DPSNN: A Differentially Private Spiking Neural Network

5 authors, including:

Zhao Dongcheng
Chinese Academy of Sciences
41 PUBLICATIONS **311** CITATIONS

Guobin Shen
Chinese Academy of Tropical Agricultural Sciences
23 PUBLICATIONS **87** CITATIONS

Yi Zeng
Chinese Academy of Sciences
182 PUBLICATIONS **1,246** CITATIONS

DPSNN: A Differentially Private Spiking Neural Network

Jihang Wang[1,2,3*]**, Dongcheng Zhao**[1,3*]**, Guobin Shen**[1,3,6]
Qian Zhang[1,2,3†]**, Yi Zeng**[1,2,3,4,5,6†]
[1]Institute of Automation, Chinese Academy of Sciences (CASIA)
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Research Center for Brain-Inspired Intelligence, CASIA
[4]National Laboratory of Pattern Recognition, CASIA
[5]Center for Excellence in Brain Sciences and Intelligence Technology,
Chinese Academy of Sciences
[6]School of Future Technology, University of Chinese Academy of Sciences
```
{wangjihang2021, zhaodongcheng2016, shenguobin2021
              q.zhang, yi.zeng}@ia.ac.cn
```

Abstract

Privacy-preserving is a key problem for the machine learning algorithm. Spiking neural network (SNN) plays an important role in many domains, such as image classification, object detection, and speech recognition, but the study on the privacy protection of SNN is urgently needed. This study combines the differential privacy (DP) algorithm and SNN and proposes differentially private spiking neural network (DPSNN). DP injects noise into the gradient, and SNN transmits information in discrete spike trains so that our differentially private SNN can maintain strong privacy protection while still ensuring high accuracy. We conducted experiments on MNIST, Fashion-MNIST, and the face recognition dataset Extended YaleB. When the privacy protection is improved, the accuracy of the artificial neural network(ANN) drops significantly, but our algorithm shows little change in performance. Meanwhile, we analyzed different factors that affect the privacy protection of SNN. Firstly, the less precise the surrogate gradient is, the better the privacy protection of the SNN. Secondly, the Integrate-And-Fire (IF) neurons perform better than leaky Integrate-And-Fire (LIF) neurons. Thirdly, a large time window contributes more to privacy protection and performance.

1 Introduction

Recently, machine learning, especially deep learning, has been widely used in many domains such as image classification (He et al., 2016), visual tracking (Bertinetto et al., 2016), and speech recognition (Nassif et al., 2019). In addition to these scenarios, it is also used in applications rich in privacy data such as advertisement recommendation (Konapure & Lobo, 2021) and health care (Shahid et al., 2019). However, the model acquires better predictions by learning from a large amount of data, which has access to sensitive information. Recent studies (Fredrikson et al., 2015; Melis et al., 2019; Wei et al., 2018; Salem et al., 2020) have shown that the original user information can be recovered from the deep neural network through privacy attacks.

Differential Privacy(DP) (Dwork, 2008) theoretically gives a strong guarantee for the risk of privacy leakage. DP adds random mechanism to the data processing, making it difficult for observers to judge

---

[*]These authors have contributed equally to this work and share first authorship
[†]Corresponding Author.

small changes in the dataset through the output results to protect privacy. The differentially private stochastic gradient descent (DP-SGD) algorithm (Abadi et al., 2016) achieves DP on deep learning with random noise added on the clipping gradient. DP-SGD has achieved great success in many domains, such as medical imaging (Ziller et al., 2021; Li et al., 2019), Generative Adversarial Networks (GANs) (Hitaj et al., 2017; Jordon et al., 2018; Xie et al., 2018), and traffic flow estimation (Cai et al., 2019). The spiking neural network(SNN), known as the third generation of artificial neural network(ANN) (Maass, 1997), has gradually attracted researchers' attention due to its brain-like information processing method. As shown in Fig. 1, the non-real-valued information transmission greatly protects privacy. Even though the input currents are different, they all reach the threshold and thus produce the same output.

The training of SNN can be mainly divided into three categories: synaptic plasticity based (Diehl & Cook, 2015; Hao et al., 2020), conversion based (Li et al., 2021; Han & Roy, 2020), and backpropagation based (Wu et al., 2018). The introduction of the surrogate gradient makes the backpropagation algorithm successfully applied to the training of SNN (Neftci et al., 2019; Wu et al., 2018, 2019; Jin et al., 2018; Zhang & Li, 2020). In addition to facilitating SNN's training, the imprecise surrogate gradient makes it difficult for attackers to recover the users' information about the samples.
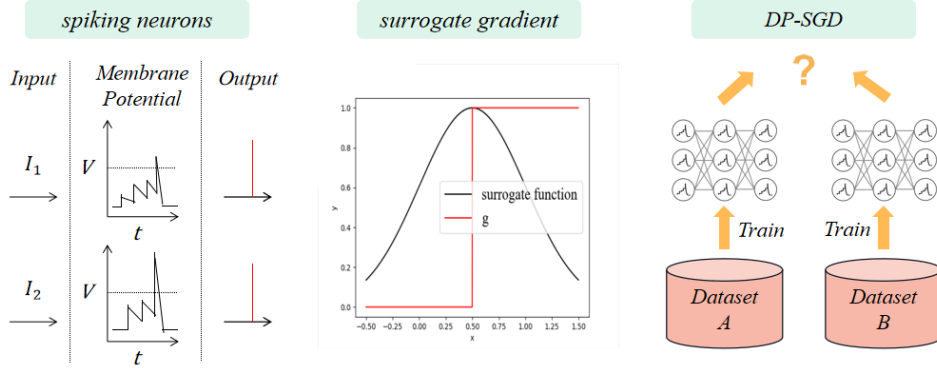

