arXiv:2302.01500v1 [cs.LG] 3 Feb 2023

脉冲突触惩罚：节能脉冲神经网络的适当惩罚项

Kazuma Suetake [*1] **Takuya Ushimaru** [*1] **Ryuji Saiin** [1] **Yoshihide Sawada** [2]

摘要

脉冲神经网络（SNNs）凭借其脉冲特性，成为一种节能型神经网络。然而，随着脉冲神经网络的脉冲发放率提高，其能耗也随之增加，从而削弱了脉冲神经网络的优势。为解决这一问题，我们在训练阶段将一种新型脉冲活动惩罚项引入目标函数。我们的方法旨在直接优化能耗指标，而无需修改网络架构。因此，在保持准确率的同时，所提出的方法比其他方法更能降低能耗。我们针对图像分类任务进行了实验，结果表明，所提出的方法有效缓解了能耗–准确率权衡的困境。

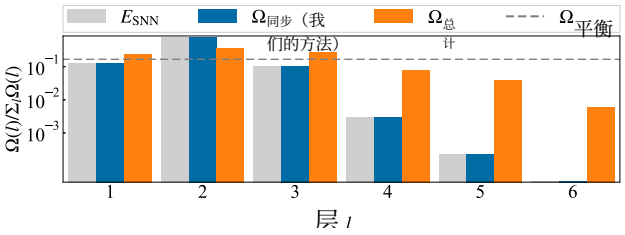

图*1*。 逐层惩罚项比较。 x轴表示层数， y轴表示某个逐层指标或惩罚项与某个总指标或惩罚项的比值。网络架构为CNN7（附录 A.1），计算细节见第3节。我们的惩罚项（蓝色）与真实值（灰色）成严格正比。

1. 引言

随着神经网络的迅速发展和普及，实现节能型神经网络已成为可持续发展的当务之急。其中一种模型是脉冲神经网络（SNN），它也被认为比普通人工神经网络（ANN）更符合生物学特性。脉冲神经网络通过异步处理脉冲信号，在神经形态芯片（Akopyan 等人，2015；Davies 等人，2018）或某些现场可编程门阵列（FPGA）（Maguire 等人，2007；Misra 和 Saha，2010）上实现节能运行。然而，随着脉冲神经网络的脉冲发放率增加，其能耗也会随之上升，因而脉冲神经网络的优势会减弱。因此，除了从人工神经网络转向脉冲神经网络之外，采用能够降低推理阶段能耗的训练方法也很有益。同时，这种训练方法应独立于网络架构，以避免限制其应用。也就是说，我们的目标是开发一种训练方法，在不受网络架构限制的情况下实现节能型脉冲神经网络。

实现节能型脉冲神经网络有多种方法，例如剪枝、量化和知识蒸馏（Kundu 等人，2021；Chowdhury 等人，2021a；Lee 等人，2021），这些方法在人工神经网络中也被广泛采用。此外，还有专门针对脉冲神经网络的方法，通过稀疏化与能耗相关的脉冲活动来降低能耗（Lee 等人，2020；Kim & Panda，2021；Naya 等人，2021），我们的方法便属于这一类。具体而言，在训练阶段对脉冲发放率施加惩罚的方法，与我们上述目标密切相关（Esser 等人，2016；Sorbaro 等人，2020；Pellegrini 等人，2021）。然而，这些方法通过任意降低脉冲发放率来间接减少能耗，而两者之间并不存在严格的正相关关系。因此，在维持准确率的同时降低能耗并非易事。

原理与构想 我们的原则是：应在训练阶段直接优化该指标。基于这一思路，我们提出在目标函数中引入一个适当的脉冲活动惩罚项，即 脉冲突触惩罚。该惩罚项经过推导，其期望值与脉冲神经网络的能耗指标严格成正比（见图 1）。尽管所提出的方法与现有方法之间的差异仅在于每个脉冲对应的缩放因子，

[*]共同贡献者 [1]AISINSOFTWARE，日本爱知 [2]AISIN 东京研究中心，日本东京。通讯作者：Kazuma Suetake <kazuma.suetake@aisin-software.com>。

预印本。版权所有 © 2023 作者。

我们证明，这一细微的修正带来了显著的改进。

主要贡献

- 我们推导出一种新颖的惩罚项，无需修改网络架构，即可直接优化脉冲神经网络的总能耗指标。
- 我们证明，所提出的方法与权重衰减兼容，后者会对网络施加隐式稀疏性（Yaguchi 等人，2018）；同时，所提出的方法比权重衰减产生更强的稀疏化效果。
- 我们证明，在保持图像分类任务准确率的同时，所提出的方法比其他方法更能降低能耗，从而缓解能耗–准确率权衡带来的困境。

2. 相关工作

2.1. 直接训练脉冲神经网络中的脉冲稀疏化

与我们提出的方法最相关的方案，是为脉冲发放率引入惩罚项，并通过替代梯度法直接训练脉冲神经网络（Esser 等人，2016；Pellegrini 等人，2021）。对脉冲发放率施加惩罚以获得节能型脉冲神经网络，是一种直接明了的思路，因为脉冲发放率是脉冲神经网络能耗指标中的一项（Lee 等人，2020；Kim & Panda，2021）。我们将脉冲发放率的降低称为脉冲稀疏化。由于不可微性，无法通过常规反向传播法优化脉冲发放率，但可以通过替代梯度法对其进行优化；训练脉冲神经网络中的脉冲神经元时也采用了同样的技术（Zenke& Ganguli，2018；Shrestha& Orchard，2018）。然而，这些惩罚项（Esser 等人(2016)；Pellegrini 等人(2021)）都无法与能耗指标精确匹配。与这些方法不同，我们提出的脉冲突触惩罚解决了这一局限。

2.2. 通过人工神经网络转换实现脉冲稀疏化

其他方法会针对相应的 ReLU 网络（具有 ReLU 激活的人工神经网络）引入惩罚项，并将其转换为脉冲神经网络（Sorbaro 等人，2020；Narduzzi 等人，2022）。尽管无法保证针对 ReLU 网络的惩罚项有助于降低转换后脉冲神经网络的能耗，但 ReLU 网络可以通过序数反向传播法进行优化。请注意，Sorbaro 等人（2020）提出了与我们的方法相同的惩罚项突触缩放因子，以降低脉冲神经网络的能耗。然而，他们未能如他们所提到的，提供证据来支持这一主张。与他们不同，我们在通过替代梯度直接训练脉冲神经网络的设置下，提供了理论和实验方面的证明。

2.3. 神经元稀疏化

神经元稀疏化是指增加激活值永久为零的数量，即死亡神经元的数量。这比脉冲稀疏化的条件更为严格，因为脉冲稀疏化并不要求神经元永久处于不活动状态。在 ReLU 网络中，使用 Adam 优化器并进行权重衰减正则化的训练会隐式地促使神经元稀疏化（Yaguchietal.，2018），因为 ReLU 激活具有不活动状态。然而，尽管脉冲神经网络通常也采用权重衰减，但这一论断尚未在脉冲神经网络的背景下得到验证；脉冲神经元同样具有不活动状态。因此，为了准确检测我们方法的效果，我们还将重点关注权重衰减。

3. 方法

本节提出脉冲突触惩罚。首先，我们介绍采用替代梯度机制的脉冲神经元模型。接着，我们说明能耗指标，该指标可用脉冲活动表示。需要注意的是，在训练阶段，我们需要同时优化准确率和能效。最后，我们指出，脉冲突触惩罚是优化能耗指标的适当惩罚项。

3.1. 神经元模型和替代梯度

在本研究中，我们使用由单步脉冲神经元构成的脉冲神经网络。对于静态任务，其训练和推理成本优于多时间步脉冲神经网络（Suetake 等人，2023）。请注意，单步脉冲神经元的设置与先前关于惩罚项的研究相同（Esser 等人，2016）。

我们将 $l \in \{l \in \boldsymbol{Z} \mid 1 \leq l \leq L\}$ 记作层，将 d_l 记作第 l 层中的神经元数量，将 $\boldsymbol{s}_0 = x \in X \subset \boldsymbol{R}^{d_0}$记作输入数据，并将任意向量的下标 i 记作其第 i个分量。于是，单步脉冲神经元定义如下（Suetake 等人，2023）：

定义 3.1。单步脉冲神经元的前向过程由两个成分构成：膜电位$\boldsymbol{u}_l \in \boldsymbol{R}^{d_l}$ 以及神经元发出的脉冲 $\boldsymbol{s}_l \in \{0,1\}^{d_l}$。它们通过 Heaviside 阶跃函数 H 定义如下：

$$\boldsymbol{u}_l := \boldsymbol{W}_l \boldsymbol{s}_{l-1}, \tag{1}$$

$$s_{l,i}\left(u_{l,i}\right) := H\left(u_{l,i} - u_{\mathrm{th}}\right) = \begin{cases} 1 & \left(u_{l,i} \geq u_{\mathrm{th}}\right), \\ 0 & \left(u_{l,i} < u_{\mathrm{th}}\right), \end{cases} \tag{2}$$

其中， $\boldsymbol{W}_l \in \boldsymbol{R}^{d_l \times d_{l-1}}$ 是突触连接的强度，

也称为权重矩阵，且 $u_{\text{th}} \in \boldsymbol{R}$ 是脉冲发放阈值（式1中 $l=1$ 对应于直接编码（Rueckauer 等人，2017）。

单步神经元本身的反向模式无法用于标准反向传播算法，因为式 2的导数几乎处处为零。因此，我们采用称为替代梯度的技术，即用某个合理的函数正式替代导数函数，例如以下函数（Suetake 等人，2023）：

$$\frac{\partial s_{l,i}}{\partial u_{l,i}}\left(u_{l,i}\right) :\simeq \begin{cases} \frac{1}{\tau}\frac{1}{u_{l,i}} & \left(u_{l,i} \geq u_{\text{th}}\right), \\ \frac{\partial \sigma_\alpha}{\partial u_{l,i}}\left(u_{l,i}\right) & \left(u_{l,i} < u_{\text{th}}\right), \end{cases} \tag{3}$$

其中，τ 和 α 是超参数，σ_α 是缩放 sigmoid 函数，其表达式如下：

$$\sigma_\alpha\left(u_{l,i}\right) := \frac{1}{1+\exp\left(\left(-u_{l,i}+u_{\text{th}}\right)/\alpha\right)}, \tag{4}$$

$$\frac{\partial \sigma_\alpha}{\partial u_{l,i}}\left(u_{l,i}\right) = \frac{1}{\alpha}\sigma_\alpha\left(u_{l,i}\right)\left(1-\sigma_\alpha\left(u_{l,i}\right)\right). \tag{5}$$

