面向高准确率与超低延迟脉冲神经网络的最优人工神经网络-脉冲神经网络转换

Tong Bu[1], Wei Fang[1], Jianhao Ding[1], PengLin Dai[2], Zhaofei Yu[1 *], Tiejun Huang[ll] 北京大学, [2] 西南交通大学 * 通讯作者：yuzf12@pku.edu.cn

arXiv:2303.04347v1 [cs.NE] 8 Mar 2023

摘要

脉冲神经网络（SNN）因其独特优势而备受关注：在神经形态硬件上功耗低、推理速度快。作为构建深度脉冲神经网络最有效的方法，人工神经网络-脉冲神经网络转换在大规模数据集上取得了与人工神经网络相当的性能。尽管如此，为了使脉冲神经网络的发放率与人工神经网络的激活相匹配，该方法仍需要较长的时间步。因此，时间步较短时，转换后的脉冲神经网络会出现严重的性能下降问题，从而妨碍脉冲神经网络的实际应用。本文从理论上分析了人工神经网络-脉冲神经网络转换误差，并推导出脉冲神经网络的估计激活函数。随后，我们提出量化-裁剪-取整-偏移激活函数，以替代源人工神经网络中的 ReLU激活函数，从而更好地逼近脉冲神经网络的激活函数。我们证明，脉冲神经网络与人工神经网络之间的期望转换误差为零，因此能够实现高准确率、超低延迟的脉冲神经网络。我们在 CIF AR-10/100 和 ImageNet 数据集上评估了该方法，结果表明，无论准确率还是时间步数，该方法均优于最先进的人工神经网络-脉冲神经网络转换方法和直接训练的脉冲神经网络。据我们所知，这是首次探索超低延迟（4 个时间步）下的高性能人工神经网络-脉冲神经网络转换。代码见 https://github.com/putshua/SNN conversion QCFS – –

1 引言

脉冲神经网络（SNNs）是基于生物神经元动态特性的、具有生物学合理性的神经网络（McCulloch 和 Pitts，1943；Izhikevich，2003）。作为人工神经网络的第三代（Maass，1997），脉冲神经网络凭借其相较于深度模拟神经网络（ANNs）的独特特性，受到广泛关注（Roy 等人，2019）。每个神经元在超过阈值时，都会传递离散脉冲以传递信息。对于大多数脉冲神经网络，脉冲神经元会在 T 个推理时间步内累积上一层的电流，并将其作为输出。二值化激活促成了神经形态计算专用硬件的发展（Pei 等人，2019；DeBole 等人，2019；Davies 等人，2018）。这类硬件在时间分辨率和能量预算方面具有显著优势。现有研究已表明，这种方法有望在大幅提升推理速度的同时节省大量能量（St¨ockl 和 Maass，2021）。

除了效率优势之外，近年来脉冲神经网络的学习算法也取得了长足进步。通过时间反向传播和人工神经网络到脉冲神经网络的转换技术训练的脉冲神经网络，在大规模数据集上的性能逐渐可与人工神经网络相媲美（Fang 等，2021；Rueckauer 等，2017）。这两种技术都受益于脉冲神经网络推理时间的设定。在反向传播中设置较长的时间步，可以使代理函数的梯度更加可靠（Wu 等，2018；Neftci 等，2019；Zenke 和 Vogels，2021）。然而，代价是训练期间会消耗大量资源。现有的基于 CUDA 的平台（如 TensorFlow 和 PyTorch ）对脉冲神经网络训练的优化有限。相比之下，人工神经网络-脉冲神经网络转换通常依赖较长的推理时间，才能获得与原始人工神经网络相当的准确率（Sengupta 等，2019），这是因为该方法基于 ReLU 激活与积分发放模型发放率之间的等价性（Cao 等，2015）。尽管延长推理时间可以进一步降低转换误差，但也会阻碍脉冲神经网络在神经形态芯片上的实际应用。

人工神经网络-脉冲神经网络转换面临的困境在于，转换理论中仍存在难以在少数几个时间步内消除的剩余电位（Rueckauer 等，2016）。尽管已有许多方法被提出以提高转换准确率，例如权重归一化（Diehl 等，2015）、阈值重缩放（Sengupta 等，2019）、软重置（Han 和 Roy，2020）以及阈值偏移（Deng & Gu，2020），但基线方法中数十到数百个时间步的要求仍令人难以承受。为获得具有超低延迟（例如 4 个时间步）的高性能脉冲神经网络，我们列出了人工神经网络-脉冲神经网络转换中的关键误差，并针对每种误差提出解决方案。我们的主要贡献总结如下：

- 我们深入分析了人工神经网络-脉冲神经网络转换中的误差，并将其归因于截断误差、量化误差和不均匀误差。我们发现，以往工作忽略的不均匀误差是由到达脉冲时序的变化引起的，可能导致脉冲数多于或少于预期。
- 我们提出量化截断-取整-平移激活函数，以替代源人工神经网络中的 ReLU 激活函数，从而更好地近似脉冲神经网络的激活函数。我们证明，脉冲神经网络与人工神经网络之间的期望转换误差为零，这表明我们能够在超低时间步下实现高性能的脉冲神经网络转换结果。
- 我们在 CIFAR-10、CIFAR-100 和 ImageNet 数据集上评估了所提方法。与人工神经网络-脉冲神经网络转换方法和反向传播训练方法相比，所提方法以更少的时间步实现了最先进的准确率。例如，我们仅用前所未有的 2 个时间步，就在 CIFAR-10 上达到了 91.18% 的 Top-1 准确率。

2 预备知识

In t本节首先简要回顾脉冲神经网络和人工神经网络的神经元模型。然后我们介绍 e the 人工神经网络-脉冲神经网络转换的基本框架。

人工神经网络的神经元模型。对于人工神经网络，模拟神经元的计算可以简化为线性变换与非线性映射的组合：

$$\boldsymbol{a}^l = h(\boldsymbol{W}^l \boldsymbol{a}^{l-1}), \quad l = 1, 2, ..., M \tag{1}$$

whe其中，向量 $\boldsymbol{a}^l$ 表示第 l 层所有神经元的输出， $\boldsymbol{W}^l$ 表示权重矩阵 x bet层 l 与层 $l-1$ 之间的权重矩阵，且 $h(\cdot)$ 是 ReLU 激活函数。

脉冲神经网络的神经元模型。与以往工作（Cao 等，2015；Diehl 等，2015；Han 等，2020）类似，我们考虑脉冲神经网络的积分发放（IF）模型。如果第 l 层的 IF神经元接收到来自上一层的输入 $\boldsymbol{x}^{l-1}(t)$，则 IF神经元的时序电位可定义为：

$$\boldsymbol{m}^l(t) = \boldsymbol{v}^l(t-1) + \boldsymbol{W}^l \boldsymbol{x}^{l-1}(t), \tag{2}$$

其中， $\boldsymbol{m}^l(t)$ 和 $\boldsymbol{v}^l(t)$ 分别表示时间步 t 触发脉冲前后的膜电位。 $\boldsymbol{W}^l$ 表示第 l 层的权重。当 $\boldsymbol{m}^l(t)$ 中的任意元素 $m_i^l(t)$ 超过发放阈值 θ^l 时，神经元便会发放脉冲，并更新膜电位 $v_i^l(t)$。为避免信息损失，我们采用“减阈值复位”机制（Rueckauer 等人，2017年；Han 等，2020），而非“归零复位”机制；也就是说，神经元发放后，膜电位 $v_i^l(t)$ 会减去阈值 θ^l。基于上述阈值触发发放机制以及放电后膜电位的“减阈值复位”机制，我们可以将膜电位的更新规则写为：

$$\boldsymbol{s}^l(t) = H(\boldsymbol{m}^l(t) - \boldsymbol{\theta}^l), \tag{3}$$

$$\boldsymbol{v}^l(t) = \boldsymbol{m}^l(t) - \boldsymbol{s}^l(t)\theta^l. \tag{4}$$

这里， $\boldsymbol{s}^l(t)$表示 l 层中所有神经元在时间 t的输出脉冲；若某个元素对应的神经元发放了脉冲，则该元素为1，否则为0。 $H(\cdot)$是海维赛德阶跃函数。 $\boldsymbol{\theta}^l$ 是发放阈值 θ^l构成的向量。

