolutional network

3.2 实验度量标准

假设 A 和 B 为原始图像和复原图像,大小为 $m \times n$ 。 $A(i, j)$ 和 $B(i, j)$ 分别为图像 A 和 B 在位置 (i, j) 的

像素值。实验采用以下3种度量标准衡量图像复原的质量^[17,25]:

1)均方误差(Mean Squared Error, MSE),以两张图像像素值的均方差衡量两张图像的相似度。MSE越小,两张图像的相似度越高。MSE定义如下:

$$\text{MSE}(A, B) = \frac{1}{mn} \sum_{i,j=1}^{m,n} \|A(i, j) - B(i, j)\|^2 \quad (9)$$

2)结构相似度(Structural Similarity, SSIM),根据两张图像的结构信息衡量相似度,取值范围为 $[0, 1]$,SSIM越大,两张图像的相似度越高。令图像 A 和 B 的像素均值分别为 μ_A 和 μ_B ,方差分别为 σ_A 和 σ_B ,协方差为 σ_{AB} , c_1 和 c_2 为参数。SSIM定义如下:

$$\text{SSIM}(A, B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \quad (10)$$

3)峰值信噪比(Peak Signal-to-Noise Ratio, PSNR),基于对应像素点的峰值误差来衡量相似度。PSNR越大,两张图像相似度越高。PSNR定义如下:

$$\text{PSNR}(A, B) = 10 \lg \left(\frac{255^2}{\text{MSE}(A, B)} \right) \quad (11)$$

3.3 实验结果

对比3种差分隐私防御算法与选择切点防御法对图像还原攻击的防御效果。选择切点防御法主选择更深的网络层作为切割点切分深度卷积网络,使得攻击者复原效果变差^[15-16]。

图4给出了在使用不同防御算法时,攻击者使用还原攻击复原的图像。模型扰动算法的防御效果最好,在隐私预算 $\epsilon=10$ 时,复原的图像已经无法看清路标的细节。输入扰动算法只有在 $\epsilon=1$ 时,才能完全抵御攻击者的图像复原。当 $\epsilon<1$ 时,在输出扰动算法保护下的复原图像像素均匀,而模型扰动算法和输入扰动算法防御下的复原图像仍存在噪点。这是因为输出扰动算法直接在输出结果中注入噪声,而另外两种扰动算法是间接扰动输出结果。当使用选择切点法进行防御(即选择更深的切割点)时,攻击者复原效果明显变差。但是,即使选择最后一层激活层作为切割点,其输出结果复原的图像仍能看到路标的细节。而且,如果选择更深的切割点,则车载终端的计算开销会显著增加,车路协同推断的效率会降低。相比之下,本文提出的3种差分隐私防御算法对图像还原攻击的防御效果更好。