请注意，替代函数的选择与我们的提议无关。

3.2. 能耗度量指标

我们用符号 $\psi_{l,\,i}$ 表示第 l 层第 i 个神经元发出的突触数量，即网络架构中不会被强制置零的（$\boldsymbol{W}_l$）$_{*,\,i} \in \boldsymbol{R}^{d_l}$ 矩阵元素数量。令 W_l，H_l 分别表示特征图的宽度和高度，C_l 表示通道大小，k_l 表示第 l 层的卷积核大小。为简单起见，我们将卷积核的宽度和高度限制为相同，均为 k_l。那么，$\psi_{l,\,i}$ 的具体形式如下，例如标准全连接（fc）层和二维卷积（conv）层，

$$\psi_{l,i} = \psi_l = \begin{cases} C_{l+1} & (\text{fc}), \\ \frac{W_{l+1}H_{l+1}}{W_l H_l} C_{l+1} k_l^2 & (\text{conv}), \end{cases} \tag{6}$$

其中，卷积层假设采用了适当的填充。请注意，某些非平凡的填充、步幅和膨胀可能会导致 i 对 $\psi_{l,i}$ 的依赖。

使用 $\psi_{l,i}$，我们可以将浮点运算次数（FLOPs）表示如下。浮点运算次数常用作衡量人工神经网络计算复杂度的指标：

$$\text{FLOPs}(l) := \sum_{i=1}^{d_l} \psi_{l,i}, \tag{7}$$

以及逐层和总计脉冲发放率，这些也是衡量脉冲活动稀疏性的重要指标

在脉冲神经网络中，如下所示：

$$R(l) := \mathop{\mathbb{E}}_{x\in X}\left[\frac{\sum_{i=1}^{d_l} s_{l,i}}{d_l}\right], \tag{8}$$

$$R := \sum_{l=1}^{L} R(l), \tag{9}$$

其中，操作 $\mathbb{E}_{x\in X}$ 表示对数据集 X 取经验期望。接下来，我们要优化的能耗指标定义如下。

定义 3.2。令 T 表示时间步的数量，E_{AC} 表示每次累加操作的能耗。那么，脉冲神经网络逐层及总能耗指标定义如下：

$$E_{\text{SNN}}(l) := TE_{\text{AC}} \mathop{\mathbb{E}}_{x\in X}\left[\sum_{i=1}^{d_l} \psi_{l,i} s_{l,i}\right], \tag{10}$$

$$E_{\text{SNN}} := \sum_{l=1}^{L} E_{\text{SNN}}(l). \tag{11}$$

请注意，T 在单步神经元模型中等于 1，我们使用 $E_{\text{AC}} = 0.9$ [pJ](Horowitz，2014)。
如果 $\psi_{l,\,i}$ 与 i ($\exists \psi_l,\ \forall i,\ \psi_{l,\,i} = \psi_l$) 无关，结合方程 7 和 8，方程 10 可改写如下：

$$\begin{aligned} E_{\text{SNN}}(l) &= TE_{\text{AC}} \sum_{i=1}^{d_l} \psi_l \mathop{\mathbb{E}}_{x\in X}\left[\frac{\sum_{i=1}^{d_l} s_{l,i}}{d_l}\right] \\ &= TE_{\text{AC}}\text{FLOPs}(l)R(l), \end{aligned} \tag{12}$$

这与Kim & Panda（2021）中使用的指标相同。为简单起见，我们考虑在第 $\psi_{l,\,i}$与 i 节中彼此独立的情况。4。

3.3. 脉冲突触惩罚

为了优化能耗 E_{SNN}（式 11），我们提出以下惩罚项。

定义 3.3。层级和总脉冲突触惩罚项定义如下：

$$\Omega_{\text{syn}}(l) = \Omega_{\text{syn}}(l, \boldsymbol{s}_l) := \frac{1}{p}\sum_{i=1}^{d_l} \psi_{l,i} s_{l,i}^p, \tag{13}$$

$$\Omega_{\text{syn}} = \Omega_{\text{syn}}(\boldsymbol{s}) := \sum_{l=1}^{L} \Omega_{\text{syn}}(l, \boldsymbol{s}_l), \tag{14}$$

where $\boldsymbol{s} := \{\boldsymbol{s}_l\}_{l=1}^{L}$ and $p \geq 1$.

总能耗指标与总脉冲突触惩罚之间的等价性可直接由二者的定义以及公式$s_{l,\,i}^p = s_{l,\,i}$推得，该公式由方程 2导出。

定理 3.4。层级和总脉冲突触惩罚的期望值与脉冲神经网络的层级和总能耗指标严格成正比：

$$pE_{AC} \mathop{\mathbb{E}}_{x\in X}[\Omega_{\text{syn}}(l)] = E_{SNN}(l), \tag{15}$$

$$pE_{AC} \mathop{\mathbb{E}}_{x\in X}[\Omega_{\text{syn}}] = E_{SNN}, \tag{16}$$

对于任意 $p \geq 1$。

这一事实意味着，优化式14 等同于优化式11。因此，我们强烈建议使用式14 作为惩罚项，以优化能耗指标。下文中，若仅称其为脉冲突触惩罚（包括式9和11），则指总脉冲突触惩罚。

备注3.5。所提出的惩罚项可以像第3.1节中的替代梯度那样进行优化，并且使用 $p \neq 1$ 选项可在脉冲未发放时控制反向信号，如下所示：

$$\begin{aligned}\frac{1}{p}\frac{\partial s_{l,i}^{p}}{\partial u_{l,i}}(u_{l,i}) &= s_{l,i}^{p-1}\frac{\partial s_{l,i}}{\partial u_{l,i}} \\ &\simeq \begin{cases} \frac{\partial s_{l,i}}{\partial u_{l,i}} & (u_{l,i} \geq u_{\text{th}}), \\ 0 & (u_{l,i} < u_{\text{th}}), \end{cases}\end{aligned} \tag{17}$$

其中我们使用了式2。根据式13 和17，当 $p > 1$时，两者没有本质区别；因此，除$p > 1$ 选项外，我们不考虑其他 $p = 2$，后者在若干研究中被普遍采用（Esser 等人，2016；Pellegrini 等人，2021）。请注意，式17的推导显然独立于替代梯度的选择。然而，开放问题依然存在，即，从理论上无法判定哪种选择更好，$p = 1$或 $p > 1$。我们针对 $p = 1$，2 进行了实验检验，如第4节所述。

3.3.1. 与其他惩罚项的区别

其他可选的惩罚项如下：

$$\Omega_{\text{total}} = \frac{1}{p}\sum_{l=1}^{L}\sum_{i=1}^{d_l} s_{l,i}^{p}, \tag{18}$$

$$\Omega_{\text{balance}} = \frac{1}{p}\sum_{l=1}^{L}\sum_{i=1}^{d_l}\frac{1}{d_l}s_{l,i}^{p}, \tag{19}$$

其中，对于 $p = 2$，Ω_{total} 和Ω_{balance} 分别与Esser 等人(2016) 和 Pellegrini 等人(2021) 中的定义相同。然而，我们认为，尽管它们在一定程度上使脉冲活动稀疏化，但二者都无法直接优化能耗，因为它们不具备比例关系（公式16）。

图。 1和 2以及表 1 展示了以下实验所用模型的能耗指标与惩罚项之间的差异。

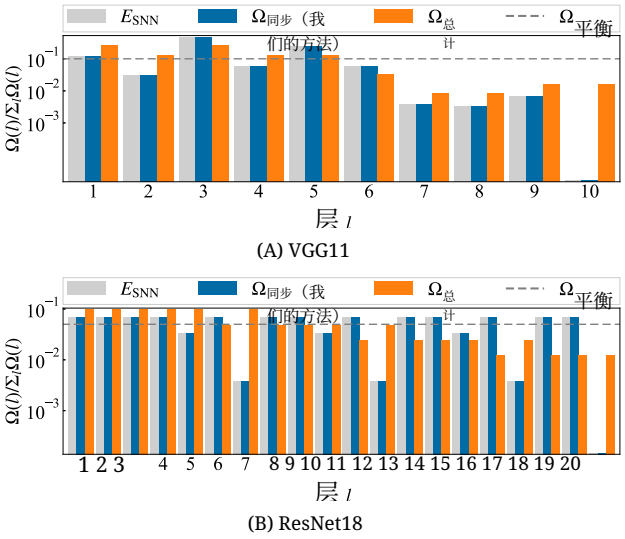

图 2. 逐层惩罚项的比较。网络架构分别为 (A) VGG11 和 (B) ResNet18（附录 A.1）。如图 1所示，我们的惩罚项（蓝色）与真实值（灰色）严格成正比。

表1。总惩罚项之间的比较。我们的惩罚项 Ω_{syn} 与真实值 $E_{\text{SNN}}/E_{\text{AC}}$完全相等。

模型	$E_{\text{SNN}}/E_{\text{AC}}$	Ω_{syn}	Ω_{total}	Ω_{balance}
CNN7	98895888	98895888	59688	6
VGG11	2526060544	2526060544	249856	10
ResNet18	553730048	553730048	671744	20

在这些图和表中，为简单起见，我们假设所有脉冲神经元都发放了即,$s_{l,i} = 1(\forall l, i)$。在这一假设下，不失一般性，真实值与浮点运算次数成正比。这些图和表表明，脉冲突触惩罚与能耗指标严格成正比，而其他惩罚项则并非如此。在下一节中，我们将通过实验验证这一主张对性能的影响。

3.3.2. 惩罚项归一化

惩罚项以其强度参数 λ 作为耦合系数 $\lambda\Omega_*$纳入目标函数，其中符号 $*$ 表示“同步”、“总计”或“平衡”。为了便于处理不同惩罚项之间或不同规模模型之间的强度参数，我们建议将惩罚项除以 $\Omega_*(\mathbf{1})$ 进行归一化，其中$\mathbf{1}$ 表示$\boldsymbol{s} = \mathbf{1}$ 即$s_{l,\,i} = 1\,\forall l,\;\; i$。请注意，将 Ω_* 替换为

$\Omega_*/\Omega_*(\mathbf{1})$ 等同于将 λ 替换为

$$\lambda' = \lambda/\Omega_*(\mathbf{1}). \tag{20}$$

因此，有时我们会采用归一化表示法 λ'，而非 λ。

4. 实验

本节将评估所提出的脉冲突触惩罚的有效性。首先，我们介绍实验设置。接着，我们展示所提出的方法能够降低能耗。最后，我们展示所提出的方法在保持准确率的同时，比其他方法更能降低能耗。具体而言，我们将说明所提出的方法不受替代梯度函数选择的影响，并且优于转换方法（Sorbaro 等人，2020）。总体而言，我们的主要目标是分析该方法的行为，而非实现最先进的性能。