Figure 1: This figure shows the privacy-preserving measures in the differentially private spiking neural network(DPSNN).

This study proposes a differentially private spiking neural network(DPSNN) which combines the characteristics of SNN and DP as well. As shown in Fig. 1, The spiking neurons transmit information using spike sequences, the inexact derivative of the surrogate gradient, and the DP-SGD all provide a strong guarantee about the users' privacy. DPSNN shows high accuracy with strong privacy guarantees. The contributions of our study can be summarized as follows:

- To our best knowledge, this study is the first work that combines DP and SNN, which keeps an excellent trade-off between performance and privacy.
- We fully analyze the impact of various components of SNN on privacy protection, including different surrogate gradients, time window, and neuron types.
- We conducted experiments on MNIST (LeCun, 1998), Fashion-MNIST (Xiao et al., 2017) and Extended YaleB datasets (Georghiades et al., 2000). The experimental results have shown demonstrate superiority of our work. Compared with ANN with the same configuration, our DPSNN maintains a high performance while guaranteeing strong privacy. We achieved 98.36% test accuracy with 0.63-GDP privacy bound on MNIST, 87.51% with 0.51-GDP on Fashion-MNIST, and 99.58% with 0.66-GDP on Extended YaleB.

2 Related Work

2.1 Privacy-preserving Methods for Machine Learning

Researchers have developed several methods for privacy-preserving machine learning. Subsequently, homomorphic encryption (HE) is proposed. HE allows models to use encrypted data as input and

ouput encrypted results, then the clients can decrypt them and get the results they want without data leakage. Many basic machine learning algorithms can be implemented with HE (Du et al., 2017; Bost et al., 2014; Fenner & Pyzer-Knapp, 2020). In some cases, data from different enterprises must be used to obtain a better model, but no one wants to leak their data. Federated learning can solve this problem by exchanging encrypted intermediate machine learning results to achieve a better joint model while preventing data leakage (Yang et al., 2019).

DP is another widely used concept for privacy-preserving, which can tell us the difficulty of distinguishing two datasets from the perspective of an attacker (Dwork et al., 2016; Dong et al., 2021). Compared to the regularization methods such as dropout (Srivastava et al., 2014) and weight decay (Krogh & Hertz, 1991), DP has been proven to be an effective method for preserving privacy (Carlini et al., 2019). Li et al. (2020) proposes a differential private gradient boosting decision tree, which has a strong guarantee of DP while the accuracy loss is less than the previous algorithm. Differential private empirical risk minimization guarantees a privacy bound of the empirical risk minimization algorithm to protect the training data (Bassily et al., 2014). The moments accountant technique can analyze the DP in deep learning (Abadi et al., 2016). Li et al. (2020) proposes deep private auto-encoders, where DP is introduced by perturbing the cross-entropy errors.

2.2 Differentially Private Stochastic Gradient Descent

The DP-SGD algorithm (Abadi et al., 2016) uses gradient perturbation in the stochastic gradient descent and has become a standard privacy-preserving method for machine learning. While the DP-SGD algorithm effectively stops known attacks in the neural network, the noise added in gradient descent reduces the accuracy of the models. There have been many related works to improve the accuracy of DP-SGD. Yu et al. (2021) proposed expected curvature to improve the utility analysis of DP-SGD for convex optimization. Bu et al. (2020) uses Gaussian Differential Privacy(GDP) to estimate tighter privacy bounds of DP-SGD in the deep learning model. Papernot et al. (2021) replaces the Rectified Linear Unit(ReLU) activation with the tempered sigmoid activation in the ANN, which can prevent the activations from exploding to improve the accuracy of differential private ANN trained by DP-SGD.

SNN is gradually applied in many domains, and the privacy of SNN needs more attention. Also, the characteristics of SNN are more conducive to building a network with high privacy protection. This study is the first attempt to effectively combine SNN and DP to build a network with a high privacy guarantee and less accuracy loss.