与 Deng 和 Gu（2020）类似，我们假设如果 $l-1$层中的突触前神经元发放了一个脉冲，那么 l层中的突触后神经元就会接收到未加权的突触后电位 θ^l，即：

$$\boldsymbol{x}^l(t) = \boldsymbol{s}^l(t)\theta^l. \tag{5}$$

表 1：本文符号汇总

符号	定义	符号	定义
l	层索引	$\boldsymbol{x}^l(t)$	未加权的突触后电位[1]
i	神经元索引	$\boldsymbol{s}^l(t)$	输出脉冲
$\boldsymbol{W}^l$	权重	$\boldsymbol{\phi}^l(T)$	时间 T 之前的平均未加权突触后电位
$\boldsymbol{a}^l$	人工神经网络激活值	$\boldsymbol{z}^l$	来自 $l-1$ 层的加权输入
t	时间步	$h(\cdot)$	ReLU函数
T	总时间步	$H(\cdot)$	海维赛德阶跃函数
$\boldsymbol{\theta}^l$	阈值	L	人工神经网络的量化步长
λ^l	人工神经网络中的可训练阈值	$\boldsymbol{Err}^l$	转换误差
$\boldsymbol{m}^l(t)$	放电前膜电位	$\widetilde{\boldsymbol{Err}}^l$	估计转换误差
$\boldsymbol{v}^l(t)$	放电后膜电位	φ	量化截断取整函数的偏移

[1] 突触后电位

人工神经网络-脉冲神经网络转换。其核心思想是将人工神经网络中模拟神经元的激活值映射为脉冲神经网络中脉冲神经元的发放率（或平均突触后电位）。具体而言，结合方程2–方程4，可以得到电位更新方程：

$$\boldsymbol{v}^l(t) - \boldsymbol{v}^l(t-1) = \boldsymbol{W}^l\boldsymbol{x}^{l-1}(t) - \boldsymbol{s}^l(t)\theta^l. \tag{6}$$

公式 6 描述了人工神经网络-脉冲神经网络转换中所用脉冲神经元的基本函数。将公式6从时间1到 T 求和，并将两边除以 T，可得：

$$\frac{\boldsymbol{v}^l(T) - \boldsymbol{v}^l(0)}{T} = \frac{\boldsymbol{W}^l\sum_{i=1}^{T}\boldsymbol{x}^{l-1}(i)}{T} - \frac{\sum_{i=1}^{T}\boldsymbol{s}^l(i)\theta^l}{T}. \tag{7}$$

若用 $\boldsymbol{\phi}^{l-1}(T) = \sum_{i=1}^{T}\boldsymbol{x}^{l-1}(i)\,T$ 表示从0到 T 这段时间内的平均突触后电位，并将公式5代入公式7，则可得：

$$\boldsymbol{\phi}^l(T) = \boldsymbol{W}^l\boldsymbol{\phi}^{l-1}(T) - \frac{\boldsymbol{v}^l(T) - \boldsymbol{v}^l(0)}{T}. \tag{8}$$

公式 8 描述了相邻层神经元平均突触后电位之间的关系。注意，$\boldsymbol{\phi}^l(T) \geqslant 0$。若将初始电位 $\boldsymbol{v}^l(0)$设为零，并忽略仿真时间步 T 足够长时的剩余项$\boldsymbol{v}^l(T)\,T$，则转换后的脉冲神经网络具有与源人工神经网络（公式 1）几乎相同的激活函数。然而，较高的 T 会导致较长的推理延迟，从而阻碍脉冲神经网络的实际应用。因此，本文旨在实现具有极低延迟的高性能人工神经网络-脉冲神经网络转换。

3 转换误差分析

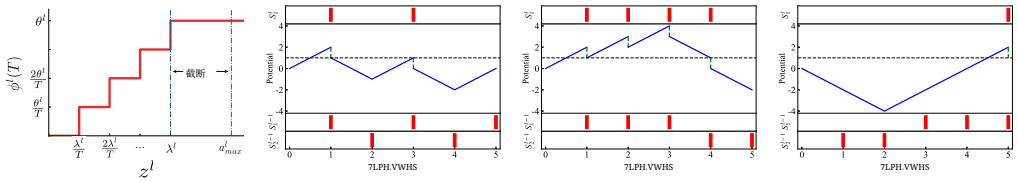
本节将详细分析源人工神经网络与转换后的脉冲神经网络在各层之间的转换误差。下文中，我们假设人工神经网络和脉冲神经网络从层 $l-1$ 接收相同的输入，即 $\boldsymbol{a}^{l-1} = \boldsymbol{\phi}^{l-1}(T)$，然后分析层 l 中的误差。为简单起见，我们用 $\boldsymbol{z}^l = \boldsymbol{W}^l\boldsymbol{\phi}^{l-1}(T) = \boldsymbol{W}^l\boldsymbol{a}^{l-1}$ 来代替人工神经网络和脉冲神经网络中来自层 $l-1$ 的加权输入。绝对转换误差恰好等于转换后的脉冲神经网络的输出减去人工神经网络的输出：

$$\boldsymbol{Err}^l = \boldsymbol{\phi}^l(T) - \boldsymbol{a}^l = \boldsymbol{z}^l - \frac{\boldsymbol{v}^l(T) - \boldsymbol{v}^l(0)}{T} - h(\boldsymbol{z}^l), \tag{9}$$

其中 $h(\boldsymbol{z}^l) = \text{ReLU}(\boldsymbol{z}^l)$。由公式 9 可知，当 $\boldsymbol{v}^l(T) - \boldsymbol{v}^l(0) \neq 0$ 且 $\boldsymbol{z}^l > 0$ 时，转换误差非零。事实上，转换误差由三个因素造成。

截断误差。脉冲神经网络的输出 $\boldsymbol{\phi}^l(T)$ 范围为 $[0, \theta^l]$，因为 $\boldsymbol{\phi}^l(T) = \sum_{i=1}^{T}\boldsymbol{x}^l(i)T = \sum_{i=1}^{T}\boldsymbol{s}^l(i)T\,\theta^l$（参见方程5）。然而，人工神经网络的输出 $\boldsymbol{a}^l$ 范围要大得多，为 $[0, a^l_{max}]$，其中 a^l_{max} 表示 $\boldsymbol{a}^l$ 的最大值。如图 1a 所示，$\boldsymbol{a}^l$ 可通过以下方程映射为 $\boldsymbol{\phi}^l(T)$：

$$\boldsymbol{\phi}^l(T) = \text{clip}\left(\frac{\theta^l}{T}\left\lfloor\frac{\boldsymbol{a}^l T}{\lambda^l}\right\rfloor, 0, \theta^l\right). \tag{10}$$



(a) 截断误差 (b) 均匀脉冲 （c）更多脉冲 （d）更少脉冲

图 1：源人工神经网络与转换后的脉冲神经网络之间的转换误差。s_1^{l-1}和s_2^{l-1}表示第 $l-1$层两个神经元的输出脉冲， s_1^l 表示第 l层一个神经元的输出脉冲。

此处，裁剪函数将上界设为 θ^l，下界设为 0。 $\lfloor \cdot \rfloor$ 表示向下取整函数。 λ^l 表示输出 $\boldsymbol{a}^l$ 的实际最大值，该输出被映射到 $\boldsymbol{\phi}^l(T)$ 的最大值 θ^l。考虑到人工神经网络中近 99.9% 的激活值都处于 $[0, \frac{a_{max}^l}{3}]$ 范围内，Rueckauer 等人（2016）建议根据 99.9% 的激活值来选择 λ^l。人工神经网络中介于 λ^l 和 a_{max}^l 之间的激活值，在脉冲神经网络中会被映射为相同的值 θ^l，这会导致一种称为截断误差的转换误差。

量化误差（向下取整误差）。输出脉冲 $\boldsymbol{s}^l(t)$ 是离散事件，因此 $\boldsymbol{\phi}^l(T)$ 具有量化分辨率 $\frac{\theta^l}{T}$，同样是离散的（见公式 10）。当将 $\boldsymbol{a}^l$ 映射到 $\boldsymbol{\phi}^l(T)$ 时，不可避免地会产生量化误差。例如，如图 1a 所示，范围在 $[\frac{\lambda^l}{T}, \frac{2\lambda^l}{T})$ 内的人工神经网络激活值会被映射到脉冲神经网络的同一个值 $\frac{\theta^l}{T}$。