4.1. 实验设置

由于单步脉冲神经元是为静态任务开发的（Suetake 等人，2023），我们在 Fashion-MNIST（Xiao 等人，2017）、CIFAR-10 和 CIFAR-100（Krizhevsky，2009）数据集上进行了实验。这些数据集广泛用于脉冲神经网络实验（Esser 等人，2016；Zhang& Li，2020；Chowdhury 等人，2021b），并采用了若干种网络架构：CNN7、VGG11 和 ResNet18（详情参见附录A.1 ）。在这些实验中，我们使用 PyTorch 框架实现程序，并使用一块图形处理器 —— NVIDIA GeForce RTX 3090（24 GB 显存）（训练时间差异请参见附录中的表B.1 ）。

在这些实验中，我们使用了以下目标函数：

$$\begin{aligned} L =& \frac{1}{n}\sum_{n=1}^{N}\left(CE(f(x_n),t_n)+\lambda\Omega_{\mathrm{syn}}(\boldsymbol{s}(x_n))\right) \\ &+\lambda_{\mathrm{WD}}\left(\|\boldsymbol{W}\|_{L_2}^2+B_{\mathrm{BN}}\|\boldsymbol{W}_{\mathrm{BN}}\|_{L_2}^2\right), \end{aligned} \tag{21}$$

其中，(x_n, t_n) 表示输入数据及其标签组成的对，f 表示某个脉冲神经网络，CE 表示交叉熵函数，$\|\boldsymbol{W}\|_{L_2}^2$ 表示针对权重的 L_2 惩罚项，即权重衰减，$\|\boldsymbol{W}_{\mathrm{BN}}\|_{L_2}^2$ 表示针对批归一化层可训练参数的 L_2惩罚项，λ 和 λ_{WD} 表示惩罚强度，且$B_{\mathrm{BN}} \in \{0,1\}$。请注意，我们将权重衰减纳入目标函数，以验证其在脉冲神经网络中的稀疏化效果（Yaguchi 等，2018）。此外，我们明确指定了批归一化层的权重衰减，因为 PyTorch 框架的默认设置中包含了该项（Paszke 等人，2019），而 Yaguchi 等人（2018）未包含。f 的训练通过以下方式完成：采用式中替代梯度的反向传播算法。3 除非另有说明。优化器选用动量随机梯度下降（mSGD）或 Adam，以验证 Yaguchi 等人（2018）。

详情请参阅附录A，了解超参数等实验设置的更多信息。

4.2. 脉冲突触惩罚带来的能耗降低

我们研究了优化脉冲突触惩罚（公式 14）是否会使能耗指标（公式 11）也得到优化，以及这是否会与目标函数（公式21）中的其他项产生冲突。设置如下：基线模型针对公式21 进行了训练，使用 $\lambda = \lambda_{\mathrm{WD}} = 0$。其他模型则从头开始训练，并采用权重衰减（$\lambda_{\mathrm{WD}} > 0$）、$L_2$ 批归一化层惩罚（$B_{\mathrm{BN}} = 1$）以及 $p = 1$ 脉冲突触惩罚（$\lambda > 0$）的若干组合。结果见表 2 和3。请注意，表中的数值取自准确率非常低（约 20%）时的点，并持续记录至准确率达到上限且不再变化。

从表2的结果中，我们可以观察到以下几点。首先，随着惩罚项强度的增加，能耗指标会降低；推理准确率也会降低。因此，强度参数 λ 控制着二者之间的权衡。其次，脉冲突触惩罚与权重衰减相结合，可以进一步降低能耗指标。因此，我们建议同时采用这两种方法。此外，我们发现，即使不使用脉冲突触惩罚，权重衰减与 Adam 优化器相结合也会引发神经元稀疏化，尽管其对能耗降低的贡献不如脉冲突触惩罚。更进一步，与 mSGD 优化器相比，Adam 优化器能使神经元稀疏化程度更高，这与 Yaguchi 等人(2018) 的论断一致（另见表 3）。需要注意的是，我们无法观察到 $B_{\mathrm{BN}} = 0$与 1之间存在显著差异。因此，为简化目标函数，在后续实验中我们采用权重衰减，并使用 $B_{\mathrm{BN}} = 0$。最后，以上结果不仅适用于 VGG11，也适用于 CNN7 和 ResNet18（见附录中的表 8–11）。

4.3. 准确率与能效之间的权衡

4.3.1. 惩罚项比较

为了了解不同惩罚项之间的差异如何影响训练结果，我们进行了对比实验。实验设置如下。为了公平比较，我们使用 λ' 记号表示惩罚项的强度参数，而非原始 λ（见式20）。基线模型按照式21 进行训练，设置为 $\lambda' = B_{\mathrm{BN}} = 0$，并对其进行了调优

表2。不同 λ、λ_{WD}和 B_{BN}下的性能。网络架构为VGG11，数据集为CIFAR-10，优化器为Adam。 E_{SNN} 表示脉冲神经网络的能耗指标（式11）， E_{baseline} 表示模型在 E_{SNN} 时的$\lambda = \lambda_{\mathrm{WD}} = B_{\mathrm{BN}} = 0$；死亡率表示死亡神经元数量与神经元总数之比，而 R 表示脉冲发放率（式9）。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	89.09 ± 0.271	100.0 ± 1.530	3.367 ± 0.221	12.11 ± 0.114
1e-08	0	0	89.23 ± 0.141	73.92 ± 0.247	5.935 ± 0.822	10.34 ± 0.050
1e-07	0	0	87.81 ± 0.458	34.07 ± 0.397	22.42 ± 1.005	6.046 ± 0.061
1e-06	0	0	79.75 ± 0.694	11.85 ± 0.492	59.01 ± 1.174	2.176 ± 0.040
1e-05	0	0	21.81 ± 14.46	1.053 ± 1.246	92.43 ± 3.149	0.174 ± 0.198
0	1e-04	0	89.15 ± 0.493	81.07 ± 4.497	9.567 ± 1.005	8.474 ± 0.771
0	1e-03	0	89.00 ± 0.832	68.64 ± 6.714	18.72 ± 1.942	6.802 ± 0.492
0	1e-02	0	84.98 ± 1.455	44.39 ± 3.540	34.13 ± 0.683	4.992 ± 0.448
0	1e-01	0	50.51 ± 36.24	17.90 ± 16.03	70.93 ± 25.66	2.988 ± 2.642
0	1e+00	0	10.00 ± 0.000	0.000 ± 0.000	100.0 ± 0.000	0.000 ± 0.000
1e-08	1e-03	0	89.61 ± 0.026	60.28 ± 1.384	20.20 ± 2.268	6.244 ± 0.155
1e-07	1e-03	0	88.06 ± 0.653	28.72 ± 0.448	35.58 ± 2.255	4.076 ± 0.064
1e-06	1e-03	0	76.00 ± 0.962	7.829 ± 0.076	68.02 ± 0.504	1.378 ± 0.016
1e-05	1e-03	0	10.02 ± 0.029	0.007 ± 0.012	99.84 ± 0.283	0.001 ± 0.002
0	1e-05	1	88.99 ± 0.626	96.29 ± 2.003	6.268 ± 0.786	9.753 ± 0.246
0	1e-04	1	89.87 ± 0.526	85.64 ± 3.412	10.24 ± 1.735	7.457 ± 0.293
0	1e-03	1	86.91 ± 0.315	51.06 ± 1.277	32.79 ± 1.257	4.376 ± 0.138
0	1e-02	1	10.00 ± 0.000	0.000 ± 0.000	100.0 ± 0.000	0.000 ± 0.000
1e-08	1e-04	1	89.41 ± 0.147	45.87 ± 0.202	18.38 ± 0.631	5.643 ± 0.019
1e-07	1e-04	1	86.76 ± 0.957	19.21 ± 0.610	41.97 ± 2.169	2.911 ± 0.043
1e-06	1e-04	1	74.38 ± 0.465	6.582 ± 0.264	71.21 ± 0.608	1.111 ± 0.028
1e-05	1e-04	1	10.00 ± 0.000	0.003 ± 0.005	99.92 ± 0.140	0.001 ± 0.001

$\lambda_{\mathrm{WD}} > 0$ 以获得最高准确率。随后，通过改变 $\lambda' > 0$ 并将 Ω_{syn} （方程）21中的Ω_{total} 或$\Omega_{\mathrm{balance}}$ 替换为从头开始训练其他模型。结果如图 3（A）所示，为以 λ' 为参数的能耗–准确率权衡曲线，其中能耗率是各模型的能耗相对于基线模型能耗的归一化值。请注意，数据点位于图的左上角较为理想。关于 A.3中 λ'的采样方法，另请参阅。此外，定量分析结果见表 4，其中所有指标的分数越高越好：曲线下面积（AUC）、斯皮尔曼等级相关系数（斯皮尔曼）和互信息（MI）。请注意，每项指标的参数代表一个截断参数，低于该准确率的数据会被排除。我们引入截断参数，是因为随着所有方法的强度参数增大，训练往往会中断。定量指标的详情请参阅附录。A.4。

从图 3(A) 和表 4的结果中，我们可以观察到以下几点。首先，对于每个 Ω_*， $p = 1$选项显然优于 $p = 2$选项。因此，我们建议采用 $p = 1$ 选项。请注意，这种差异源于式 17中的反向控制。然而，它影响训练结果的机制仍不清楚，这将留待未来研究。其次，对于 $p = 1$，Ω_{syn} 的权衡曲线最佳，其次依次是Ω_{total} 和$\Omega_{\mathrm{balance}}$。

因此，我们通过实验阐明了系数$\psi_{l,\ i}$对于式 13的重要性；这一问题在Sorbaro 等人(2020)提出的方法中一直未得到解决。最后，上述所有结果不仅适用于 CNN7，也适用于 VGG11 和 ResNet18（见图 5 和附录中的表 12–17 ）。

4.3.2. 替代梯度的独立性

为了了解替代梯度函数的选择如何影响训练结果，我们进行了与第 4.3.1节相同的实验，唯一不同的是函数的选择。我们没有采用公式 3，而是采用了分段线性函数（Esser 等人，2016）和缩放 sigmoid 函数（Pellegrini 等人，2021）作为替代梯度，如下所示：

$$\frac{\partial s}{\partial u} \simeq \max\left(1 - |u - u_{\mathrm{th}}|, 0\right), \tag{22}$$

$$\frac{\partial s}{\partial u} \simeq \frac{\partial \sigma_\alpha}{\partial u}, \tag{23}$$