3 Methods

3.1 Differential Privacy Theory

Two datasets are adjacent if there is only one different record between each other. For the definition of (ε, δ)-DP, we refer to Dwork's previous work (Dwork et al., 2016): For any pair of adjacent datasets D, D' and any event E, a mechanism M is (ε, δ)-DP if $\mathbb{P}(M(D) \in E) \leqslant e^{\varepsilon}\mathbb{P}(M(D') \in E) + \delta$ (Dwork et al., 2016).

This study performs privacy analysis in the framework of GDP theory (Dong et al., 2021). A mechanism M is μ-GDP if $T(M(D), M(D')) \geqslant T(N(0,1), N(\mu,1))$, where T means trade-off function (Dong et al., 2021). This means distinguishing between adjacent datasets D and D' is harder than distinguishing between Gaussian distribution $N(0,1)$ and $N(\mu,1)$.

3.2 Spiking Neural Network

The Integrate-And-Fire (IF) model is commonly used to describe the neuronal dynamics of SNN (Wu et al., 2018; Jin et al., 2018; Zhang & Li, 2020). The cell membrane is treated as a capacitor and the dynamics of IF neurons can be governed by a differential equation:

$$C\frac{dV(t)}{dt} = I(t) \tag{1}$$

where C is the membrane capacitance, $V(t)$ is the membrane potential at time t. $I(t) = \sum_{j=1}^{l} w_{ij}o_j$ denotes the pre-synaptic input current. When the membrane potential reaches a certain threshold,

the neurons fire a spike and return to the resting potential (we set the resting potential as zero in our model).

For better simulation, we convert the Eq. 1 into a discrete form with the Euler method into Eq.2:

$$V_i^{t+1,n} = V_i^{t,n}(1 - o_i^{t,n}) + \sum_{j=1}^{l(n-1)} w_{ij}^n o_j^{t+1,n-1} \qquad (2)$$

$o_i^{t,n}$ indicates whether the neuron i in layer n fires a spike at time t ($0 \le t \le T$, T is the time window). w_{ij}^n represents the weight of synaptic connections between the j-neuron in layer $n-1$ and i-neuron in layer n. $l(n-1)$ denotes the number of neurons in layer $n-1$. $V_i^{t,n}$ means the membrane potential of the i-neuron in layer n at time t.

The neuron will fire a spike when the membrane potential exceeds the threshold V_{th}, which is shown in Eq. 3. g is the spiking activation function.

$$o_i^{t+1,n} = g(V_i^{t+1,n}) = \begin{cases} 0, & V_i^{t+1,n} < V_{th} \\ 1, & V_i^{t+1,n} \ge V_{th} \end{cases} \qquad (3)$$

The biggest limitation that restricts the use of BP in SNN is the non-differentiable spike activation function g. The surrogate gradient utilizes the inexact gradients near the threshold to enable the BPTT algorithm to be successfully used in the training of SNN. Herein, we focus on three commonly used surrogate gradient functions as shown in Eq.4, 5, 6, and we call them Gate Function, Darctanh Function, and Normal Function, respectively. The influence of different surrogate functions on the privacy protection is discussed in detail in the Discussion section.

$$\frac{dg}{dV} = sign(|V - V_{th}| < \frac{1}{2}) \qquad (4)$$

$$\frac{dg}{dV} = \frac{4}{e^{-2(V-V_{th})} + e^{2(V-V_{th})} + 2} \qquad (5)$$

$$\frac{dg}{dV} = e^{2(V-V_{th})^2} \qquad (6)$$

3.3 Differentially Private Spiking Neural Network

Here, we propose DPSNN, a differentially private spiking neural network. DPSNN uses the iterative IF neurons as Eq. 2~3. For the output layer, we used the accumulated membrane potential as the final output as used in (Kim & Panda, 2020). The entire pipeline of our DPSNN is shown in Fig. 2. The network propagates the information in the forward process as the normal SNN. Furthermore, we clipped the gradient and added noise in the backpropagation process.

The loss function is set with cross-entropy loss.

$$L = -\sum_{i=1}^{l(d)} y_i \log\left(\frac{exp(V_i^{T,d}/T)}{\sum_{j=1}^{l(d)} exp(V_i^{T,d}/T)}\right) \qquad (7)$$

y_i is the true label. For the backward process of the network, we first calculated the gradient of each parameter. The details are shown below.

$$\frac{\partial L}{\partial w_{ij}^n} = \frac{\partial L}{\partial V_i^{t,n}} \frac{\partial V_i^{t,n}}{\partial w_{ij}^n} = \frac{\partial L}{\partial V_i^{t,n}} o_j^{t,n-1} \qquad (8)$$

We used the surrogate gradient mentioned above combined with the backpropagation through time (BPTT) to get the gradient of the corresponding time and layer.