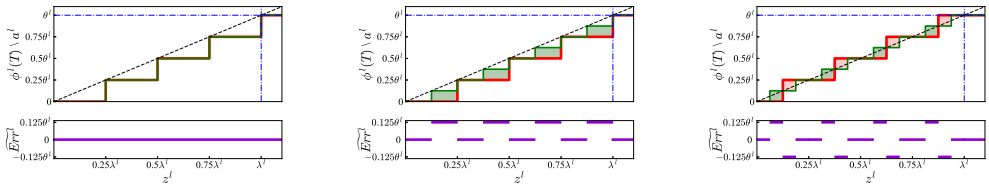
不均匀误差。不均匀误差是由输入脉冲的不均匀性引起的。如果到达脉冲的时间发生变化，输出发放率可能会发生变化，从而导致转换误差。存在两种情况：脉冲数多于预期或少于预期。为说明这一点，在源人工神经网络中，假设第 $l-1$ 层的两个模拟神经元与第 l 层的一个模拟神经元相连，权重分别为 2 和 -2，并且第 $l-1$ 层神经元的输出向量 $\boldsymbol{a}^{l-1}$ 为 $[0.6, 0.4]$。此外，在转换后的脉冲神经网络中，假设第 $l-1$ 层的两个脉冲神经元分别在 5 个时间步（T=5）内发放 3 个脉冲和 2 个脉冲，且阈值为 $\theta^{l-1}=1$。因此， $\boldsymbol{\phi}^{l-1}(T) = \sum_{i=1}^{T} \boldsymbol{s}^{l-1}(i)$ $\theta^{l-1} = [0.6, 0.4]$。尽管 $\boldsymbol{\phi}^{l-1}(T) = \boldsymbol{a}^{l-1}$，并且人工神经网络与脉冲神经网络的权重相同，但如果到达脉冲的时间发生变化， $\boldsymbol{\phi}^l(T)$ 可能不同于 $\boldsymbol{a}^l$。根据公式 1，人工神经网络输出 $\boldsymbol{a}^l = \boldsymbol{W}^l \boldsymbol{a}^{l-1} = [2, -2][0.6, 0.4]^T = 0.4$。对于脉冲神经网络，假设阈值为 $\theta^l = 1$，则存在三种可能的输出发放率，如图 1（b）—（d）所示。如果两个突触前神经元分别以 $t=1$、3、5 和 $t=2$、4（红色条）的发放率发放脉冲，权重为 2 和 -2，则 $\frac{1}{T}\sum_{i=1}^{T} \boldsymbol{\phi}^l(i) = 0$（图 1b）。如果突触前神经元分别在 $t=1$、2、3 和 $t=4$、5 处发放脉冲，那么突触后神经元将在 $t=1$、2、3、4 和 $t=5$ 处发出四个脉冲 $\boldsymbol{\phi}^l(T) = 0.8 > \boldsymbol{a}^l$。如果突触前神经元分别在 $t=3$、4、5 和 $t=1$、2 处发放脉冲，那么突触后神经元将在 $t=5$ 处仅发出一个脉冲，并且 $\boldsymbol{\phi}^l(T) = 0.2 < \boldsymbol{a}^l$。

请注意，截断误差和量化误差由 Li等（2021）提出。上述三种误差之间存在相互依赖关系。具体而言，如果 $\boldsymbol{v}^l(T)$ 处于 $[0, \theta^l]$ 的范围内，不均匀误差将退化为量化误差。假设电位 $\boldsymbol{v}^l(T)$ 落入 $[0, \theta^l]$，即可在忽略不均匀误差影响的情况下估计脉冲神经网络的激活函数。因此，转换后的脉冲神经网络中的输出值 $\boldsymbol{\phi}^l(T)$ 可通过结合裁剪函数和向下取整函数来估计，即：

$$\boldsymbol{\phi}^l(T) \approx \theta^l \operatorname{clip}\left(\frac{1}{T}\left\lfloor \frac{\boldsymbol{z}^l T + \boldsymbol{v}^l(0)}{\theta^l} \right\rfloor, 0, 1\right). \tag{11}$$

详细推导见附录。借助对脉冲神经网络输出的这一估计，可以从公式 9 推导出$\widetilde{\ }^l$估计的转换误差 $\boldsymbol{Err}$：

$$\widetilde{\boldsymbol{Err}}^l = \theta^l \operatorname{clip}\left(\frac{1}{T}\left\lfloor \frac{\boldsymbol{z}^l T + \boldsymbol{v}^l(0)}{\theta^l} \right\rfloor, 0, 1\right) - h(\boldsymbol{z}^l) \approx \boldsymbol{Err}^l. \tag{12}$$



(a) $L = T = 4$　　(b) $L = 4$、$T = 8$　　(c) $L = 4$、$T = 8$、$\varphi = 0.5$

图2：具有相同输入 $\boldsymbol{z}^l$ 时，脉冲神经网络输出 $\boldsymbol{\phi}^l(T)$ 与人工神经网络输出 $\boldsymbol{a}^l$ 的比较

4 最优人工神经网络-脉冲神经网络转换

4.1 量化截断激活函数

根据方程12中的转换误差，自然可以认为，如果将常用的 ReLU激活函数 $h(\boldsymbol{z}^l)$ 替换为具有给定量化步数 L 的裁剪取整函数（类似于方程11），则时间步 $T = L$ 上的转换误差将被消除。这样，低延迟下的性能下降问题就能得到解决。如公式 13 所示，我们提出了量化裁剪取整激活函数来训练人工神经网络。

$$\boldsymbol{a}^l = \bar{h}(\boldsymbol{z}^l) = \lambda^l \operatorname{clip}\left(\frac{1}{L}\left\lfloor\frac{\boldsymbol{z}^l L}{\lambda^l}\right\rfloor, 0, 1\right), \tag{13}$$

其中，超参数 L 表示人工神经网络的量化步数，可训练参数 λ^l 决定了人工神经网络中

$\boldsymbol{a}^l$　　$\boldsymbol{\phi}^l\ T$　　$\boldsymbol{z}^l$ =映射到脉冲神经网络中 ( ) 最大值的最

$\boldsymbol{W}^l\boldsymbol{\phi}^{l-1}\ T\ = \boldsymbol{W}^l\boldsymbol{a}^{l-1}$

大值。注意，( )。采用这一新的激活函数，我们可以证明脉冲神经网络与人工神经网络之间的估计转换误差为零，并得到以下定理。

定理1. 使用激活函数 (13) 的人工神经网络可转换为权重相同的脉冲神经网络。如果$T = L$、 $\theta^l = \lambda^l$和 $\boldsymbol{v}^l(0) = \boldsymbol{0}$，则:

$$\widetilde{\boldsymbol{Err}}^l = \boldsymbol{\phi}^l(T) - \boldsymbol{a}^l = \boldsymbol{0}. \tag{14}$$

证明。根据公式 12，以及条件 $T = L$、 $\theta^l = \lambda^l$、 $\boldsymbol{v}^l(0) = \boldsymbol{0}$，我们有 $\widetilde{\boldsymbol{Err}}^l = \boldsymbol{\phi}^l(T) - \boldsymbol{a}^l = \theta^l$截断 $\left(\frac{1}{T}\left\lfloor\frac{\boldsymbol{z}^l T + \boldsymbol{v}^l(0)}{\theta^l}\right\rfloor, 0, 1\right) - \lambda^l$截断 $\left(\frac{1}{L}\left\lfloor\frac{\boldsymbol{z}^l L}{\lambda^l}\right\rfloor, 0, 1\right) = 0$。 □

定理1表明，如果转换后的脉冲神经网络的时间步 T 与源人工神经网络的量化步数 L 相同，则转换误差为零。图2a给出了一个示例，其中 $T = L = 4, \theta^l = \lambda^l$。红色曲线表示转换后的脉冲神经网络在不同输入 $\boldsymbol{z}^l$下的估计输出 $\boldsymbol{\phi}^l(T)$，绿色曲线则表示源人工神经网络在不同输入 $\boldsymbol{z}^l$下的输出 $\boldsymbol{a}^l$。由于两条曲线相同，估计的转换$\frown$ l误差 $\boldsymbol{Err}$ 为零。然而，在实际应用中，我们关注脉冲神经网络在不同时间步下的性能。无法保证当 T不等于 L时转换误差为零。如图 2b 所示，其中 $L = 4$ 和 $L = 8$，我们可以发现，对于某些 $\boldsymbol{z}^l$，转换误差大于零。这种误差会逐层传递，最终降低转换后的脉冲神经网络的准确率。解决这一问题的一种方法是，针对不同的量化步数训练多个源人工神经网络，再将它们转换为具有不同时间步的脉冲神经网络，但这样做的成本相当高。下一节中，我们将提出一种带偏移项的量化裁剪取整激活函数来解决这一问题。这种方法无需增加额外的计算成本，即可在不同时间步下实现高准确率。

4.2 量化裁剪-下限-偏移激活函数

我们提出量化裁剪-取整-偏移激活函数，用于训练人工神经网络。

$$\boldsymbol{a}^l = \widehat{h}(\boldsymbol{z}^l) = \lambda^l \operatorname{clip}\left(\frac{1}{L}\left\lfloor\frac{\boldsymbol{z}^l L}{\lambda^l} + \boldsymbol{\varphi}\right\rfloor, 0, 1\right). \tag{15}$$

与公式 13 相比，这里存在一个超参数向量 $\boldsymbol{\varphi}$，用于控制激活函数的偏移。当 $L \neq T$ 时，我们无法保证转换误差为 0。不过，我们可以估计转换误差的期望。与（Deng & Gu, 2020）类似，我们假设 z_i^l 在区间 $[(t-1)\lambda^l/T、(t)\lambda^l/T]$ 和 $[(l-1)\lambda^l/L、(l)\lambda^l/L]$ 内均匀分布，对于 $t=1、2、...、T$ 以及 $L=1、2、...、L$，可得以下定理。