其中 σ_α 与公式 5相同。结果见附录中的图 3 （B）和（C），以及表 18 和 19 。

从图3（B）和（C）的结果来看，第 4.3.1 节中的观察结果同样成立。也就是说， $p = 1$ 选项显然优于 $p = 2$ 选项；Ω_{syn} 的权衡曲线表现最佳，其次依次是Ω_{total} 和$\Omega_{\mathrm{balance}}$。因此，无论选择哪种替代梯度函数，脉冲突触惩罚均能发挥作用。

表3. 不同 λ、λ_{WD} 和 B_{BN}条件下的性能。优化器为 mSGD。其余说明与表 2相同。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	89.14 ± 0.158	100.0 ± 2.548	1.678 ± 0.167	14.19 ± 0.277
1e-09	0	0	89.36 ± 0.159	78.84 ± 1.166	2.428 ± 0.188	13.35 ± 0.214
1e-08	0	0	88.81 ± 0.224	40.00 ± 0.335	11.04 ± 0.139	10.07 ± 0.097
1e-07	0	0	86.80 ± 0.348	18.58 ± 0.384	33.62 ± 0.875	4.476 ± 0.143
1e-06	0	0	24.32 ± 7.158	73.27 ± 98.19	78.72 ± 21.06	4.207 ± 4.165
0	1e-04	0	88.59 ± 0.413	84.40 ± 6.843	2.705 ± 0.754	12.73 ± 0.344
0	1e-03	0	88.60 ± 2.134	70.96 ± 8.041	3.968 ± 1.326	11.97 ± 0.582
0	1e-02	0	70.80 ± 2.795	45.18 ± 3.008	19.57 ± 2.636	10.85 ± 0.651
0	1e-01	0	20.23 ± 17.72	18.52 ± 13.99	92.18 ± 10.94	3.322 ± 4.056
1e-09	1e-03	0	90.96 ± 0.202	64.62 ± 0.802	3.976 ± 0.196	11.82 ± 0.085
1e-08	1e-03	0	90.06 ± 0.387	35.65 ± 0.076	13.51 ± 0.195	9.043 ± 0.068
1e-07	1e-03	0	87.74 ± 0.257	16.32 ± 0.166	37.68 ± 0.598	4.163 ± 0.040
1e-06	1e-03	0	21.42 ± 9.467	26.13 ± 24.14	87.16 ± 4.910	3.255 ± 1.576
0	1e-07	1	89.30 ± 0.263	99.75 ± 2.576	1.798 ± 0.260	14.17 ± 0.270
0	1e-06	1	88.85 ± 0.277	94.59 ± 2.186	1.838 ± 0.117	13.63 ± 0.108
0	1e-05	1	88.80 ± 0.235	95.85 ± 1.806	1.834 ± 0.134	13.43 ± 0.299
0	1e-04	1	88.51 ± 0.295	87.16 ± 1.614	2.657 ± 0.105	10.12 ± 0.151
0	1e-03	1	84.47 ± 0.757	51.21 ± 2.576	16.90 ± 0.388	5.005 ± 0.310
0	1e-02	1	19.53 ± 16.50	1.786 ± 3.093	97.64 ± 4.088	0.164 ± 0.284
1e-09	1e-07	1	89.32 ± 0.417	78.35 ± 1.694	2.606 ± 0.078	13.24 ± 0.143
1e-08	1e-07	1	88.77 ± 0.181	39.70 ± 0.844	11.29 ± 0.428	10.04 ± 0.173
1e-07	1e-07	1	86.95 ± 0.236	18.60 ± 0.285	33.49 ± 0.418	4.478 ± 0.130
1e-06	1e-07	1	28.05 ± 1.125	78.52 ± 105.4	81.75 ± 13.47	4.500 ± 4.411

表4. 定量对比 对应图3 （A）的比较结果。分数越高越好。最佳和次佳结果分别以粗体和下划线 d
标出，分别 请参阅附录A.4，了解定量指标的详细信息。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\mathrm{syn}}(p=1)$（我们的方法）	**68.02**	**79.60**	**0.9861**	**0.9865**	**3.465**	**3.610**
$\Omega_{\mathrm{syn}}(p=2)$（我们的方法）	61.62	72.69	0.9474	0.9709	3.233	<u>3.476</u>
$\Omega_{\mathrm{total}}(p=1)$	<u>63.30</u>	<u>76.05</u>	0.9766	<u>0.9767</u>	<u>3.244</u>	3.319
$\Omega_{\mathrm{total}}(p=2)$	55.16	64.63	<u>0.9831</u>	0.9701	3.218	3.295
$\Omega_{\mathrm{balance}}(p=1)$	54.23	67.16	0.9412	0.9412	2.978	2.978
$\Omega_{\mathrm{balance}}(p=2)$	31.47	42.94	0.8500	0.8946	2.708	2.833

4.3.3. 优于转换方法

为了了解训练方法的差异如何影响训练结果，我们还绘制了转换方法（Sorbaro 等人，2020）对应的权衡曲线。我们训练了一个 QReLU 网络（具有量化 ReLU 激活的人工神经网络），逐步提高惩罚强度，并评估不同时间步下、各惩罚强度对应的转换后脉冲神经网络：$T=1$，5 和10 （详情请参阅原论文（Sorbaro 等人，2020））。结果见图3 （D）和附录中的表20，其中能耗以基线在 Ω_{syn}下的能耗进行归一化。

从图3 (D) 中的结果可以看出，转换方法的能耗和准确率均低于替代梯度方法。这是因为转换过程会降低准确率，而且 QReLU 网络的惩罚项无法直接优化脉冲神经网络的能耗指标。因此，我们应直接使用替代梯度和脉冲突触惩罚来训练脉冲神经网络，以避免这种性能下降。

5. 结论

我们从替代梯度的角度研究了获取节能型脉冲神经网络的训练方法。基于应直接优化指标本身这一原则，我们推导出了脉冲突触惩罚，以优化能耗指标。随后，我们通过实验表明，脉冲突触惩罚（尤其是对于 $p=1$）优于现有惩罚项和转换方法。此外，其有效性不受网络类型

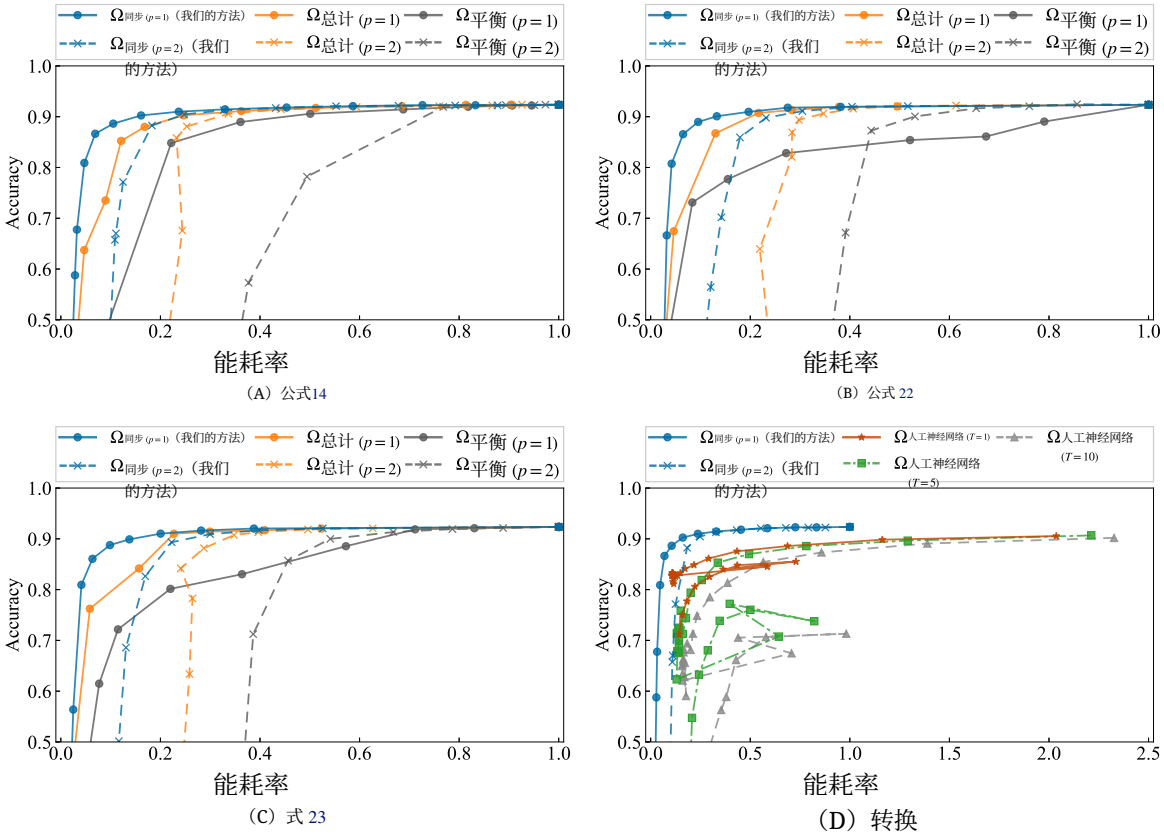

图*3*。 能耗–准确率权衡曲线。网络架构为 CNN7，数据集为 Fashion-MNIST，优化器为 Adam。能耗率是各模型的能耗与基线模型能耗的比值。从（A）到（C）：每个模型均采用各图下方标示的替代梯度进行训练。（D）：与转换方法（Sorbaro 等人，2020）的比较。Ω_{ANN} 表示由 QReLU 网络转换得到的脉冲神经网络。由于转换后的脉冲神经网络的能耗高于脉冲神经网络基线， x轴的数值超过 1，适用于Ω_{syn}。

架构和替代梯度的影响。我们得出结论，我们的原则行之有效。

一个明显的局限性是，脉冲突触惩罚的定义取决于能耗指标的定义。然而，如果目标指标发生变形，则应根据我们的原则相应地调整惩罚，即使该指标与能耗无关。另一个局限性是，尽管目标指标被直接纳入目标函数，但它只是通过替代梯度进行间接优化的。为克服这一问题，我们需要一种替代训练方法，例如平衡训练（Xiao 等人，2021）。

我们进一步列出一些尚待解决的问题。首先，尚不清楚为什么脉冲突触惩罚的$p = 1$ 与 2 的训练结果存在差异。阐明这种差异的机制有助于我们理解替代梯度。其次，我们没有关注脉冲稀疏化与剪枝之间的协同作用。剪枝感知稀疏化训练将有助于我们获得更节能的脉冲神经网络。最后，应验证脉冲突触惩罚的广泛适用性，例如在真实数据集、大型网络和其他任务中的表现。解决这些问题后，我们便能为实现真正环保的脉冲神经网络贡献力量。

为促进进一步研究，我们将很快发布代码。

参考文献

Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., Merolla, P., Imam, N., Nakamura, Y., Datta, P., Nam, G.-J. 等。TrueNorth：一款 65 mW、拥有 100 万个可编程神经元的神经突触芯片的设计与工具流程。IEEE 集成电路与系统计算机辅助设计汇刊，34(10):1537–1557，2015。

Chowdhury, S. S., Garg, I., 和 Roy, K. 面向低延迟脉冲神经网络的时空剪枝与量化

。载于 2021 年国际神经网络联合会议 (IJCNN)，第 1–9 页。IEEE，2021a。

Chowdhury, S. S., Rathi, N., 和 Roy, K. 一步时间步即可满足需求：以超低延迟训练脉冲神经网络。arXiv 预印本 arXiv:2110.05929, 2021b.