$$\frac{\partial L}{\partial o_i^{t,n}} = \sum_{j=1}^{l(n+1)} \frac{\partial L}{\partial o_j^{t,n+1}} \frac{\partial o_j^{t,n+1}}{\partial o_i^{t,n}} + \frac{\partial L}{\partial o_i^{t+1,n}} \frac{\partial o_i^{t+1,n}}{\partial o_i^{t,n}}$$

$$= \sum_{j=1}^{l(n+1)} \frac{\partial L}{\partial o_j^{t,n+1}} \frac{\partial g}{\partial V_j^{t,n+1}} w_{ji}^{n+1} - \frac{\partial L}{\partial o_i^{t+1,n}} \frac{\partial g}{\partial V_i^{t+1,n}} V_i^{t,n} \qquad (9)$$
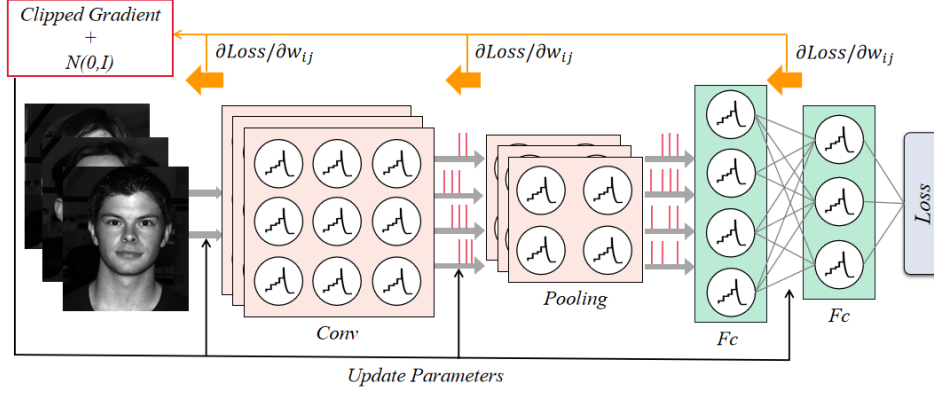
Figure 2: The forward process transmits information as the normal SNN. The backward process clipped the gradient and added noise to the clipped gradient (gradient perturbation).

$$
\begin{aligned}
\frac{\partial L}{\partial V_i^{t,n}} &= \frac{\partial L}{\partial o_i^{t,n}} \frac{\partial o_i^{t,n}}{\partial V_i^{t,n}} + \frac{\partial L}{\partial V_i^{t+1,n}} \frac{\partial V_i^{t+1,n}}{\partial V_i^{t,n}} \\
&= \frac{\partial L}{\partial o_i^{t,n}} \frac{\partial g}{\partial V_i^{t,n}} + \frac{\partial L}{\partial V_i^{t+1,n}} (1 - o_i^{t,n})
\end{aligned}
\tag{10}
$$

We can get the gradients of parameters $\frac{\partial L}{\partial \theta_m}$ where θ_m means the vector of all the w_{ij}^n after m times parameters update. To restrict the influence of each sample to the privacy, we clipped the gradient with a certain range with l_2 norm. The clipping operation ensured the l_2 norm of gradients was bounded by R

$$
\bar{v}_k = \frac{\partial L}{\partial \theta_m} \Big/ \max\left\{1, \left\|\frac{\partial L}{\partial \theta_m}\right\|_2 \frac{1}{R}\right\}
\tag{11}
$$

Then the algorithm injected Gaussian noise controlled by the noise scale σ into the clipped gradients, and updated the weights of the network with stochastic gradient descent (SGD) as shown in Eq. 12.

$$
\theta_{m+1} = \theta_m - \frac{\eta_t}{B}\left(\sum_{k=1}^{B} \bar{v}_k + \sigma R \cdot N(0, I)\right)
\tag{12}
$$

Where B is the mini-batch size and η_t is the learning rate. DPSNN iteratively optimizes the initial parameters θ_0 to θ_E. E is the number of iterations.

4 Experiments

In this section, we conducted experiments on several commonly used datasets to illustrate the superiority of our DPSNN. All the experiments were conducted on the PyTorch Framework with Intel(R) Core(TM) i5-7300HQ CPU. Moreover, to demonstrate the power of DPSNN for privacy protection, we compared experiments with ANN of the same structure using DP. To prove the generalization of the experiments, experiments were conducted on two different structures Net1 and Net2. The Net1 structure is 16C8S2-P2S1-32C4S2-P2S1-32-10, and the Net2 structure is 16C7S1-P2S2-32C4S1-P2S2-32-10, where 16C8S2 means 16 conventional filters of 8×8 and strides are 2, and P2S1 means 2×2 max-pooling and strides are 1. The activation function in the ANN is set with tanh, which is the same as Papernot's work (Papernot et al., 2021). Experiments compared the test accuracy under different noise scales σ to illustrate the privacy bound. The mini-batch size was 250, learning rate $\eta_t = 0.25$, and gradient norm bound $R = 1.5$.

The gradient approximation curve in the experiments was Eq. 4 and the time window $T = 5$. We discuss the influence of different surrogate functions, different neuron types, and time window in Section 5.

5

Table 1: The test accuracy of Net2-SNN and Net2-ANN with same privacy bound on MNIST dataset.