定理 2。使用激活函数 (15) 的人工神经网络可转换为权重相同的脉冲神经网络。若 $\theta^l=\lambda^l$, $\boldsymbol{v}^l(0)=\theta^l\boldsymbol{\varphi}$，则对于任意 T和 L，当源人工神经网络中的偏移项 $\boldsymbol{\varphi}$ 满足 1 2 时，转换误差的期望达到 0。

$$\forall\, T, L \quad \mathbb{E}_z\left(\widetilde{\boldsymbol{Err}}^l\right)\Big|_{\boldsymbol{\varphi}=\frac{1}{2}}=\mathbf{0}. \tag{16}$$

证明见附录。定理 2 表明，偏移项 1 2 能够优化转换误差的期望。通过比较图 2b 和图 2c，我们可以发现，加入偏移项 $\boldsymbol{\varphi}=\mathbf{0.5}$ 后，即使 $L \neq T$，平均转换误差也会达到零。这些结果表明，我们能够在超低时间步数下实现高性能的转换后脉冲神经网络。

L 是量化截断移位激活函数唯一未确定的超参数。当$T=L$时，转换误差为零。因此我们自然会认为，参数 L 应尽可能设小，以便在较少时间步下获得更好的性能。然而，激活函数的量化程度过低会降低模型容量，并在时间步数相对较大时进一步造成精度损失。选择合适的 L，是在低延迟下的准确率与脉冲神经网络的最佳准确率之间进行权衡。我们将在实验部分进一步分析量化步数 L的影响。

4.3 量化截断-取整-平移激活函数的训练算法

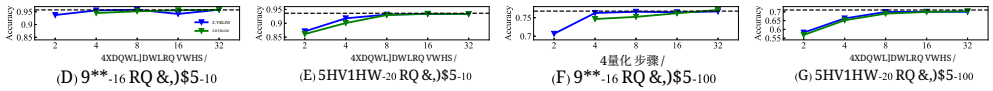
使用量化截断-取整-平移激活函数代替 ReLU 来训练人工神经网络同样是一个棘手的问题。为了直接训练人工神经网络，我们使用直通估计器（Bengio 等，2013）来估计向下取整函数的导数，即 $\frac{\mathrm{d}\lfloor x \rfloor}{\mathrm{d}x}=1$。完整的求导规则见公式 17。

$$\frac{\partial \widehat{h}_i(\boldsymbol{z}^l)}{\partial z_i^l}=\begin{cases}1, & -\frac{\lambda^l}{2L}<z_i^l<\lambda^l-\frac{\lambda^l}{2L}\\ 0, & \text{otherwise}\end{cases}, \frac{\partial \widehat{h}_i(\boldsymbol{z}^l)}{\partial \lambda^l}=\begin{cases}\frac{\widehat{h}_i(\boldsymbol{z}^l)-z_i^l}{\lambda^l}, & -\frac{\lambda^l}{2L}\leqslant z_i^l<\lambda^l-\frac{\lambda^l}{2L}\\ 0, & z_i^l<-\frac{\lambda^l}{2L}\\ 1, & z_i^l\geqslant\lambda^l-\frac{\lambda^l}{2L}\end{cases} \tag{17}$$

这里，z_i^l 是 $\boldsymbol{z}^l$的第 i 个元素。然后，我们可以使用随机梯度下降算法（Bottou，2012）训练采用量化截断-取整-平移激活函数的人工神经网络。

5 相关工作

人工神经网络-脉冲神经网络转换的研究最早由 Cao 等人（2015年）开展。随后，Diehl 等人（2015年）采用基于数据和基于模型的归一化方法，将一个三层卷积神经网络转换为脉冲神经网络。为了在复杂数据集和更深层网络上获得高性能的脉冲神经网络，Rueckauer 等人（2016年）和 Sengupta 等人（2019年）分别提出了更精确的缩放方法，用于归一化权重和缩放阈值；之后研究证明，这两种方法是等价的（Ding 等，2021）。然而，由于第 3 节分析的转换误差，转换后的深层脉冲神经网络需要数百个时间步才能得到准确结果。为解决潜在的信息损失问题，Rueckauer 等人（2016年）和 Han 等人（2020年）建议使用“减法复位”神经元，而非“归零复位”神经元。近期，人们提出了许多消除转换误差的方法。Rueckauer 等人（2016年）建议将激活值的 99.9 百分位数用作缩放因子，Ho 和 Chang（2020年）则加入了可训练裁剪层。此外，Han 等人（2020年）重新缩放脉冲神经网络的阈值，以避免脉冲神经元激活不当。Massa 等人（2020年）和 Singh 等人（2021年）评估了转换后的脉冲神经网络在 Loihi 神经形态处理器上的性能。我们的工作与 Deng 和 Gu（2020）及 Li 等（2021）类似，也关注转换误差。Deng 和 Gu（2020）通过在转换后的脉冲神经网络偏置之外引入额外偏置，最小化逐层误差。Li 等（2021）进一步提出采用量化微调对权重和偏置进行校准。他们在 16 和 32 个时间步下取得了良好结果



(D) 9**-16 RQ &,)$5-10 (E) 5HV1HW-20 RQ &,)$5-10 (F) 9**-16 RQ &,)$5-100 (G) 5HV1HW-20 RQ &,)$5-100

图3：比较人工神经网络的准确率。

但未尝试更极端的时间步数。相比之下，我们的工作旨在通过消除上述转换误差的技术，使人工神经网络适配脉冲神经网络。我们通过端到端训练量化层来提升整体性能。我们的偏移校正能够使单个脉冲神经网络在超低时间步数和较大时间步数下都表现良好。即使采用时间反向传播（BPTT）等监督学习方法，要在极少的时间步内维持脉冲神经网络的性能也很困难。由于经过充分训练，时间反向传播通常所需的时间步较少，但代价是大量的 GPU 计算（Wu 等人，2018；2019；Lee 等人，2016；Neftci 等人，2019；Lee 等人，2020；Zenke 和 Vogels，2021）。基于时序的反向传播方法（Bohte 等人，2002；Tavanaei 等人，2019；Kim 等人，2020）能够在非常短的时间窗口内训练脉冲神经网络，例如 5–10 个时间步。然而，这些方法通常仅适用于 MNIST（Kheradpisheh 和 Masquelier，2020）及 CIFAR10（Zhang 和 Li，2020）等简单数据集。Rathi 等人（2019）先通过转换方法初始化脉冲神经网络，再采用脉冲时序依赖可塑性对其进行调优，从而缩短仿真步数。本文所提方法能够以超低延迟（4 个时间步）实现高性能的脉冲神经网络。

6 实验

本节中，我们验证所提方法的有效性，并将其与其他最先进方法进行比较，评估它们在 CIFAR-10（LeCun 等人，1998年）、CIFAR-100（Krizhevsky 等人，2009年）和 ImageNet 数据集（Deng 等人，2009年）上的图像分类性能。与以往工作类似，我们采用 VGG-16（Simonyan 和 Zisserman，2014年）、ResNet-18（He 等人，2016年）和 ResNet-20 网络结构作为源人工神经网络。我们将所提方法与最先进的人工神经网络-脉冲神经网络转换方法进行比较，包括 Rathi 等人（2019年）提出的混合转换（HC）、Han 等人（2020年）提出的 RMP、Han 和 Roy（2020年）提出的 TSC、Ding 等人（2021年）提出的 RNL、Deng 和 Gu（2020年）提出的 ReLUThreshold-Shift（RTS），以及 Li等（2021年）提出的高级流水线脉冲神经网络转换（SNNC-AP）。我们还比较了不同的脉冲神经网络训练方法，以彰显低延迟推理的优势，包括 Rathi 等人（2019年）提出的混合转换-STDB（HC-STDB）、Wu等人（2018年）提出的 STBP、Wu 等人（2019年）提出的直接训练（DT），以及 Zhang 和 Li（2020年）提出的 TSSL。所提人工神经网络-脉冲神经网络算法的详细信息及训练配置见附录。

6.1 采用量化截断-取整-平移激活函数的人工神经网络测试准确率

我们首先比较采用量化截断激活（绿色曲线）的人工神经网络、采用量化截断-取整-平移激活函数（蓝色曲线）的人工神经网络，以及采用 ReLU 激活的原始人工神经网络（黑色虚线）的性能。图3(a)–(d)分别报告了 VGG-16 在 CIFAR-10 上、ResNet-20 在 CIFAR-10 上、VGG-16 在 CIFAR-100 上以及 ResNet-20 在 CIFAR-100 上的结果。采用量化截断-取整-平移激活函数的人工神经网络，其性能优于采用量化截断激活的人工神经网络。当 $L > 4$ 时，这两种人工神经网络均能达到与采用 ReLU 激活的原始人工神经网络相同的性能。这些结果表明，我们提出的量化裁剪-取整-偏移激活函数几乎不会影响人工神经网络的性能。