Cubuk, E. D., Zoph, B., Shlens, J., 和 Le, Q. V. Randaugment：通过缩减搜索空间实现实用的自动化数据增强。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第 702–703 页，2020.

Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S. 等。Loihi：一种支持片上学习的神经形态众核处理器。IEEE Micro, 38(1):82–99, 2018.

Esser, S. K., Merolla, P. A., Arthur, J. V., Cassidy, A. S., Ap- puswamy, R., Andreopoulos, A., Berg, D. J., McKinstry, J. L., Melano, T., Barch, D. R., 等。用于快速、节能神经形态计算的卷积网络。美国国家科学院院刊，113(41):11441，2016。

Horowitz, M. 1.1 计算的能耗问题（以及我们能为此做些什么）。载于《2014 IEEE国际固态电路会议技术论文摘要》（ISSCC），第 10–14 页，2014。

Kim, Y. 和 Panda, P. 重新审视批归一化：从头开始训练低延迟深度脉冲神经网络。Frontiers in Neuroscience，15:1638，2021。

Krizhevsky, A. 从微小图像中学习多层特征。技术报告，2009。

Kundu, S., Datta, G., Pedram, M., 和 Beerel, P. A. Spike-thrift：通过注意力引导压缩限制脉冲活动，实现节能型深度脉冲神经网络。载于《IEEE/CVF冬季计算机视觉应用会议论文集》，第 3953–3962 页，2021。

Lee, C., Sarwar, S. S., Panda, P., Srinivasan, G., 和 Roy, K. 实现基于脉冲的反向传播，以训练深度神经网络架构。《Frontiers in Neuroscience》，14:119，2020。

Lee, D., Park, S., Kim, J., Doh, W., 和 Yoon, S. 面向脉冲神经网络的节能型知识蒸馏。arXiv 预印本 arXiv:2106.07172，2021。

Loshchilov, I. 和 Hutter, F. SGDR：带热重启的随机梯度下降。发表于 2017 年国际学习表征会议。

Maguire, L. P., McGinnity, T. M., Glackin, B., Ghani, A., Belatreche, A. 和 Harkin, J. 在现场可编程门阵列上大规模实现脉冲神经网络所面临的挑战。《Neurocomputing》，71(1-3):13–29，2007 年。Misra, J. 和 Saha, I. 硬件中的人工神经网络：二十年进展综述。《Neurocomputing》，74(1-3):239–255，2010 年。

Narduzzi, S., Bigdeli, S. A., Liu, S.-C. 和 Dunbar, L. A. 使用活动正则化优化脉冲神经网络的能耗。载于《ICASSP 2022—2022 IEEE 国际声学、语音与信号处理会议（ICASSP）》，第 61–65 页。IEEE，2022 年。

Naya, K., Kutsuzawa, K., Owaki, D. 和 Hayashibe, M. 脉冲神经网络在深度强化学习中发现节能的六足运动。IEEE Access，9:150345–150354，2021 年。

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. 等。PyTorch：一种命令式、高性能深度学习库。神经信息处理系统进展，32:8024–8035，2019 年。https://pytorch.org/。

Pellegrini, T., Zimmer, R., and Masquelier, T. 应用于语音命令识别的低活动监督式卷积脉冲神经网络。载于 2021 年 IEEE口语技术研讨会 (SLT)，第 97–103 页。IEEE，2021。

Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., and Liu, S.-C. 将连续值深度网络转换为用于图像分类的高效事件驱动网络。Frontiers in Neuroscience，11:682，2017。

Shrestha, S. B. and Orchard, G. SLAYER：脉冲层误差时间重分配。神经信息处理系统进展，31:1419–1428，2018。

Sorbaro, M., Liu, Q., Bortone, M., andSheik, S. 面向神经形态应用优化脉冲神经网络的能耗。Frontiers in Neuroscience，14:662，2020。

斯皮尔曼，C. 两个事物之间关联性的证明与测量。《美国心理学杂志》，15：72–101，1904。

Suetake，K.、Ikegawa，S.-i.、Saiin，R. 和 Sawada，Y. S3NN：用于训练高能效单步脉冲神经网络的脉冲替代梯度时间步缩减。《神经网络》，159：208–219，2023。

Xiao, H., Rasul, K., 和 Vollgraf, R. Fashion-MNIST：用于基准测试机器学习算法的新型图像数据集。arXiv 预印本 arXiv:1708.07747，2017。

Xiao, M., Meng, Q., Zhang, Z., Wang, Y., 和 Lin, Z. 通过对平衡态进行隐式微分训练反馈脉冲神经网络。神经信息处理系统进展，34:14516–14528，2021。

Yaguchi, A., Suzuki, T., Asano, W., Nitta, S., Sakata, Y., 和 Tanizawa, A. Adam 在整流器神经网络中诱导出隐式权重稀疏性。载于 2018 年第 17 届 IEEE国际机器学习与应用会议 (ICMLA)，第 318–325 页。IEEE，2018 年。

Zenke, F. 和 Ganguli, S. SuperSpike：多层脉冲神经网络中的监督学习。神经计算，30(6)：1514–1541，2018 年。

Zhang, W. 和 Li, P. 通过反向传播进行时序脉冲序列学习，以构建深度脉冲神经网络。《神经信息处理系统进展》，33：12022–12033，2020 年。

表 5。网络架构。C(i_1)$_{\mathrm{k}(i_2)}$s(i_3)p(i_4) 是一个二维卷积层，其通道大小为 $=i_1$、卷积核大小为 $=i_2$、步幅大小为 $=i_3$、填充大小为 $=i_4$；FC(x) 是一个全连接层，其通道大小为 $=x$；S 是脉冲激活函数（公式 2）；BN 是批归一化层；DO(x) 是一个丢弃层，丢弃率为 $=x$；Avg 是一个 2×2 平均池化层；GAP 是全局平均池化层。ResA 和 ResB 都是特定的残差模块——ResA(x) 由两条路径组成：[BN–S–C(x)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–BN–S–C($x$)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$] 和恒等函数；ResB($x$) 由两条路径组成：[BN–S–C($x$)$_{\mathrm{k}(3)\mathrm{s}(2)\mathrm{p}(1)}$–BN– S–C($x$)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$] 和 [BN–S–C($x$)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$]。

Name	网络架构
CNN7	C(64)$_{\mathrm{k}(3)\mathrm{s}(2)\mathrm{p}(0)}$–批归一化–S–Dropout(0.1)–C(128)$_{\mathrm{k}(6)\mathrm{s}(1)\mathrm{p}(0)}$–批归一化–S–Dropout(0.2)–C(256)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(0)}$–批归一化–S–Dropout(0.3)–C(128)$_{\mathrm{k}(1)\mathrm{s}(1)\mathrm{p}(0)}$–批归一化–S–Dropout(0.2)–C(64)$_{\mathrm{k}(1)\mathrm{s}(1)\mathrm{p}(0)}$–批归一化–S–Dropout(0.1)–C(10)$_{\mathrm{k}(1)\mathrm{s}(1)\mathrm{p}(0)}$–批归一化–S–C(10)$_{\mathrm{k}(1)\mathrm{s}(1)\mathrm{p}(0)}$–全局平均池化
VGG11	C(64)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Dropout(0.2)–C(128)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Avg–C(256)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Avg–C(512)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Dropout(0.2)–C(512)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Avg–C(512)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Avg–C(512)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–Dropout(0.2)–C(512)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–批归一化–S–全连接层(4096)–批归一化–S–Dropout(0.2)–全连接层(4096)–批归一化–S–Dropout(0.2)–全连接层(10)
ResNet18	C(64)$_{\mathrm{k}(3)\mathrm{s}(1)\mathrm{p}(1)}$–{ResA(64)–ResA(64)}–{ResB(128)–ResA(128}–{ResB(256)–ResA(256)}–{ResB(512)–ResA(512)}–批归一化–S–C(10)$_{\mathrm{k}(1)\mathrm{s}(1)\mathrm{p}(0)}$–全局平均池化

A. 实验设置详情

A.1. 网络架构

我们在实验中采用的网络架构如表 5 所示。请注意，在这些网络架构中，$\psi_{l,\,i}$ 独立于 i，且所有二维卷积层和全连接层均不含偏置项。

批归一化层对膜电位（式 1）的影响如下：

$$\boldsymbol{u}_l := \frac{\boldsymbol{\alpha}_l}{\boldsymbol{\sigma}_l}\left(\boldsymbol{W}_l\boldsymbol{s}_{l-1} - \boldsymbol{\mu}_l\right) + \boldsymbol{\gamma}_l, \tag{24}$$

其中，$\boldsymbol{\mu}_l$ 和 $\boldsymbol{\sigma}_l \in \boldsymbol{R}^{d_l}$ 分别表示突触后电流 $\boldsymbol{W}_l\boldsymbol{s}_{l-1}$ 的运行平均值和标准差值，$\boldsymbol{\alpha}_l$ 和 $\boldsymbol{\gamma}_l \in \boldsymbol{R}^{d_l}$ 则是批归一化层的可训练仿射参数。

A.2. 数据集

我们使用了三个数据集作为基准，并将每个数据集划分为训练集、验证集和测试集，具体如下：每个数据集的（#训练集、#验证集、#测试集）分别为 (54000，6000，10000)（Fashion-MNIST），(45000，5000，10000)（CIFAR-10），以及 (45000，5000，10000)（CIFAR-100）。我们采用随机增强作为数据增强技术（Cubuk 等人，2020）。