σ	SNN-Accuracy(%)	ANN-Accuracy(%)	MA-ε	CLT-ε	CLT-μ
1.6	97.96	95.35	1.46	1.32	0.35
1.3	98.19	96.57	1.93	1.74	0.44
1.0	98.36	97.31	2.93	2.60	0.63
0.7	98.55	97.95	6.55	5.31	1.18
0.5	98.63	98.45	23.20	21.65	3.66

4.1 MNIST

The MNIST (LeCun, 1998) is a hand-written digit dataset that contains 60,000 labeled training images and 10,000 labeled testing images. Each digit sample is a 28×28 grayscale image sourced from the postal codes of 0-9.

Based on the above network structures, we conducted experiments on ANN and SNN, respectively, denoted as Net1-ANN, Net1-SNN, Net2-ANN, and Net2-SNN. The number of training epochs was set to 60. As shown in Fig. 3, we compared the different accuracy curves of three different privacy protection scenarios, including no privacy protection and noise scale σ as 1.0, 1.6.
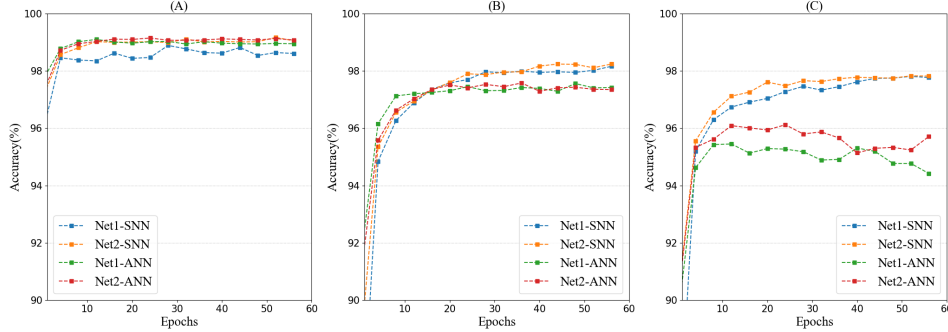

Figure 3: (A) is the test accuracy of the networks on MNIST without differentially private stochastic gradient descent (DP-SGD). (B) and (C) are the test accuracy with the noise scale σ is 1.0 and 1.6.

As shown in the Fig. 3, when the DP-SGD is not applied, both ANN and SNN guarantee high accuracy. When the degree of noise added is low, such as $\sigma = 1.0$, SNN has a higher accuracy rather than ANN. When the noise scale increases from $\sigma = 1.0$ to $\sigma = 1.6$, with the increase of the training epochs, the performance of ANN gradually decreases, but the SNN always maintains a high performance. This shows that ANN based on real-valued neurons is considerably affected by noise, while our DPSNN can guarantee a great trade-off between performance and privacy protection.

Also, we performed a quantitative analysis of privacy bounds on Net2. The ε of the (ε, δ)-DP privacy bounds are calculated by the moments accountant (Abadi et al., 2016) and central limit theorem (Bu et al., 2020) and are denoted as MA-ε and CLT-ε. δ is set with 10^{-5}. We also compute the μ-GDP privacy bound with the central limit theorem and denote the cost as CLT-μ. The smaller the ε and μ, the better the privacy protection. In Table 1, the test accuracy of Net2 on SNN and ANN with the same privacy bounds are shown. It can be seen that both ANN and SNN demonstrate great performance under poor privacy protection, which the CLT-ε = 21.65. However, when the privacy protection is gradually improved, the accuracy of SNN does not change much, while for ANN, the accuracy drops from 98.45% to 95.35%. Our DPSNN guaranteed a high accuracy while having strong privacy protection and achieved the 98.36% performance at the 0.63-GDP.

4.2 Fashion-MNIST

The Fashion-MNIST (Xiao et al., 2017) is more complex compared with MNIST, which has the same image size as MNIST. It has 60,000 training images and 10,000 test images. Each example is associated with a label from 10 classes.
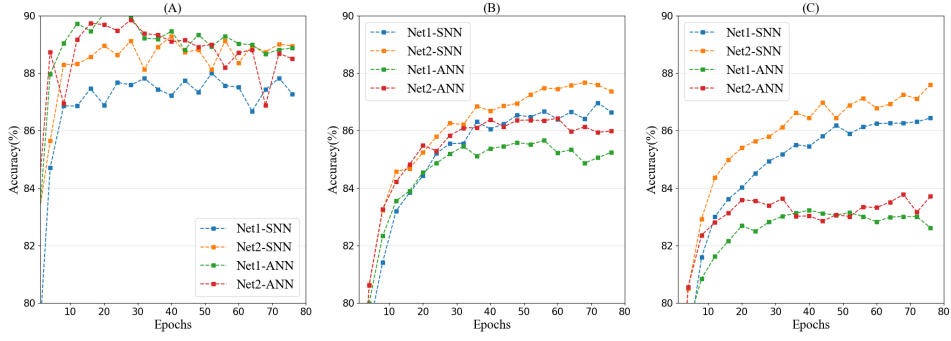
Figure 4: (A) is the test accuracy of the networks on Fashion-MNIST without differentially private stochastic gradient descent (DP-SGD). (B) and (C) are the test accuracy with the noise scale σ is 1.0 and 1.6.