6.2 与最先进方法的比较

表 2 比较了我们的方法与在 CIFAR-10 上最先进的人工神经网络-脉冲神经网络转换方法。在低延迟推理（T ≤ 64）方面，在相同时间步设置下，我们的模型优于所有其他方法。对于 T $= 32$，我们方法的准确率略高于人工神经网络（95.54% 对比 95.52%），而 RMP、RTS、RNL 和 SNNC-AP 方法的精度损失分别为 33.3%、19.48%、7.42% 和 2.01%。此外，我们仅用 4 个时间步就达到了 93.96 % 的准确率，比需要 32 个时间步的 SNNC-AP 快 8 倍。对于 ResNet-20，我们用 4 个时间步达到了 83.75% 的准确率。值得注意的是，我们的超低延迟性能

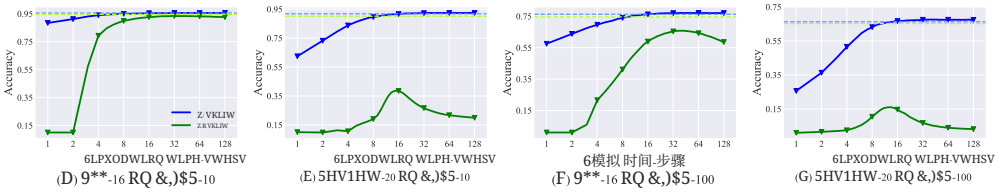


图 4：比较有无偏移项时的量化截断下限激活

表 2：所提方法与以往工作在 CIFAR-10 数据集上的比较。

架构	方法	ANN	T=2	T=4	T=8	T=16	T=32	T=64	T>512
VGG-16	RMP	93.63%	-	-	-	-	60.30%	90.35%	93.63%
	TSC	93.63%	-	-	-	-	-	92.79%	93.63%
	RTS	95.72%	-	-	-	-	76.24%	90.64%	95.73%
	RNL	92.82%	-	-	-	57.90%	85.40%	91.15%	92.95%
	SNNC-AP	95.72%	-	-	-	-	93.71%	95.14%	95.79%
	Ours	95.52%	91.18%	93.96%	94.95%	95.40%	95.54%	95.55%	95.59%
ResNet-20	RMP	91.47%	-	-	-	-	-	-	91.36%
	TSC	91.47%	-	-	-	-	-	69.38%	91.42%
	Ours	91.77%	73.20%	83.75%	89.55%	91.62%	92.24%	92.35%	92.41%
ResNet-18	RTS [1]	95.46%	-	-	-	-	84.06%	92.48%	94.42%
	SNNC-AP [1]	95.46%	-	-	-	-	94.78%	95.30%	95.45%
	Ours	96.04%	75.44%	90.43%	94.82%	95.92%	96.08%	96.06%	96.06%

[1] RTS 和 SNNC-AP 使用经过修改的 ResNet-18，而我们的方法使用标准 ResNet-18。

可与其他最先进的监督训练方法相媲美，如附录中的表 S3 所示。

我们进一步测试了所提方法在大规模数据集上的性能。表 3 报告了在 ImageNet 上的结果，我们的方法在准确率和超低延迟方面均优于其他方法。对于 ResNet-34，当 $T = 32$ 时，所提方法的准确率比 SNNC-AP 高 4.83%，比 RTS 高 69.28%。当时间步为 16 时，我们仍可达到 59.35 % 的准确率。对于 VGG-16，当 $T = 32$ 时，所提方法的准确率比 SNNC-AP 高 4.83%，比 RTS 高 68.356%。当时间步为 16 时，我们仍可达到 50.97% 的准确率。这些结果表明，我们的方法优于以往的转换方法。关于 CIFAR-100 的更多实验结果见附录中的表 S4。

6.3 量化截断激活与量化截断移位激活的比较

在这里，我们进一步比较分别由采用量化截断激活的人工神经网络和采用量化截断移位激活的人工神经网络转换得到的脉冲神经网络的性能。在第 4 节中，我们证明，采用量化截断-取整-平移激活函数时，无论 T 和 L 是否相同，转换误差的期望值都为 0。为验证这一点，我们将 L 设为 4，并分别训练采用量化截断激活和量化截断移位激活的人工神经网络。图 4 展示了转换后的脉冲神经网络的准确率随时间步数 T 的变化情况。由采用量化截断激活的人工神经网络（绿色虚线）转换得到的脉冲神经网络（绿色曲线）的准确率随着时间步数增加先上升，随后迅速下降，因为当 T 不等于 L时，我们无法保证转换误差为零。其最佳性能仍低于源人工神经网络（绿色虚线）。相比之下，由采用量化截断移位激活的人工神经网络转换得到的脉冲神经网络（蓝色曲线）的准确率随着 T增加而上升。当时间步数大于 16 时，其准确率与源人工神经网络（蓝色虚线）相同。

6.4 量化步数的影响

在我们的方法中，量化步数 L 是一个超参数，会影响转换后的脉冲神经网络的准确率。为分析 L 的影响并更好地确定其最佳取值，我们使用不同的量化步数 L（包括 2、4、8、16 和 32），训练采用量化截断-取整-平移激活函数的 VGG-16/ResNet-20 网络，然后将其转换为脉冲神经网络。实验结果

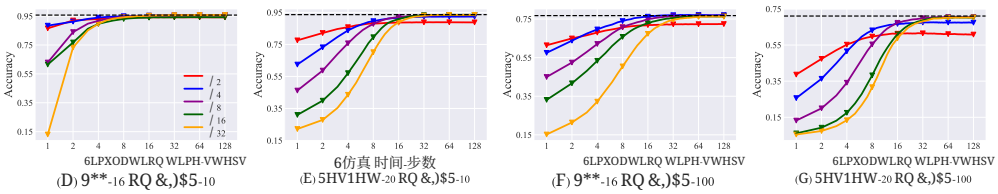


图 5：不同量化步数的影响

表 3：所提方法与以往工作在 ImageNet 数据集上的比较 t.

架构	方法	ANN	T=16	T=32	T=64	T=128	T=256	T>1024
ResNet-34	RMP	70.64%	-	-	-	-	-	65.47%
	TSC	70.64%	-	-	-	-	61.48%	65.10%
	RTS	75.66%	-	0.09%	0.12%	3.19%	47.11%	75.08%
	SNNC-AP	75.66%	-	64.54%	71.12%	73.45%	74.61%	75.45%
	Ours	74.32%	59.35%	69.37%	72.35%	73.15%	73.37%	73.39%
VGG-16	RMP	73.49%	-	-	-	-	48.32%	73.09%
	TSC	73.49%	-	-	-	-	69.71%	73.46%
	RTS	75.36%	-	0.114%	0.118%	0.122%	1.81%	73.88%
	SNNC-AP	75.36%	-	63.64%	70.69%	73.32%	74.23%	75.32%
	Ours	74.29%	50.97%	68.47%	72.85%	73.97%	74.22%	74.32%

在 CIFAR-10/100 数据集上的实验结果如表 S2 和图 5 所示，其中黑色虚线表示人工神经网络的准确率，彩色曲线表示转换后的脉冲神经网络的准确率。为权衡低延迟与高准确率，我们主要从两个方面评估转换后的脉冲神经网络的性能。首先，关注超低延迟（不超过 4 个时间步）下的脉冲神经网络准确率；其次，考察脉冲神经网络的最佳准确率。显然，随着 L 增大，超低延迟下的脉冲神经网络准确率会下降。然而， L 过小会降低模型容量，进而导致精度损失。当 $L = 2$ 时，脉冲神经网络的最佳准确率与源人工神经网络之间存在明显差距。当 $L > 4$ 时，脉冲神经网络的最佳准确率接近源人工神经网络。总之，参数 L 的设定主要取决于对低延迟或最佳准确率的追求。推荐的量化步长 L 为 4 或 8，这样转换后的脉冲神经网络在较少和较多时间步下均能取得出色性能。

7 讨论与结论

本文提出了一种人工神经网络到脉冲神经网络的转换方法，可实现高准确率、超低延迟的深度脉冲神经网络。我们提出量化截断-取整-平移激活函数来替代 ReLU 激活函数，该方法几乎不影响人工神经网络的性能，且更接近脉冲神经网络的激活方式。此外，我们证明，无论脉冲神经网络的时间步数与人工神经网络的量化步数是否相同，期望转换误差均为零。我们在 CIFAR-10、CIFAR-100 和 ImageNet 数据集上，以更少的时间步数实现了最先进方法的准确率。我们的研究成果有助于在神经形态硬件上实现相关应用，并为脉冲神经网络的大规模应用铺平道路。