A.3. 超参数

我们使用了以下超参数。权重采用针对 ReLU 函数的 He 初始化（尽管我们采用的是脉冲神经元），并使用 Adam (β =(0.9, 0.999) $\epsilon = 10^{-8}$) 或 mSGD（momentum$=0.9$）优化器进行优化。小批量大小为 100，训练轮数为 150，脉冲发放阈值为 $u_{\mathrm{th}} = 1$，式 3 中替代梯度的学习率和 (α, τ) 汇总于表 6，学习率则通过余弦退火 ($T_{\max} = 150, \eta_{\min} = 0.0$) 进行调度（Loshchilov 和 Hutter，2017）。请注意，惩罚项 Ω_* 采用线性调度，即将 λ 乘以当前训练轮数与总训练轮数之比。在权衡曲线实验（第4.3节）中，归一化强度参数 λ'（式 20）从 14 种模式中选取——$\{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192\}$。所有实验均使用三个随机种子进行。

表 6. 超参数。除非另有说明，实验中使用了以下超参数。学习率、 λ_{WD}、α τ 和通过网格搜索确定。

模型	优化器	替代梯度	学习率	λ_{WD}	B_{BN}	α	τ
CNN7	Adam	式 3	1e-3	1e-4	0	0.25	0.6
CNN7	Adam	式 22	1e-3	1e-6	0	-	-
CNN7	Adam	式 23	1e-2	1e-7	0	0.45	-
CNN7	mSGD	式 3	1e-2	1e-4	0	0.35	0.6
VGG11	Adam	式 3	1e-3	1e-3	0	0.25	0.6
VGG11	mSGD	公式 3	1e-2	1e-3	0	0.35	0.8
ResNet18	Adam	公式 3	1e-3	1e-4	0	0.35	1.0
ResNet18	mSGD	公式 3	1e-2	1e-3	0	0.35	1.0

算法1 曲线下面积(P)

输入：数据 $X = \{(x_i, y_i)\}_{i=1}^{N} \subset \boldsymbol{R}_{\geq 0}^{2}$，截断值 $P' = P/100$ (
$0 \leq P < 100$)。$X \leftarrow$ 排序$(X, x_{i-1} \leq x_i)$;$X \leftarrow \{(x_i, y_i) \in X \mid x_i \leq 1\}$;
$X \leftarrow (x_0 = x_1, y_0 = 0) \cup X$;$X \leftarrow X \cup (x_\infty = 1; y_\infty =$ 最大值$(\{y_i \in \boldsymbol{R} \mid$
$(x_i, y_i) \in X)\})$;$X \leftarrow \{(x_i, y_i) \in X \mid y_j \leq y_i\ (j < i)\}$;$c \leftarrow$ 的线性插值曲
线;$A(P) \leftarrow$ 在 $y = P'$ 上、范围为 $x \in [0, 1]$ 的 c 曲线下面积；返回 A
$(P)/(1 - P')$。

A.4. 指标

在第4.3节的定量分析中，我们使用了三种定量评分：曲线下面积（曲线下面积）、斯皮尔曼等级相关系数（斯皮尔曼）和互信息（互信息）。评分的参数 P表示该评分是针对准确率高于 P的数据计算得出的。曲线下面积(P)指能耗–准确率权衡曲线下的归一化面积，其中 x轴和 y轴的范围分别为 $[0, 1]$ 和 $[P/100, 1]$。因此，曲线下面积越高，维持较高准确率时的能耗指标就越低。有关曲线下面积的详细计算方法，请参见算法1和图4。请注意，我们的曲线下面积算法会高估非单调曲线，即其他方法的结果往往会被高估。斯皮尔曼等级相关系数 ρ 描述了两个成分与单调函数的拟合程度（斯皮尔曼，1904）。因此，斯皮尔曼 ρ 越高，强度参数作为权衡控制器的适用性就越强。互信息也适合作为权衡控制器的指标，因为它描述了两个成分之间的相互依赖关系，评分越高越好。

B. 实验结果详情

B.1. 训练时间差异

表7 展示了各方法相对于基线（$\lambda = 0$）的训练时间比。从表中可以看出，训练时间没有显著差异。需要注意的是，在其他设置中也观察到了类似趋势。

表 7。训练时间差异 。网络架构为 ResNet18，数据集为 CIFAR-10，优化器为 Adam，且 $p = 1$.

惩罚项	None	Ω_{syn}	Ω_{total}	$\Omega_{\mathrm{balance}}$
训练时间速率	1.000	0.998	0.968	0.999

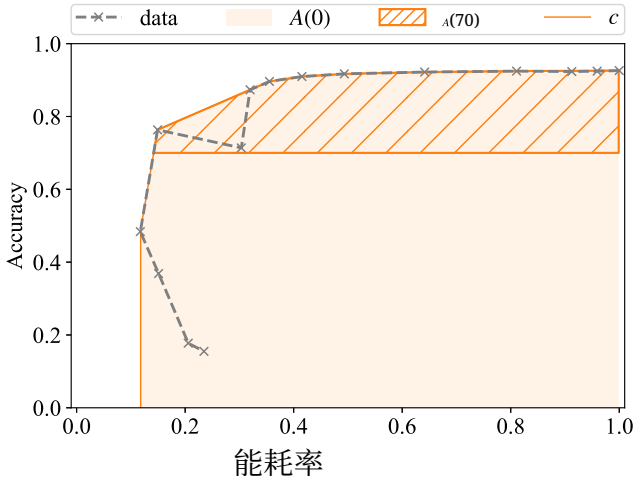

图*4.* 曲线下面积示例(P) 计算（ResNet18 / CIFAR-10 / mSGD）。 $A(P)$ 表示归一化前曲线下方的面积，而 c 表示线性插值曲线。详细定义请参阅算法1 。

B.2. 脉冲突触惩罚的验证

第4.2 节的结果也列于表3–11中。需要注意的是，这些表中的数值取自 λ 和 λ_{WD}准确率非常低（约为20%）时的点，直到准确率达到上限并停止变化。

B.3. 能耗–准确率权衡曲线

第 4.3 节的结果也展示在图 5 和表 12–20中。对于所有方法，mSGD 优化器的训练结果往往比 Adam 优化器的训练结果对惩罚强度更加敏感。

表*8*。 在不同 λ、λ_{WD}和 B_{BN} 条件下的性能。网络架构为 CNN7，数据集为 Fashion-MNIST，优化器为 Adam。其余说明与表2相同。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	89.14 ± 0.158	100.0 ± 2.548	1.678 ± 0.167	14.19 ± 0.277
1e-09	0	0	89.36 ± 0.159	78.84 ± 1.166	2.428 ± 0.188	13.35 ± 0.214
1e-08	0	0	88.81 ± 0.224	40.00 ± 0.335	11.04 ± 0.139	10.07 ± 0.097
1e-07	0	0	86.80 ± 0.348	18.58 ± 0.384	33.62 ± 0.875	4.476 ± 0.143
1e-06	0	0	24.32 ± 7.158	73.27 ± 98.19	78.72 ± 21.06	4.207 ± 4.165
0	1e-04	0	88.59 ± 0.413	84.40 ± 6.843	2.705 ± 0.754	12.73 ± 0.344
0	1e-03	0	88.60 ± 2.134	70.96 ± 8.041	3.968 ± 1.326	11.97 ± 0.582
0	1e-02	0	70.80 ± 2.795	45.18 ± 3.008	19.57 ± 2.636	10.85 ± 0.651
0	1e-01	0	20.23 ± 17.72	18.52 ± 13.99	92.18 ± 10.94	3.322 ± 4.056
1e-09	1e-03	0	90.96 ± 0.202	64.62 ± 0.802	3.976 ± 0.196	11.82 ± 0.085
1e-08	1e-03	0	90.06 ± 0.387	35.65 ± 0.076	13.51 ± 0.195	9.043 ± 0.068
1e-07	1e-03	0	87.74 ± 0.257	16.32 ± 0.166	37.68 ± 0.598	4.163 ± 0.040
1e-06	1e-03	0	21.42 ± 9.467	26.13 ± 24.14	87.16 ± 4.910	3.255 ± 1.576
0	1e-07	1	89.30 ± 0.263	99.75 ± 2.576	1.798 ± 0.260	14.17 ± 0.270
0	1e-06	1	88.85 ± 0.277	94.59 ± 2.186	1.838 ± 0.117	13.63 ± 0.108
0	1e-05	1	88.80 ± 0.235	95.85 ± 1.806	1.834 ± 0.134	13.43 ± 0.299
0	1e-04	1	88.51 ± 0.295	87.16 ± 1.614	2.657 ± 0.105	10.12 ± 0.151
0	1e-03	1	84.47 ± 0.757	51.21 ± 2.576	16.90 ± 0.388	5.005 ± 0.310
0	1e-02	1	19.53 ± 16.50	1.786 ± 3.093	97.64 ± 4.088	0.164 ± 0.284
1e-09	1e-07	1	89.32 ± 0.417	78.35 ± 1.694	2.606 ± 0.078	13.24 ± 0.143
1e-08	1e-07	1	88.77 ± 0.181	39.70 ± 0.844	11.29 ± 0.428	10.04 ± 0.173
1e-07	1e-07	1	86.95 ± 0.236	18.60 ± 0.285	33.49 ± 0.418	4.478 ± 0.130
1e-06	1e-07	1	28.05 ± 1.125	78.52 ± 105.4	81.75 ± 13.47	4.500 ± 4.411