Table 2: The test accuracy of Net2-SNN and Net2-ANN with same privacy bound on Fashion-MNIST dataset.

σ	SNN-Accuracy(%)	ANN-Accuracy(%)	MA-ε	CLT-ε	CLT-μ
1.6	87.31	83.82	1.61	1.46	0.38
1.3	87.51	84.75	2.23	2.02	0.51
1.0	87.93	85.90	3.55	3.18	0.76
0.7	88.34	87.72	8.28	7.02	1.49
0.5	88.37	88.33	24.98	23.77	3.93

The training epochs are set to 80. The test accuracy of two structures on Fashion-MNIST with different privacy guarantees is shown in Fig. 4. For Net1 and Net2, when there is no privacy protection, both ANN and SNN show strong performance. However, when the injected noise increases, the learning curve for the SNN does not change much, while for ANN, the learning curve drops significantly. Our DPSNN maintains high performance while ensuring strong privacy protection.

For Net2, we calculate the accuracy under various privacy bounds, as shown in Table 2. Both ANN and SNN show superior performance when privacy protection is low. When the μ drops from 3.93-GDP to 0.38-GDP, the accuracy of ANN drops 4.5% while SNN only drops approximately 1%. Compared with MNIST, Fashion-MNIST is more complex and is more susceptible to noise. When privacy protection is improved, the accuracy of Fashion-MNIST is more likely to decrease. DPSNN can maintain a good balance between privacy protection and performance on more complex datasets. DPSNN is an effective way to improve privacy-utility trade-off.

4.3 Extended YaleB

In addition to the common datasets, privacy protection has received particular attention in the field of personal privacy data, especially facial recognition. This study performs face recognition experiments on the Extended YaleB dataset. The Extended YaleB dataset contains face images of 29 subjects under 9 poses and 64 illumination conditions. Every subject has 576 grayscale images, and the size of each image is 640×480.

In this study, 128 images are randomly selected from each subject's data as the testing set, while the rest are the training set. We construct SNN for privacy-guaranteed facial recognition. The

Table 3: The best test accuracy in 80 epochs under different noise scale on Extended YaleB dataset.

σ	Accuracy(%)	MA-ε	CLT-ε	CLT-μ
2.5	99.30	2.21	1.96	0.49
2.0	99.58	3.05	2.71	0.66
1.5	99.55	4.42	3.89	0.9

network structure is 16C7S1-P2S2-32C4S1-P2S2-128-29. The gradient norm bound R is 2.0, and the mini-batch size is 256.

Table 3 shows the best test accuracy in 80 epochs under different noise scales. The network can achieve a test accuracy of about 99.58%, and the model is only 0.66-GDP, realizing the facial recognition under solid privacy guarantees.

5 Discussion

The privacy of SNN trained based on backpropagation is affected by multiple settings, such as the surrogate function setup, the simulation time window, and using IF neurons or LIF neurons. Based on the Fashion-MNIST dataset and Net2 structures, we then analyze the impact of different settings on the privacy protection of SNN.

5.1 The Influence of Surrogate Function

The learning curves of different surrogate functions 4,5,6 under different noise scale σ has been shown in the Fig. 5.
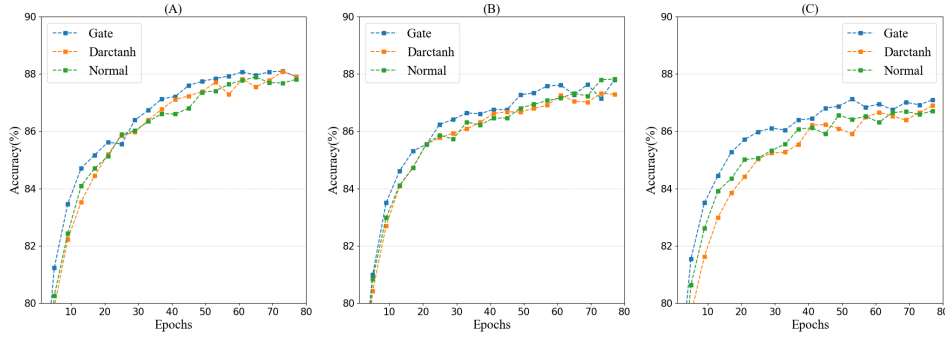

Figure 5: The performance of the differentially private spiking neural network (DPSNN) using different surrogate functions. (A)∽(C) are the test accuracy when the noise scale is 0.5, 1.0, 1.6 respectively.