与 Deng 和 Gu（2020）的工作不同，他们将转换后的脉冲神经网络的偏置用于偏移理论上的人工神经网络-脉冲神经网络曲线，以尽可能减小量化误差；我们则在量化裁剪取整激活函数中加入偏移项，并使用这一量化-截断-取整-移位函数训练源人工神经网络。我们表明，当时间步与量化步数不匹配时，偏移项能够克服性能下降问题。由于存在不均匀误差，即使 $L = T$，人工神经网络准确率与脉冲神经网络准确率之间仍有差距。此外，当时间步 $T = 1$ 时，很难实现高性能的人工神经网络-脉冲神经网络转换。这些问题都值得进一步研究。基于转换的方法的一个优势是，它们能够在保持与源人工神经网络性能相当的同时，降低总体计算成本。将基于转换的方法与模型压缩相结合，可能有助于大幅降低神经元活动

从而降低能耗，且不会造成精度损失（Kundu 等人，2021；Rathi 和 Roy，2021），这是一个很有前景的方向。

致谢

本研究得到国家自然科学基金资助，项目编号为 No.62176003 和 No.62088102。

参考文献

Yoshua Bengio、Nicholas L´eonard 和 Aaron Courville。针对条件计算，通过随机神经元估计或传播梯度。arXiv 预印本 arXiv:1308.3432，2013 年。

Sander M Bohte、Joost N Kok 和 Han La Poutre。脉冲神经元时间编码网络中的误差反向传播。Neurocomputing，48(1-4)：17–37，2002 年。

L´eon Bottou。随机梯度下降技巧。载于《神经网络：技巧与诀窍》，第421–436页。Springer，2012年。

Yongqiang Cao、Yang Chen 和 Deepak Khosla。用于节能目标识别的脉冲深度卷积神经网络。《International Journal of Computer Vision》，113(1)：54–66，2015年。

Ekin D Cubuk、Barret Zoph、Dandelion Mane、Vijay Vasudevan 和 Quoc V Le。AutoAugment：从数据中学习数据增强策略。载于 IEEE计算机视觉与模式识别会议，第113–123页，2019年。

Mike Davies、Narayan Srinivasa、Tsung-Han Lin、Gautham Chinya、Yongqiang Cao、Sri Harsha Choday、Georgios Dimou、Prasad Joshi、Nabil Imam、Shweta Jain 等。Loihi：一种具备片上学习功能的神经形态多核处理器。《IEEE Micro》，38(1)：82–99，2018。

Michael V DeBole、Brian Taba、Arnon Amir、Filipp Akopyan、Alexander Andreopoulos、William P Risk、Jeff Kusnitz、Carlos Ortega Otero、Tapan K Nayak、Rathinakumar Appuswamy 等。TrueNorth：十年间从零发展到 6400 万个神经元。Computer，52(5)：20–29，2019年。

Jia Deng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei。ImageNet：大型分层图像数据库。载于 IEEE计算机视觉与模式识别会议，页 248–255。Ieee，2009年。

Shikuang Deng 和 Shi Gu。将传统人工神经网络最优转换为脉冲神经网络。载于国际学习表征会议，2020。

Terrance DeVries 和 Graham W Taylor。利用 Cutout 改进卷积神经网络的正则化。arXiv 预印本 arXiv:1708.04552，2017 年。

Peter U Diehl、Daniel Neil、Jonathan Binas、Matthew Cook、Shih-Chii Liu 和 Michael Pfeiffer。通过权重与阈值平衡，实现快速分类、高准确率的脉冲深度网络。发表于国际神经网络联合会议，第 1–8 页，2015年。

Jianhao Ding、Zhaofei Yu、Yonghong Tian 和 Tiejun Huang。用于深层脉冲神经网络快速、准确推理的最优人工神经网络-脉冲神经网络转换。发表于国际人工智能联合会议，第 2328–2336 页，2021。

Wei Fang、Zhaofei Yu、Yanqi Chen、Tiejun Huang、Timoth´ee Masquelier 和 Yonghong Tian。脉冲神经网络中的深度残差学习。arXiv 预印本 arXiv:2102.04159，2021。

Bing Han 和 Kaushik Roy。深层脉冲神经网络：通过基于时间的编码提升能量效率。发表于欧洲计算机视觉会议，第 388–404 页，2020。

Bing Han、Gopalakrishnan Srinivasan 和 Kaushik Roy。RMP-SNN：残差膜电位神经元，用于实现更深层、高准确率和低延迟的脉冲神经网络。发表于 IEEE计算机视觉与模式识别会议，第 13558–13567 页，2020。

Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。《用于图像识别的深度残差学习》。载于 IEEE计算机视觉与模式识别会议，第 770–778 页，2016年。

Nguyen-Dong Ho 和 Ik-Joon Chang。Tcl：一种带有可训练裁剪层的人工神经网络到脉冲神经网络转换。arXiv 预印本 arXiv:2008.04509，2020。Eugene M Izhikevich。脉冲神经元的简单模型。《IEEE 神经网络汇刊》，14(6):1569–1572，2003。Saeed Reza Kheradpisheh 和 Timoth´ee Masquelier。每个神经元仅发放一个脉冲的脉冲神经网络时间反向传播。《国际神经系统杂志》，30(06):2050027，2020。Jinseok Kim、Kyungsu Kim 和 Jae-Joon Kim。统一脉冲神经网络的基于激活和基于时序的学习规则。载于《神经信息处理系统进展》，第 19534–19544 页，2020。Alex Krizhevsky、Geoffrey Hinton 等。学习微小图像的多层特征。2009。Souvik Kundu、Gourav Datta、Massoud Pedram 和 Peter A Beerel。Spike-thrift：通过注意力引导的压缩限制脉冲活动，打造高能效的深层脉冲神经网络。载于 IEEE/CVF 冬季计算机视觉应用会议（WACV），第 3953–3962 页，2021。Yann LeCun、L´eon Bottou、Yoshua Bengio 和 Patrick Haffner。将基于梯度的学习应用于文档识别。《IEEE 论文集》，86(11):2278–2324，1998。Chankyu Lee、Syed Shakib Sarwar、Priyadarshini Panda、Gopalakrishnan Srinivasan 和 Kaushik Roy。实现基于脉冲的反向传播，以训练深度神经网络架构。《神经科学前沿》，14，2020。Jun Haeng Lee、Tobi Delbruck 和 Michael Pfeiffer。使用反向传播训练深层脉冲神经网络。《神经科学前沿》，10:508，2016。Yuhang Li、Shikuang Deng、Xin Dong、Ruihao Gong 和 Shi Gu。来自人工神经网络的免费午餐：迈向高效、准确的脉冲神经网络校准。载于国际机器学习会议，第 6316–6325 页，2021。Ilya Loshchilov 和 Frank Hutter。随机梯度下降热重启：带热重启的随机梯度下降。载于国际表征学习会议，2016。Wolfgang Maass。脉冲神经元网络：第三代神经网络模型。Neural Networks，10(9):1659–1671，1997。Riccardo Massa、Alberto Marchisio、Maurizio Martina 和 Muhammad Shafique。一种高效的脉冲神经网络：利用 Loihi 神经形态处理器和 DVS 相机识别手势。载于国际神经网络联合会议，第 1–9 页，2020。Warren S McCulloch 和 Walter Pitts。神经活动中内在观念的逻辑演算。《数学生物物理学通报》，5(4):115–133，1943。Paul A Merolla、John V Arthur、Rodrigo Alvarez-Icaza、Andrew S Cassidy、Jun Sawada、Filipp Akopyan、Bryan L Jackson、Nabil Imam、Chen Guo、Yutaka Nakamura 等。一种具有可扩展通信网络和接口的百万脉冲神经元集成电路。《科学》，345 (6197):668–673，2014。Emre O Neftci、Hesham Mostafa 和 Friedemann Zenke。脉冲神经网络中的替代梯度学习：将基于梯度的优化能力引入脉冲神经网络。《IEEE 信号处理杂志》，36(6):51–63，2019。Jing Pei、Lei Deng、Sen Song、Mingguo Zhao、Youhui Zhang、Shuang Wu、Guanrui Wang、Zhe Zou、Zhenzhi Wu、Wei He 等。借助混合型天机芯片架构迈向通用人工智能。《自然》，572(7767):106–111，2019。Ning Qiao、Hesham Mostafa、Federico Corradi、Marc Osswald、Fabio Stefanini、Dora Sumislawska 和 Giacomo Indiveri。一种可重构在线学习脉冲神经形态处理器，包含 256 个神经元和 128K 个突触。《神经科学前沿》，9:141，2015。

Nitin Rathi 和 Kaushik Roy。Diet-snn：一种具有直接输入编码以及泄漏与阈值优化的低延迟脉冲神经网络。IEEE 神经网络与学习系统汇刊，2021。

Nitin Rathi、Gopalakrishnan Srinivasan、Priyadarshini Panda 和 Kaushik Roy。通过混合转换和依赖脉冲时序的反向传播实现深层脉冲神经网络。载于《国际学习表征会议》，2019。