表 *9*。 不同 λ、λ_{WD} 和 B_{BN} 下的性能。网络架构为 CNN7，数据集为 Fashion-MNIST，优化器为 mSGD。其余说明与表 2中的相同。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	92.02 ± 0.142	100.0 ± 1.288	11.17 ± 0.749	14.34 ± 0.643
1e-08	0	0	91.96 ± 0.072	94.65 ± 0.996	11.98 ± 0.602	14.28 ± 0.755
1e-07	0	0	91.62 ± 0.119	61.72 ± 2.295	19.16 ± 1.284	13.12 ± 0.677
1e-06	0	0	90.16 ± 0.225	21.72 ± 0.324	45.37 ± 1.028	10.26 ± 0.781
1e-05	0	0	77.14 ± 0.217	4.651 ± 0.270	79.80 ± 0.902	4.570 ± 0.257
0	1e-05	0	92.11 ± 0.114	98.84 ± 2.827	11.41 ± 0.633	14.44 ± 0.763
0	1e-04	0	92.01 ± 0.130	98.50 ± 0.888	11.90 ± 0.786	14.76 ± 0.774
0	1e-03	0	91.84 ± 0.229	85.42 ± 2.323	14.92 ± 0.359	15.75 ± 0.979
0	1e-02	0	89.70 ± 0.605	73.98 ± 0.855	22.17 ± 0.477	15.02 ± 0.647
0	1e-01	0	62.45 ± 1.900	19.16 ± 1.442	62.08 ± 1.776	13.46 ± 0.259
1e-08	1e-04	0	92.13 ± 0.168	91.49 ± 2.061	12.72 ± 0.450	14.62 ± 0.856
1e-07	1e-04	0	92.01 ± 0.135	61.60 ± 1.440	19.45 ± 0.833	13.56 ± 0.793
1e-06	1e-04	0	90.48 ± 0.420	21.20 ± 0.451	45.99 ± 1.018	10.84 ± 0.665
1e-05	1e-04	0	73.40 ± 7.835	4.413 ± 0.958	80.46 ± 3.399	5.385 ± 1.278
0	1e-05	1	92.15 ± 0.154	100.6 ± 1.111	11.38 ± 0.742	14.44 ± 0.712
0	1e-04	1	92.24 ± 0.021	103.0 ± 1.589	10.89 ± 0.781	14.77 ± 0.877
0	1e-03	1	91.64 ± 0.161	83.64 ± 1.326	16.90 ± 1.336	13.89 ± 1.016
0	1e-02	1	79.16 ± 3.171	30.91 ± 1.941	70.40 ± 1.285	8.008 ± 0.381
0	1e-01	1	10.00 ± 0.000	0.000 ± 0.000	100.0 ± 0.000	0.000 ± 0.000
1e-08	1e-05	1	91.97 ± 0.077	93.33 ± 1.838	12.43 ± 0.928	14.26 ± 0.624
1e-07	1e-05	1	91.86 ± 0.165	61.27 ± 1.850	19.44 ± 0.982	13.24 ± 0.634
1e-06	1e-05	1	90.20 ± 0.170	21.53 ± 0.835	45.55 ± 1.673	10.45 ± 0.608
1e-05	1e-05	1	73.70 ± 2.893	4.297 ± 0.546	81.37 ± 2.846	4.293 ± 0.441

表*10.* 关于不同 λ、λ_{WD}和 B_{BN}的性能。网络架构为 ResNet18，数据集为 CIFAR-10，优化器为 Adam。其余说明与表2相同。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	91.32 ± 0.209	100.0 ± 0.293	1.074 ± 0.033	17.74 ± 0.067
1e-09	0	0	91.27 ± 0.175	93.06 ± 0.486	1.325 ± 0.020	16.75 ± 0.023
1e-08	0	0	91.40 ± 0.293	58.03 ± 0.222	3.846 ± 0.067	11.69 ± 0.055
1e-07	0	0	89.46 ± 0.215	19.13 ± 0.247	25.36 ± 0.825	4.936 ± 0.010
1e-06	0	0	78.76 ± 0.239	4.328 ± 0.192	65.38 ± 0.283	1.629 ± 0.015
1e-05	0	0	19.53 ± 2.170	0.010 ± 0.004	99.36 ± 0.094	0.040 ± 0.008
0	1e-05	0	92.15 ± 0.032	92.79 ± 0.547	2.088 ± 0.077	16.55 ± 0.086
0	1e-04	0	92.21 ± 0.196	79.67 ± 0.721	5.377 ± 0.084	14.22 ± 0.097
0	1e-03	0	91.64 ± 0.135	70.58 ± 0.114	12.08 ± 0.185	12.38 ± 0.064
0	1e-02	0	86.88 ± 0.466	50.96 ± 1.071	32.84 ± 1.270	10.59 ± 0.099
0	1e-01	0	80.94 ± 0.264	26.48 ± 0.815	58.03 ± 1.090	7.163 ± 0.073
0	1e+00	0	60.29 ± 2.386	13.99 ± 0.264	76.32 ± 1.462	5.239 ± 0.136
1e-09	1e-04	0	92.44 ± 0.093	74.74 ± 0.310	5.851 ± 0.116	13.57 ± 0.078
1e-08	1e-04	0	92.12 ± 0.177	52.82 ± 0.206	8.275 ± 0.161	10.20 ± 0.012
1e-07	1e-04	0	90.18 ± 0.135	15.27 ± 0.075	31.96 ± 0.398	3.895 ± 0.063
1e-06	1e-04	0	75.38 ± 0.766	2.385 ± 0.449	74.48 ± 1.881	1.320 ± 0.054
1e-05	1e-04	0	15.95 ± 3.741	0.002 ± 0.001	99.72 ± 0.046	0.032 ± 0.014
0	1e-05	1	91.87 ± 0.065	89.63 ± 0.447	2.467 ± 0.022	15.55 ± 0.043
0	1e-04	1	91.97 ± 0.183	66.36 ± 0.600	10.98 ± 0.076	10.32 ± 0.079
0	1e-03	1	85.93 ± 0.244	23.50 ± 0.551	54.07 ± 0.387	3.321 ± 0.039
0	1e-02	1	19.22 ± 6.452	0.574 ± 0.040	99.19 ± 0.492	0.114 ± 0.011
1e-09	1e-04	1	91.79 ± 0.215	61.58 ± 0.537	11.45 ± 0.088	9.699 ± 0.034
1e-08	1e-04	1	91.85 ± 0.187	38.43 ± 0.512	15.40 ± 0.274	6.616 ± 0.097
1e-07	1e-04	1	88.97 ± 0.227	12.29 ± 0.292	40.56 ± 0.295	2.735 ± 0.039
1e-06	1e-04	1	73.40 ± 0.533	2.153 ± 0.115	77.12 ± 0.622	0.996 ± 0.038
1e-05	1e-04	1	15.79 ± 3.494	0.001 ± 0.001	99.74 ± 0.071	0.027 ± 0.016

表 *11.* 不同 λ、λ_{WD}和 B_{BN} 条件下的性能。网络架构为 ResNet18，数据集为 CIFAR-10，优化器为 mSGD。其余说明与表 2 相同。

λ	λ_{WD}	B_{BN}	准确率 [%]	$E_{\mathrm{SNN}}/E_{\mathrm{baseline}}$ [%]	死亡率 [%]	R [%]
0	0	0	90.15 ± 0.291	100.0 ± 0.583	1.188 ± 0.060	17.40 ± 0.074
1e-09	0	0	90.31 ± 0.294	95.21 ± 0.552	1.275 ± 0.032	16.70 ± 0.086
1e-08	0	0	89.83 ± 0.092	60.41 ± 0.201	3.120 ± 0.072	11.57 ± 0.040
1e-07	0	0	87.86 ± 0.274	16.01 ± 0.067	26.68 ± 0.865	4.135 ± 0.022
1e-06	0	0	54.23 ± 3.671	2.911 ± 2.568	90.89 ± 0.931	1.389 ± 0.410
1e-05	0	0	15.65 ± 3.972	0.613 ± 0.490	99.04 ± 0.247	0.176 ± 0.034
0	1e-04	0	90.56 ± 0.098	98.25 ± 0.253	1.157 ± 0.134	17.26 ± 0.038
0	1e-03	0	92.59 ± 0.150	86.46 ± 0.818	1.380 ± 0.092	15.96 ± 0.113
0	1e-02	0	91.86 ± 0.036	74.17 ± 0.965	3.932 ± 0.277	14.19 ± 0.213
0	1e-01	0	88.24 ± 0.432	61.39 ± 2.329	19.10 ± 1.422	12.66 ± 0.274
0	1e+00	0	14.65 ± 1.956	17.60 ± 1.014	54.54 ± 2.851	4.290 ± 0.438
1e-09	1e-03	0	92.44 ± 0.268	81.41 ± 0.328	1.588 ± 0.028	15.23 ± 0.026
1e-08	1e-03	0	92.48 ± 0.124	48.00 ± 0.160	4.646 ± 0.111	10.37 ± 0.056
1e-07	1e-03	0	89.05 ± 0.245	11.51 ± 0.320	43.32 ± 0.177	3.997 ± 0.050
1e-06	1e-03	0	65.07 ± 1.377	0.794 ± 0.100	91.15 ± 0.777	1.105 ± 0.044
1e-05	1e-03	0	12.85 ± 2.377	0.036 ± 0.051	99.63 ± 0.335	0.065 ± 0.066
0	1e-05	1	89.98 ± 0.142	98.41 ± 0.691	1.163 ± 0.027	17.06 ± 0.091
0	1e-04	1	90.24 ± 0.211	80.90 ± 0.597	1.886 ± 0.058	13.80 ± 0.071
0	1e-03	1	89.84 ± 0.162	32.15 ± 0.172	26.02 ± 0.286	5.147 ± 0.030
0	1e-02	1	10.04 ± 0.064	0.563 ± 0.523	99.68 ± 0.279	0.111 ± 0.099
1e-09	1e-04	1	90.17 ± 0.192	74.70 ± 0.739	2.154 ± 0.056	12.91 ± 0.114
1e-08	1e-04	1	90.08 ± 0.071	44.94 ± 0.286	5.625 ± 0.230	8.558 ± 0.045
1e-07	1e-04	1	88.17 ± 0.151	13.89 ± 0.066	31.53 ± 0.324	3.586 ± 0.016
1e-06	1e-04	1	59.17 ± 3.370	1.119 ± 0.522	91.11 ± 0.966	0.992 ± 0.092
1e-05	1e-04	1	16.01 ± 0.654	1.087 ± 1.670	99.05 ± 1.117	0.236 ± 0.257

表*12.* 对应于（CNN7 / Fashion-MNIST / mSGD）的定量比较。描述与表 4相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\mathrm{syn}}(p=1)$（我们的方法）	**65.67**	**77.18**	**0.9652**	0.7486	**3.355**	**3.454**
$\Omega_{\mathrm{syn}}(p=2)$（我们的方法）	58.19	70.15	0.9387	0.7794	3.332	3.434
$\Omega_{\mathrm{total}}(p=1)$	60.90	74.22	0.9557	0.7705	3.178	3.295
$\Omega_{\mathrm{total}}(p=2)$	52.73	66.27	0.9313	**0.9562**	3.062	3.233
$\Omega_{\mathrm{balance}}(p=1)$	45.42	59.26	0.9100	0.9229	2.871	2.926
$\Omega_{\mathrm{balance}}(p=2)$	28.53	38.09	0.8694	0.8694	2.813	2.813