When the privacy protection is gradually strengthened, the Gate Function has higher accuracy than the other two surrogate functions. The main reason is that the Gate Function is more imprecise in the derivation of the gradient than the other two functions. The imprecise surrogate gradient provides strong privacy guarantees.

5.2 The influence of different neuron types

In addition to the IF neurons, leaky Integrate-And-Fire (LIF) neuron is another commonly used neuron model in the deep SNN. The dynamics of LIF neurons can be described as follows:

$$\tau \frac{dV(t)}{dt} = -V(t) + RI(t) \tag{13}$$

where $\tau = RC$ is the time constant, R is the membrane resistance. The iterative LIF model adds a parameter $\lambda = 1 - \frac{1}{\tau}$ into Eq. 2 to describe the leak rate of the membrane potential as follows:

$$V_i^{t+1,n} = \lambda V_i^{t,n}(1 - o_i^{t,n}) + \sum_{j=1}^{l(n-1)} w_{ij}^n o_j^{t+1,n-1} \tag{14}$$

Here, we set $\lambda = 0.2$ and compare the performance of LIF neurons and IF neurons. As can be seen in Fig. 6, when the noise scale is 0.5, 1.0 and 1.6, we plot learning curves of IF neurons and LIF neurons.
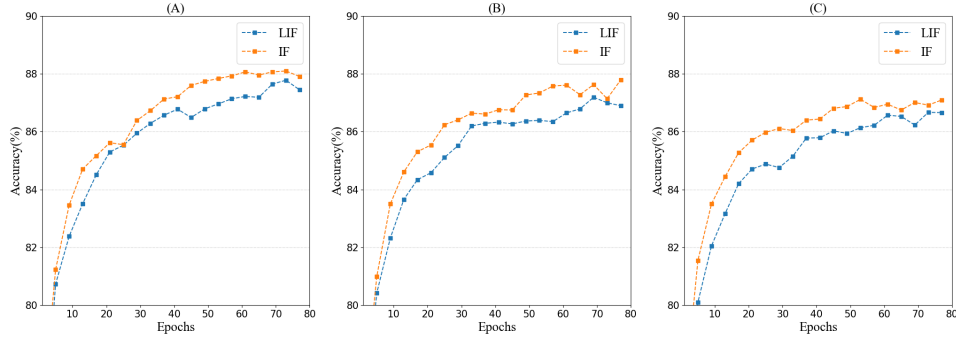
8

Figure 6: The performance of the differentially private spiking neural network (DPSNN) using leaky Integrate-And-Fire (LIF) neurons and Integrate-And-Fire (IF) neurons. (A)⌒(C) are the test accuracy when the noise scale is 0.5, 1.0, 1.6 respectively.

Under different noise scales, the privacy of the IF neuron network is better. The leak rate enables the network to pay attention to different influences of information at different time steps. The farther the time step, the smaller the influence. The impact of this temporal information increases the risks of privacy leakage to a certain extent, and consequently, when the membrane potential does not decay, the network's privacy can be significantly protected.

5.3 The Influence of Time Window

Also, the simulation time window is the key point for the SNN training. We give the corresponding privacy bound of when the SNN reaches the best accuracy under different time windows as shown in Table 4 when the noise scale $\sigma = 1.0$.

Table 4: The accuracy and privacy bound of differentially private spiking neural network (DPSNN) under different time windows

		Accuracy(%)	MA-ε	CLT-ε	CLT-μ
	5	87.93	3.55	3.18	0.76
Time Window T	10	88.49	3.43	3.06	0.73
	15	89.0	3.45	3.09	0.74

When the time window increases from 5 to 10, the performance is improved, and privacy protection is enhanced. With the time window increasing from 10 to 15, the performance is further improved, while the privacy bound does not change much. A large time window allows the information to spread over different time steps. When the network is attacked, gradients at multiple moments confuse the attacker, making it hard to accurately recover the sample's real information. When $T = 1$, it is equivalent to ANN to a certain extent, the network is easier to attack, and the privacy protection is worse.

6 Conclusion

This study is the first to combine the DP algorithm with SNN, and propose a DPSNN. DPSNN maintains a good trade-off between privacy protection and performance. We conducted experiments on MNIST, Fashion-MNIST, and the facial recognition dataset, Extended YaleB, and found that our DPSNN approach can achieve excellent privacy protection while maintaining high accuracy. Furthermore, we analyzed the influence of different factors on the privacy protection of SNN. We found that the imprecise surrogate function (Gate Function), the large time window, and using IF neurons instead of LIF neurons significantly contribute to constructing a high privacy-preserving and high-performance SNN.

9

References

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

Bassily, R., Smith, A., and Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 464–473. IEEE, 2014.

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pp. 850–865. Springer, 2016.

Bost, R., Popa, R. A., Tu, S., and Goldwasser, S. Machine learning classification over encrypted data. *Cryptology ePrint Archive*, 2014.

Bu, Z., Dong, J., Long, Q., and Su, W. J. Deep learning with gaussian differential privacy. *Harvard data science review*, 2020(23), 2020.