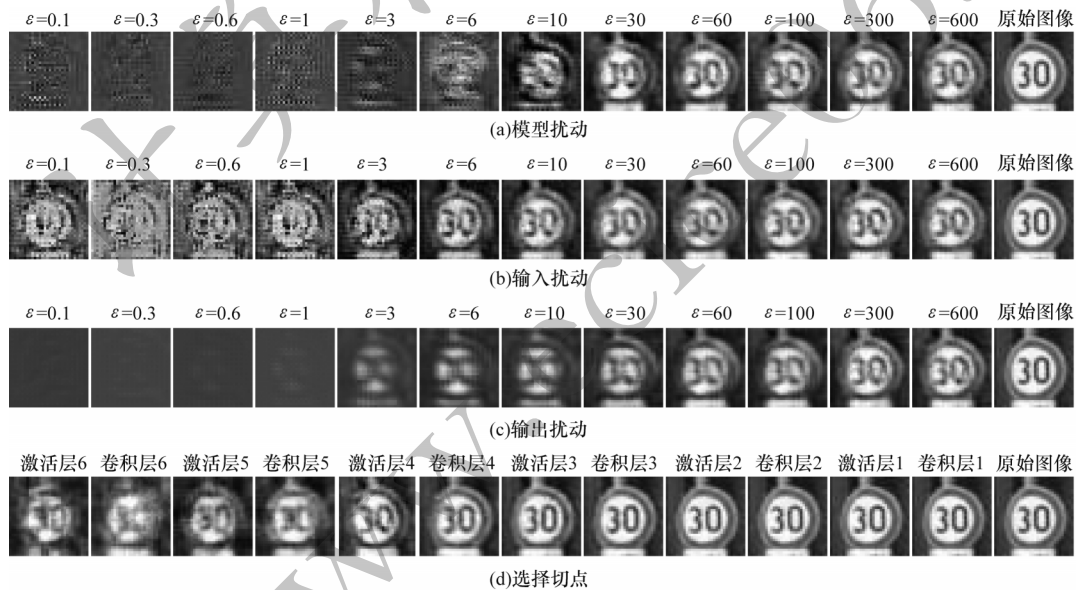


图4 不同防御算法保护下的复原图像

Fig.4 Recovered images protected by different defense algorithms

图5、图6和图7分别给出了对于不同的隐私预算,在3种防御算法干扰下复原图像的MSE、PSNR和SSIM。模型扰动算法的防御效果明显最优,在其干扰下的图像还原质量最低,其次是输出扰动算法,最后是输入扰动算法。但是当 $\epsilon<10^0$ 时,随着隐私预算降低,输入扰动算法防御下的复原图像MSE急剧升高,PSNR急剧下降。这是因为在输入图像添加噪声尺度太大,复原图像像素极度不均匀。

图8表示不同隐私预算对推断精确度的影响。

当 $\epsilon \geq 10^1$ 时,添加噪声对推断精确度的影响较小。当 $\epsilon < 10^1$ 时,推断精确度急剧下降。模型扰动算法比另外两种算法可达到更高的推断精确度,当 $\epsilon=10^1$ 时,其推断精确度为0.9。总体来说,当 ϵ 取值为 $[10^1, 10^2]$ 时,差分隐私保护算法可以有效降低黑盒攻击还原图像质量,并确保了一定的推断精确度。

图9、图10表示推断精确度和还原图像质量之间的关系。当PSNR和SSIM越小,推断精确度越低,即还原图像质量越差时,对黑盒图像还原攻击防

御效果越好,同时对深度卷积网络模型的推断精确度影响也越大。在达到相同的防御效果时,模型扰动算法达到的推断精确度最高,其次是输出扰动算法,最后是输入扰动算法。因此,在3种差分隐私保护算法中,模型扰动算法在均衡隐私保护和推断精确度方面获得最优效果。

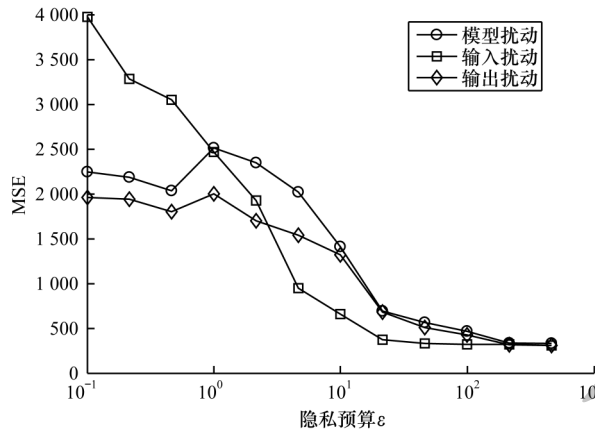


图5 不同隐私预算对应的复原图像 MSE
Fig.5 MSE of recovered images corresponding to different privacy budgets

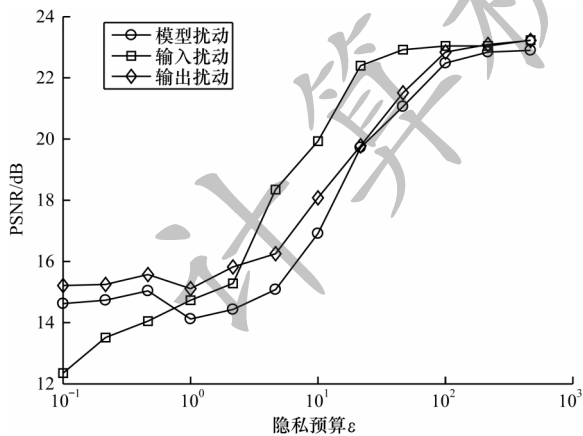


图6 不同隐私预算对应的复原图像 PSNR
Fig.6 PSNR of recovered images corresponding to different privacy budgets

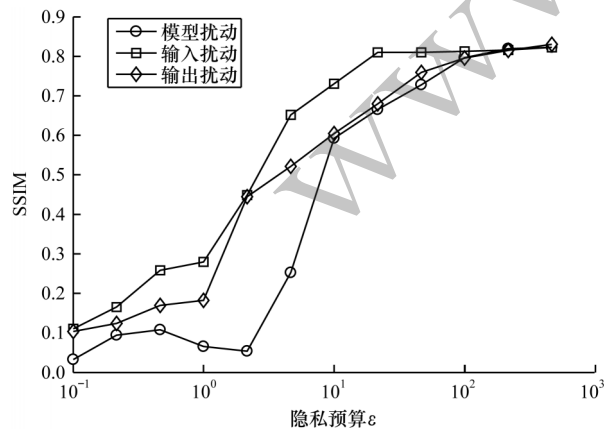


图7 不同隐私预算对应的复原图像 SSIM
Fig.7 SSIM of recovered images corresponding to different privacy budgets