Kaushik Roy、Akhilesh Jaiswal 和 Priyadarshini Panda。借助神经形态计算迈向基于脉冲的机器智能。《自然》，575(7784)：607–617，2019年。

Bodo Rueckauer、Iulia-Alexandra Lungu、Yuhuang Hu 和 Michael Pfeiffer。模拟脉冲卷积神经网络转换的理论与工具。arXiv 预印本 arXiv:1612.04052，2016年。

Bodo Rueckauer、Iulia-Alexandra Lungu、Yuhuang Hu、Michael Pfeiffer 和 Shih-Chii Liu。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。《神经科学前沿》，11:682，2017年。

Olga Russakovsky、Jia Deng、Hao Su、Jonathan Krause、Sanjeev Satheesh、Sean Ma、Zhiheng Huang、Andrej Karpathy、Aditya Khosla、Michael Bernstein、Alexander C. Berg 和 Li Fei-Fei。ImageNet 大规模视觉识别挑战赛。《International Journal of Computer Vision》（IJCV），115(3):211–252，2015年。doi：10.1007/s11263-015-0816-y。

Abhronil Sengupta、Yuting Ye、Robert Wang、Chiao Liu 和 Kaushik Roy。深入探索脉冲神经网络：VGG 与残差架构。《神经科学前沿》，13:95，2019年。

Karen Simonyan 和 Andrew Zisserman。用于大规模图像识别的超深卷积网络。arXiv 预印本 arXiv:1409.1556，2014年。

Sonali Singh、Anup Sarma、Sen Lu、Abhronil Sengupta、Vijaykrishnan Narayanan 和 Chita R Das。Gesture-snn：针对神经形态视觉传感器，对脉冲神经网络的准确率、延迟和能量进行协同优化。载于 IEEE/ACM 国际低功耗电子与设计研讨会，第 1–6 页，2021。

Christoph St¨ockl 和 Wolfgang Maass。优化的脉冲神经元通过使用两个脉冲的时间编码，以高准确率对图像进行分类。《自然·机器智能》，3(3)：230–238，2021。

Amirhossein Tavanaei、Masoud Ghodrati、Saeed Reza Kheradpisheh、Timoth´ee Masquelier 和 Anthony Maida。脉冲神经网络中的深度学习。Neural Networks，111：47–63，2019年。

Yujie Wu、Lei Deng、Guoqi Li、Jun Zhu 和 Luping Shi。用于训练高性能脉冲神经网络的时空反向传播。《神经科学前沿》，12:331，2018。

Yujie Wu、Lei Deng、Guoqi Li、Jun Zhu、Yuan Xie 和 Luping Shi。脉冲神经网络的直接训练：更快、更大、更优。载于 AAAI人工智能大会，第 1311–1318 页，2019年。

Friedemann Zenke 和 Tim P Vogels。替代梯度学习在脉冲神经网络中实现复杂函数方面具有非凡的稳健性。Neural Computation，33(4):899–925，2021。

Wenrui Zhang 和 Peng Li。通过反向传播进行时间脉冲序列学习，以训练深层脉冲神经网络。载于《神经信息处理系统进展》，第 12022–12033 页，2020。

A 附录

A.1 网络结构与训练配置

训练人工神经网络之前，我们首先用平均池化替换最大池化，然后用所提出的量化截断-取整-平移激活函数（公式 15）替换 ReLU 激活。训练完成后，我们将源人工神经网络中的所有权重复制到转换后的脉冲神经网络，并将转换后的脉冲神经网络中每一层的阈值 θ^l 设为源人工神经网络对应层的最大激活值 λ^l。此外，我们将转换后的脉冲神经网络中的初始膜电位 $\boldsymbol{v}^l(0)$设为 $\boldsymbol{\theta^l}/2$，以匹配源人工神经网络中量化截断-取整-平移激活函数的最优偏移 $\varphi = \frac{1}{2}$。
除了常规的数据归一化外，我们还采用了一些数据预处理技术。对于CIFAR数据集，我们将图像调整为 32×32；对于ImageNet数据集，我们将图像调整为224×224。此外，我们对所有数据集使用随机裁剪、Cutout（DeVries & Taylor，2017）和AutoAug-ment（Cubuk等，2019）。

我们使用随机梯度下降优化器（Bottou，2012），动量参数设为0.9。对于CIFAR-10和ImageNet，初始学习率设为0.1；对于CIFAR-100，设为0.02。我们使用余弦衰减调度器（Loshchilov & Hutter，2016）来调整学习率。对于CIFAR数据集，我们采用5×10^{-4} 权重衰减；对于ImageNet，则采用 1×10^{-4} 权重衰减。所有模型均训练300个训练轮次。在CIFAR-10以及CIFAR-100 数据集上的VGG-16、ResNet-18训练过程中，量化步数 L 设为4。在CIFAR-100上训练ResNet-20时，参数 L 设为8。在ImageNet上训练ResNet-34和VGG-16时，参数 L 分别设为8和16。评估转换后的脉冲神经网络时，我们使用恒定输入。

A.2 数据集介绍

CIFAR-10。CIFAR-10 数据集（Krizhevsky 等人，2009年）包含 60000 32×32 张图像，分为 10 个类别。其中有 50000 张训练图像和 10000 张测试图像。
CIFAR-100。CIFAR-100 数据集（Krizhevsky 等人，2009年）包含 60000 32×32 张图像，分为 100 个类别。其中有 50000 张训练图像和 10000 张测试图像。
ImageNet。我们使用 ILSVRC 2012 数据集（Russakovsky 等人，2015年），该数据集包含 1,281,167 张训练图像和 50000 张测试图像。

A.3 方程12的推导及定理 2 的证明

方程11的推导
类似于，，我们定义

$$\boldsymbol{u}^l(t) = \boldsymbol{W}^l \boldsymbol{x}^{l-1}(t). \tag{S1}$$

我们分别用 $u_i^l(t)$ 和 z_i^l 表示向量 $\boldsymbol{u}^l(t)$ 和 $\boldsymbol{z}^l$ 中的第 i 个元素。为推导方程11，需要对人工神经网络激活值与脉冲神经网络突触后电位之间的关系作出一些额外假设，这些假设见方程S2。

$$\begin{cases} \text{if } z_i^l < 0, \ \text{then } \forall t \ u_i^l(t) < 0, \\ \text{if } 0 \leqslant z_i^l \leqslant \theta_l, \ \text{then } \forall t \ 0 \leqslant u_i^l(t) \leqslant \theta_l, \\ \text{if } z_i^l > \theta_l, \ \text{then } \forall t \ u_i^l(t) > \theta_l. \end{cases} \tag{S2}$$

基于上述假设，我们可以讨论神经元在每个时间步的发放行为。当 $z_i^l < 0$ 或 $z_i^l > \theta_l$时，神经元将永远不会发放，或在所有时间步都发放，这意味着 $\phi_i^l(T) = 0$或 $\phi_i^l(T) = \theta^l$。在这种情况下，我们可以使用裁剪函数表示 $\phi_i^l(T)$。

$$\phi_i^l(T) = \text{clip}(z_i^l, 0, \theta^l). \tag{S3}$$

当 $0 < z_i^l < \theta_l$时，脉冲神经网络中来自突触前神经元的每个输入都落在 $[0, \theta^l]$内，因此我们有$\forall t$, $v_i^l(t) \in [0, \theta]$。我们可以将公式 8 改写为以下方程。

$$\frac{\phi_i^l(T)T}{\theta^l} = \frac{z_i^l T + v_i^l(0)}{\theta^l} - \frac{v_i^l(T)}{\theta^l}. \tag{S4}$$

考虑到 $\frac{\phi_i^l}{\theta^l}(T) = \sum_{t=1}^{T} s_i^l(t) \in \mathbb{N}$ 和 $0 < \frac{v_i^l(T)}{\theta^l} < 1$，公式 S4 可改写为：

$$\phi_i^l(T) = \frac{\theta^l}{T} \left\lfloor \frac{z_i^l T + v_i^l(0)}{\theta^l} \right\rfloor. \tag{S5}$$

我们将这两种情况（公式 S3 和公式 S4）结合起来，得到：

$$\boldsymbol{\phi}^l(T) = \theta^l \operatorname{clip}\left(\frac{1}{T} \left\lfloor \frac{\boldsymbol{z}^l T + \boldsymbol{v}^l(0)}{\theta^l} \right\rfloor, 0, 1 \right). \tag{S6}$$