Cai, Z., Zheng, X., and Yu, J. A differential-private framework for urban traffic flows estimation via taxi companies. *IEEE Transactions on Industrial Informatics*, 15(12):6492–6499, 2019.

Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., and Song, D. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX Security Symposium (USENIX Security 19)*, pp. 267–284, 2019.

Diehl, P. U. and Cook, M. Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in computational neuroscience*, 9:99, 2015.

Dong, J., Roth, A., and Su, W. Gaussian differential privacy. *Journal of the Royal Statistical Society*, 2021.

Du, Y., Gustafson, L., Huang, D., and Peterson, K. Implementing ml algorithms with he. *MIT Course*, 6:857, 2017.

Dwork, C. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pp. 1–19. Springer, 2008.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality*, 7(3):17–51, 2016.

Fenner, P. and Pyzer-Knapp, E. Privacy-preserving gaussian process regression–a modular approach to the application of homomorphic encryption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3866–3873, 2020.

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

Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings fourth ieee international conference on automatic face and gesture recognition (cat. no. pr00580)*, pp. 277–284. IEEE, 2000.

Han, B. and Roy, K. Deep spiking neural network: Energy efficiency through time based coding. In *European Conference on Computer Vision*, pp. 388–404. Springer, 2020.

Hao, Y., Huang, X., Dong, M., and Xu, B. A biologically plausible supervised learning method for spiking neural networks using the symmetric stdp rule. *Neural Networks*, 121:387–395, 2020.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

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

Jin, Y., Zhang, W., and Li, P. Hybrid macro/micro level backpropagation for training deep spiking neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 7005–7015, 2018.

Jordon, J., Yoon, J., and Van Der Schaar, M. Pate-gan: Generating synthetic data with differential privacy guarantees. In *International conference on learning representations*, 2018.

Kim, Y. and Panda, P. Revisiting batch normalization for training low-latency deep spiking neural networks from scratch. *Frontiers in neuroscience*, pp. 1638, 2020.

Konapure, R. and Lobo, L. Video content-based advertisement recommendation system using classification technique of machine learning. In *Journal of Physics: Conference Series*, volume 1854, pp. 012025. IOP Publishing, 2021.

Krogh, A. and Hertz, J. A simple weight decay can improve generalization. *Advances in neural information processing systems*, 4, 1991.

LeCun, Y. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

Li, Q., Wu, Z., Wen, Z., and He, B. Privacy-preserving gradient boosting decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 784–791, 2020.

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., et al. Privacy-preserving federated brain tumour segmentation. In *International workshop on machine learning in medical imaging*, pp. 133–141. Springer, 2019.

Li, Y., Zeng, Y., and Zhao, D. Bsnn: Towards faster and better conversion of artificial neural networks to spiking neural networks with bistable neurons. *arXiv preprint arXiv:2105.12917*, 2021.

Maass, W. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997.

Melis, L., Song, C., De Cristofaro, E., and Shmatikov, V. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 691–706. IEEE, 2019.

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., and Shaalan, K. Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7:19143–19165, 2019.

Neftci, E. O., Mostafa, H., and Zenke, F. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 2019.

Papernot, N., Thakurta, A., Song, S., Chien, S., and Erlingsson, Ú. Tempered sigmoid activations for deep learning with differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9312–9321, 2021.

Salem, A., Bhattacharya, A., Backes, M., Fritz, M., and Zhang, Y. {Updates-Leak}: Data set inference and reconstruction attacks in online learning. In *29th USENIX Security Symposium (USENIX Security 20)*, pp. 1291–1308, 2020.

Shahid, N., Rappon, T., and Berta, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS one*, 14(2):e0212356, 2019.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1): 1929–1958, 2014.

Wei, L., Luo, B., Li, Y., Liu, Y., and Xu, Q. I know what you see: Power side-channel attack on convolutional neural network accelerators. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pp. 393–406, 2018.

Wu, Y., Deng, L., Li, G., Zhu, J., and Shi, L. Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in neuroscience*, 12:331, 2018.

Wu, Y., Deng, L., Li, G., Zhu, J., Xie, Y., and Shi, L. Direct training for spiking neural networks: Faster, larger, better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1311–1318, 2019.

Xiao, H., Rasul, K., and Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Xie, L., Lin, K., Wang, S., Wang, F., and Zhou, J. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*, 2018.

Yang, Q., Liu, Y., Chen, T., and Tong, Y. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

Yu, D., Zhang, H., Chen, W., Yin, J., and Liu, T.-Y. Gradient perturbation is underrated for differentially private convex optimization. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 3117–3123, 2021.

Zhang, W. and Li, P. Temporal spike sequence learning via backpropagation for deep spiking neural networks. *Advances in Neural Information Processing Systems*, 33:12022–12033, 2020.

Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., and Kaissis, G. Medical imaging deep learning with differential privacy. *Scientific Reports*, 11(1):1–8, 2021.