表 *13*。对应于（VGG11 / CIFAR-10 / Adam）的定量比较。描述与表 4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\mathrm{syn}}(p=1)$（我们的方法）	**52.56**	68.63	**0.9365**	**0.9497**	**3.610**	**3.688**
$\Omega_{\mathrm{syn}}(p=2)$（我们的方法）	46.85	59.73	0.7758	0.8445	3.536	3.671
$\Omega_{\mathrm{total}}(p=1)$	52.21	**68.69**	0.8883	0.9121	3.583	3.663
$\Omega_{\mathrm{total}}(p=2)$	47.38	60.58	0.7400	0.7953	3.506	3.592
$\Omega_{\mathrm{balance}}(p=1)$	41.21	55.66	0.8770	0.8770	3.344	3.344
$\Omega_{\mathrm{balance}}(p=2)$	40.68	53.07	0.8633	0.9006	3.436	3.556

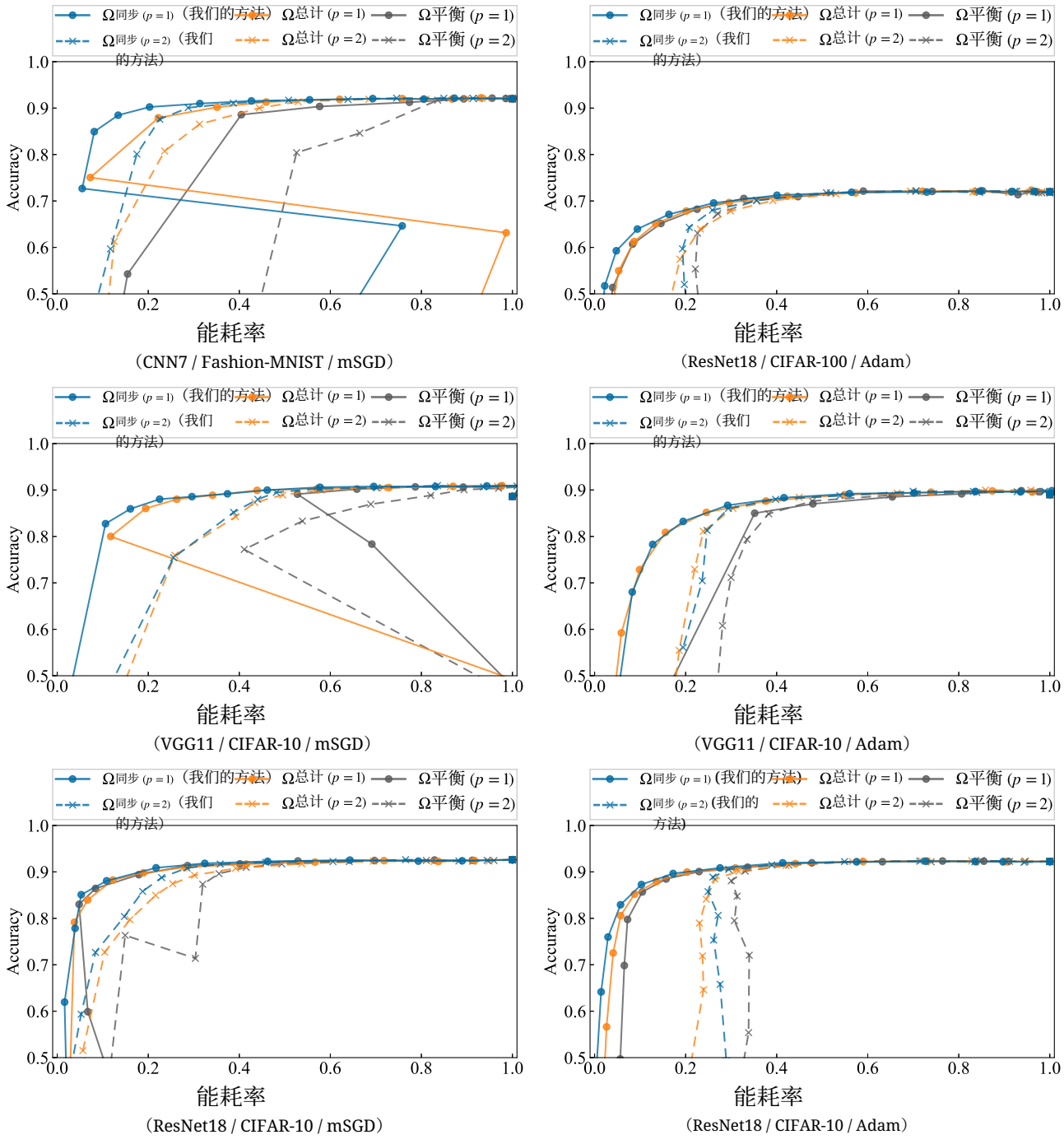

图 *5.* 能耗–准确率权衡曲线。能耗率指各模型的能耗相对于基线模型能耗的归一化值。每幅图下方标注了网络架构 / 数据集 / 优化器。

表*14*。对应于（VGG11 / CIFAR-10 / mSGD）的定量比较。描述与表4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$（我们的方法）	**58.82**	**72.98**	0.9576	0.9613	**3.454**	**3.485**
$\Omega_{\text{syn}}(p=2)$（我们的方法）	45.73	60.29	0.9348	0.9348	3.308	3.308
$\Omega_{\text{total}}(p=1)$	56.62	69.23	**0.9631**	0.9631	3.401	3.401
$\Omega_{\text{total}}(p=2)$	44.82	59.09	0.9373	0.9373	3.401	3.401
$\Omega_{\text{balance}}(p=1)$	35.94	48.02	0.8525	0.4420	3.075	3.163
$\Omega_{\text{balance}}(p=2)$	32.15	46.92	0.9591	**0.9633**	3.295	3.332

表*15*。对应于（ResNet18 / CIFAR-10 / Adam）的定量比较。描述与表4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$ (我们的方法)	**67.68**	**80.07**	0.9021	0.9261	3.308	3.412
$\Omega_{\text{syn}}(p=2)$（我们的方法）	54.60	62.74	0.8169	0.7917	**3.454**	**3.573**
$\Omega_{\text{total}}(p=1)$	66.51	78.67	0.9321	0.9475	3.296	3.400
$\Omega_{\text{total}}(p=2)$	55.08	64.29	0.8751	0.8846	3.412	3.536
$\Omega_{\text{balance}}(p=1)$	65.19	76.63	**0.9484**	**0.9572**	3.434	3.526
$\Omega_{\text{balance}}(p=2)$	51.07	58.70	0.7931	0.7441	3.412	3.506

表*16*。 对应于 (ResNet18 / CIFAR-10 / mSGD) 的定量比较。描述与表4相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$（我们的方法）	**68.85**	**80.56**	0.9009	0.9274	3.244	3.355
$\Omega_{\text{syn}}(p=2)$（我们的方法）	62.15	75.04	**0.9660**	**0.9730**	3.454	3.545
$\Omega_{\text{total}}(p=1)$	67.47	79.15	0.9573	0.9573	3.308	3.308
$\Omega_{\text{total}}(p=2)$	59.02	72.34	0.9469	0.9552	**3.496**	**3.555**
$\Omega_{\text{balance}}(p=1)$	67.57	78.56	0.9389	0.9350	3.295	3.401
$\Omega_{\text{balance}}(p=2)$	56.18	68.44	0.9408	0.9201	3.319	3.389

表*17*。对应于（ResNet18 / CIFAR-100 / Adam）的定量比较。描述与表4中的相同。请注意，带有 $P=70$ 的指标不纳入比较，因为最高准确率约为70%。

方法	曲线下面积(50)[%]	斯皮尔曼(50)	互信息(50)
$\Omega_{\text{syn}}(p=1)$（我们的方法）	**38.98**	0.9091	**3.496**
$\Omega_{\text{syn}}(p=2)$（我们的方法）	33.86	0.8162	3.496
$\Omega_{\text{total}}(p=1)$	38.05	**0.9520**	3.454
$\Omega_{\text{total}}(p=2)$	33.15	0.8143	3.355
$\Omega_{\text{balance}}(p=1)$	38.00	0.8536	3.412
$\Omega_{\text{balance}}(p=2)$	32.69	0.7170	3.401

表*18*。 使用公式对（CNN7 / Fashion-MNIST / Adam）进行的定量比较。 22。说明与表 4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$（我们的方法）	**68.67**	**79.98**	**0.9831**	**0.9863**	**3.091**	**3.258**
$\Omega_{\text{syn}}(p=2)$（我们的方法）	60.19	71.01	0.9608	0.9496	2.833	3.135
$\Omega_{\text{total}}(p=1)$	64.78	77.19	0.9765	0.9835	2.772	2.890
$\Omega_{\text{total}}(p=2)$	52.99	63.03	0.9492	0.9609	3.091	3.178
$\Omega_{\text{balance}}(p=1)$	46.25	65.31	0.7549	0.8211	2.833	2.944
$\Omega_{\text{balance}}(p=2)$	40.83	49.20	0.9588	0.9789	2.772	2.995

表*19*。 使用公式对应于（CNN7 / Fashion-MNIST / Adam）的定量比较 23。描述与表 4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$ (我们的方法)	**68.66**	**80.05**	**0.9935**	**0.9948**	3.044	**3.178**
$\Omega_{\text{syn}}(p=2)$ (我们的方法)	60.04	71.01	0.9676	0.9797	2.772	3.091
$\Omega_{\text{total}}(p=1)$	64.03	76.76	0.9365	0.9460	2.813	2.871
$\Omega_{\text{total}}(p=2)$	54.16	62.96	0.9123	0.9123	**3.075**	3.075
$\Omega_{\text{balance}}(p=1)$	49.28	66.39	0.9236	0.9390	2.890	3.044
$\Omega_{\text{balance}}(p=2)$	39.97	48.82	0.9500	0.9628	2.772	2.890

表*20*。与 Sorbaro 等人 (2020) 的定量比较，对应于（CNN7 / Fashion-MNIST / Adam）。描述与表 4中的相同。

方法	曲线下面积(70)[%]	曲线下面积(50)[%]	斯皮尔曼(70)	斯皮尔曼(50)	互信息(70)	互信息(50)
$\Omega_{\text{syn}}(p=1)$（我们的方法）	**68.02**	**79.60**	**0.9861**	**0.9865**	3.465	3.610
$\Omega_{\text{syn}}(p=2)$（我们的方法）	61.62	72.69	0.9474	0.9709	3.233	3.476
ANN2SNN（$T=1$）	53.51	67.80	0.6863	0.6275	**4.316**	4.355
ANN2SNN（$T=5$）	47.40	63.17	0.8548	0.7015	4.030	4.385
ANN2SNN（$T=10$）	37.93	56.30	0.8602	0.6044	3.784	**4.406**