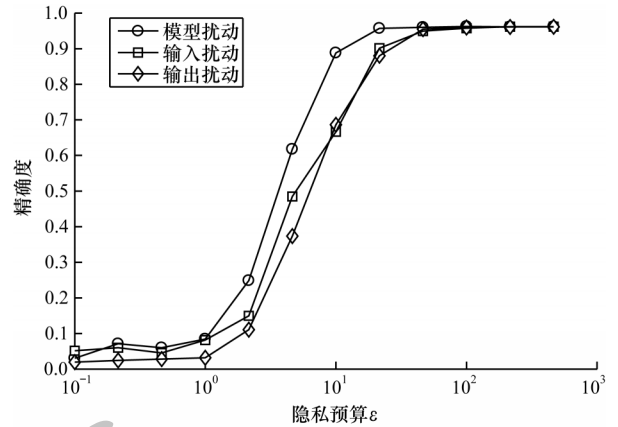


图8 不同隐私预算对推断精确度的影响
Fig.8 Impact of different accuracy budgets on inference accuracy

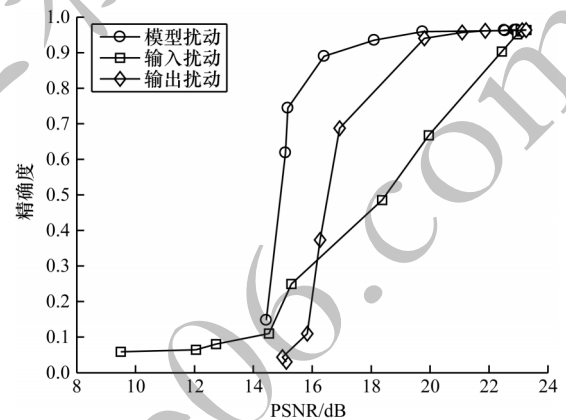


图9 推断精确度与 PSNR 的关系
Fig.9 Relationship between inference accuracy and PSNR

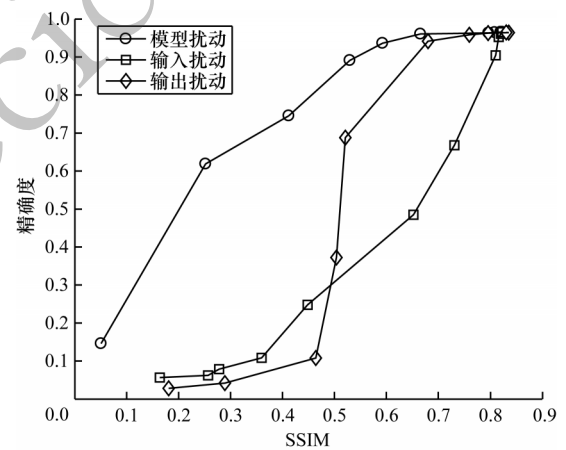


图10 推断精确度与 SSIM 的关系
Fig.10 Relationship between inference accuracy and SSIM

4 结束语

本文针对车路协同推断中的黑盒图像还原攻击,提出3种基于差分隐私的防御算法,分别在车载终端深度卷积网络的模型参数、输入图像、输出结果中注入随机生成的拉普拉斯噪声。通过理论分析得出3种算法均满足 ϵ 差分隐私保护的结论。实验结果证明,3种算法在有效防御黑盒图像还原攻击的同

时保证了车路协同推断的精确度。后续将结合传输压缩等方法设计更高效的防御算法,进一步提高车路协同推断的效率和精确度,实现用户隐私保护。

参考文献

- [1] HUANG X M, YU R, YE D D, et al. Efficient workload allocation and user-centric utility maximization for task scheduling in collaborative vehicular edge computing[J]. IEEE Transactions on Vehicular Technology, 2021, 70(4): 3773-3787.
- [2] LIM W Y B, HUANG J Q, XIONG Z H, et al. Towards federated learning in UAV-enabled Internet of vehicles: a multi-dimensional contract-matching approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(8): 5140-5154.
- [3] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.
- [4] MUHAMMAD K, ULLAH A, LLORET J, et al. Deep learning for safe autonomous driving: current challenges and future directions[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(7): 4316-4336.
- [5] WU M Q, HUANG X M, TAN B H, et al. Hybrid sensor network with edge computing for AI applications of connected vehicles[J]. Journal of Internet Technology, 2020, 21: 1503-1516.
- [6] LIU L, CHEN C, PEI Q Q, et al. Vehicular edge computing and networking: a survey[J]. Mobile Networks and Applications, 2021, 26(3): 1145-1168.
- [7] LIU Y J, WANG S G, ZHAO Q L, et al. Dependency-aware task scheduling in vehicular edge computing[J]. IEEE Internet of Things Journal, 2020, 7(6): 4961-4971.
- [8] ZHANG J, LETAIEF K B. Mobile edge intelligence and computing for the Internet of vehicles[J]. Proceedings of the IEEE, 2020, 108(2): 246-261.
- [9] KANG Y P, HAUSWALD J, GAO C, et al. Neurosurgeon: collaborative intelligence between the cloud and mobile edge[J]. ACM SIGPLAN Notices, 2017, 52(4): 615-629.
- [10] LI E, ZHOU Z, CHEN X. Edge intelligence: on-demand deep learning model co-inference with device-edge synergy[C]//Proceedings of 2018 Workshop on Mobile Edge Communications. New York, USA: ACM Press, 2018: 31-36.
- [11] WANG Q, LI Z Y, NAI K, et al. Dynamic resource allocation for jointing vehicle-edge deep neural network inference[J]. Journal of Systems Architecture, 2021, 117: 102133.
- [12] TAN X R, LI H J, WANG L M, et al. End-edge coordinated inference for real-time BYOD malware detection using deep learning[C]//Proceedings of IEEE Wireless Communications and Networking Conference. Washington D. C., USA: IEEE Press, 2020: 1-6.
- [13] LI E, ZENG L K, ZHOU Z, et al. Edge AI: on-demand accelerating deep neural network inference via edge computing[J]. IEEE Transactions on Wireless Communications, 2020, 19(1): 447-457.
- [14] HE Z C, ZHANG T W, LEE R B. Model inversion attacks against collaborative inference[C]//Proceedings of the 35th Annual Computer Security Applications Conference. New York, USA: ACM Press, 2019: 148-162.
- [15] SHI C S, CHEN L X, SHEN C, et al. Privacy-aware edge computing based on adaptive DNN partitioning[C]//Proceedings of IEEE Global Communications Conference. Washington D. C., USA: IEEE Press, 2019: 1-6.
- [16] HE Z C, ZHANG T W, LEE R B. Attacking and protecting data privacy in edge-cloud collaborative inference systems[J]. IEEE Internet of Things Journal, 2021, 8(12): 9706-9716.
- [17] RYU J, ZHENG Y F, GAO Y S, et al. Can differential privacy practically protect collaborative deep learning inference for the Internet of Things?[EB/OL]. [2021-08-17]. <https://arxiv.org/abs/2104.03813>.
- [18] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York, USA: ACM Press, 2016: 308-318.
- [19] ZHAO P, ZHANG G L, WAN S H, et al. A survey of local differential privacy for securing Internet of vehicles[J]. The Journal of Supercomputing, 2020, 76(11): 8391-8412.
- [20] 郝晨艳, 彭长根, 张盼盼. 重复攻击下差分隐私保护参数 ϵ 的选取方法[J]. 计算机工程, 2018, 44(7): 145-149.
- [21] HAO C Y, PENG C G, ZHANG P P. Selection method of differential privacy protection parameter ϵ under repeated attack[J]. Computer Engineering, 2018, 44(7): 145-149. (in Chinese)
- [22] WU M Q, YE D D, DING J H, et al. Incentivizing differentially private federated learning: a multidimensional contract approach[J]. IEEE Internet of Things Journal, 2021, 8(13): 10639-10651.
- [23] 王丹, 龙士工. 权重社交网络隐私保护中的差分隐私算法[J]. 计算机工程, 2019, 45(4): 114-118.
- [24] WANG D, LONG S G. Differential privacy algorithm for privacy protection in weighted social network[J]. Computer Engineering, 2019, 45(4): 114-118. (in Chinese)
- [25] NIU B, CHEN Y H, WANG B Y, et al. AdaPDP: adaptive personalized differential privacy[C]//Proceedings of IEEE Conference on Computer Communications. Washington D. C., USA: IEEE Press, 2021: 1-10.
- [26] YIN S H, DENG J C, ZHANG D W, et al. Traffic sign recognition based on deep convolutional neural network[M]. Berlin, Germany: Springer, 2017.
- [27] 雷蕾, 郭东恩, 靳峰. 基于谱归一化条件生成对抗网络的图像修复算法[J]. 计算机工程, 2021, 47(1): 230-238.
- [28] LEI L, GUO D E, JIN F. Image inpainting algorithm based on conditional generative adversarial network with spectral normalization[J]. Computer Engineering, 2021, 47(1): 230-238. (in Chinese)

编辑 陆燕菲