定理 2 的证明

在证明定理 2 之前，我们先介绍引理 1。

引理 1. 若随机变量 $x \in [0, \theta]$ 在每个小区间 $[m_t, m_{t+1}]$上均匀分布，其概率密度函数为 $p_t(t = 0, 1, ..., T)$，其中 $m_0 = 0, m_{T+1} = \theta, m_t = (t - \frac{1}{2}) \frac{\theta}{T}$，对于 $t = 1, 2, ..., T, p_0 = p_T$，则可以得出

$$\mathbb{E}_x \left(x - \frac{\theta}{T} \left\lfloor \frac{Tx}{\theta} + \frac{1}{2} \right\rfloor \right) = 0. \tag{S7}$$

证明。

$$\begin{aligned}
&\mathbb{E}_x \left(x - \frac{\theta}{T} \left\lfloor \frac{Tx}{\theta} + \frac{1}{2} \right\rfloor \right) = \int_0^{\theta/2T} p_0 \left(x - \frac{\theta}{T} \left\lfloor \frac{xT}{\theta} + \frac{1}{2} \right\rfloor \right) \mathrm{d}x \\
&+ \sum_{t=1}^{T-1} \int_{(2t-1)\theta/2T}^{(2t+1)\theta/2T} p_t \left(x - \frac{\theta}{T} \left\lfloor \frac{xT}{\theta} + \frac{1}{2} \right\rfloor \right) \mathrm{d}x \\
&+ \int_{(2T-1)\theta/2T}^{\theta} p_T \left(x - \frac{\theta}{T} \left\lfloor \frac{xT}{\theta} + \frac{1}{2} \right\rfloor \right) \mathrm{d}x \\
&= p_0 \int_0^{\theta/2T} x \, \mathrm{d}x + \sum_{t=1}^{T-1} p_t \int_{(2t-1)\theta/2T}^{(2t+1)\theta/2T} (x - \frac{t\theta}{T}) \, \mathrm{d}x + p_T \int_{(2T-1)\theta/2T}^{\theta} (x - \theta) \, \mathrm{d}x \\
&= p_0 \frac{\theta^2}{8T^2} + 0 - p_T \frac{\theta^2}{8T^2} = (p_0 - p_T) \frac{\theta^2}{8T^2} = 0.
\end{aligned} \tag{S8}$$

□

定理 2. 将激活函数为 (15) 的人工神经网络转换为权重相同的脉冲神经网络。若 $\theta^l = \lambda^l$, $\boldsymbol{v}^l(0) = \theta^l \boldsymbol{\varphi}$, 则对于任意 T和 L，当源人工神经网络中的偏移项 $\boldsymbol{\varphi}$ 为 $\mathbf{1}_2$ 时，转换误差的期望达到 0。

$$\forall\, T, L \quad \mathbb{E}_z \left(\widetilde{\boldsymbol{Err}}^l \right) \Big|_{\boldsymbol{\varphi} = \frac{1}{2}} = \mathbf{0}. \tag{S9}$$

证明。

$$\mathbb{E}_z \left(\widetilde{\boldsymbol{Err}}^l \right) \Big|_{\boldsymbol{\varphi} = \frac{1}{2}} = \mathbb{E}_z \left(\frac{\theta^l}{T} \left\lfloor \frac{\boldsymbol{z}^l T + \boldsymbol{v}^l(0)}{\theta^l} \right\rfloor - \frac{\lambda^l}{L} \left\lfloor \frac{\boldsymbol{z}^l L}{\lambda} + \boldsymbol{\varphi} \right\rfloor \right). \tag{S10}$$

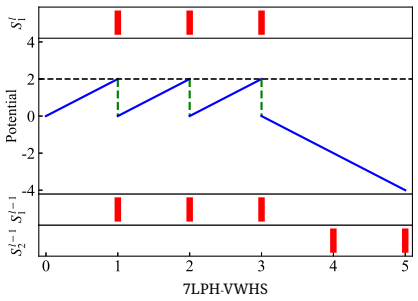


图 S1：采用设置最大激活值的方法时，脉冲数量超出预期。

由于向量 $\boldsymbol{z}$ 中的每个元素都相同，我们只需考虑其中一个元素。

$$\begin{aligned}&\mathbb{E}_{z_i}\left(\frac{\theta^l}{T}\left\lfloor\frac{z_i^l T+v_i^l(0)}{\theta^l}\right\rfloor-\frac{\lambda^l}{L}\left\lfloor\frac{z_i^l L}{\lambda}+\varphi_i\right\rfloor\right)\\&=\mathbb{E}_{z_i}\left(\frac{\theta^l}{T}\left\lfloor\frac{z_i^l T+v_i^l(0)}{\theta^l}\right\rfloor-z_i^l\right)+\mathbb{E}_{z_i}\left(z_i^l-\frac{\lambda^l}{L}\left\lfloor\frac{z_i^l L}{\lambda}+\varphi_i\right\rfloor\right).\end{aligned}\tag{S11}$$

根据引理 1，我们有

$$\mathbb{E}_{z_i}\left(\frac{\theta^l}{T}\left\lfloor\frac{z_i^l T+v_i^l(0)}{\theta^l}\right\rfloor-z_i^l\right)\Big|_{v_i^l(0)=1/2}=0,\tag{S12}$$

$$\mathbb{E}_{z_i}\left(z_i^l-\frac{\lambda^l}{L}\left\lfloor\frac{z_i^l L}{\lambda}+\varphi_i\right\rfloor\right)\Big|_{\varphi=1/2}=0.\tag{S13}$$

因此，两项之和也等于零。 □

A.4 CIFAR-100 数据集上使用或不使用动态阈值的方法比较

本文使用训练参数 λ^l 来决定人工神经网络激活值的最大值。以往工作建议将训练后人工神经网络激活值的最大值设为阈值。如果我们设置 $\theta^l=\max_{\boldsymbol{s}\in\{0,1\}^n}\left(\max(\theta^{l-1}\boldsymbol{W}^l\boldsymbol{s})\right)$，那么预期中脉冲更少的情况永远不会发生，因为我们可以证明 $\boldsymbol{v}^l(T)<\theta^l$（见定理 3）。尽管如此，预期中脉冲更多的情况仍然存在。图 S1 给出了一个示例。这里我们考虑与图 1 相同的示例。在源人工神经网络中，假设层$l-1$ 中的两个模拟神经元通过权重 2 和 -2 与层 l 中的一个模拟神经元相连，且层 $l-1$ 中神经元的输出向量 $\boldsymbol{a}^{l-1}$ 为 $[0.6,0.4]$。此外，在转换后的脉冲神经网络中，假设层 $l-1$ 中的两个脉冲神经元分别在 5 个时间步（T=5）内发放 3 个脉冲和 2 个脉冲，且阈值为 $\theta^{l-1}=1$。因此，$\boldsymbol{\phi}^{l-1}(T)=\sum_{i=1}^{T}\boldsymbol{s}^{l-1}(i)^T\theta^{l-1}=[0.6,0.4]$。根据公式 1，人工神经网络的输出为 $\boldsymbol{a}^l=\boldsymbol{W}^l\boldsymbol{a}^{l-1}=[2,-2][0.6,0.4]^T=0.4$。对于脉冲神经网络，假设突触前神经元分别在 $t=1$、2、3 和 $t=4$、5 时发放。即使我们将阈值 $\theta^l=1$ 设为最大激活值 2，突触后神经元仍会在 $t=1$、2、3 时发放三个脉冲，且 $\boldsymbol{\phi}^l(T)=0.6>\boldsymbol{a}^l$。

此外，将 $\max_{\boldsymbol{s}\in\{0,1\}^n}\left(\max(\theta^{l-1}\boldsymbol{W}^l\boldsymbol{s})\right)$ 设为阈值还会带来另外两个问题。首先，由于阈值较大，脉冲神经元需要较长时间才能发放脉冲，因此难以在较少的时间步内维持脉冲神经网络的性能。其次，量化误差会较大，因为它与阈值成正比。如果某一层的转换误差不为零，它就会逐层传播，并因较大的量化误差而被放大。我们在 CIFAR-100 数据集上比较了我们的方法与将最大激活值设为阈值的方法。结果如表 S1 所示，其中 DT 表示我们方法中的动态阈值。结果表明，我们的方法能够取得更好的性能。

表 S1：我们的方法与设置最大激活值的方法之间的比较。

DT[1]	无偏移	T=4	T=8	T=16	T=32	T=64	T=128	T=256	T≥512
CIFAR-100 上的 VGG-16（L=4）									
✓	✓	69.62%	73.96%	76.24%	77.01%	77.10%	77.05%	77.08%	77.08%
✓	×	21.57%	41.13%	58.92%	65.38%	64.19%	58.60%	52.99%	49.41%
×	✓	1.00%	0.96%	1.00%	1.10%	2.41%	13.76%	51.70%	77.10%
×	×	1.00%	1.00%	0.90%	1.00%	1.01%	2.01%	19.59%	70.86%

[1] 动态阈值。

定理 3. 如果将阈值设为人工神经网络激活值的最大值，即 $\theta^l = \max_{\boldsymbol{s}\in\{0,1\}^n}(\max(\theta^{l-1}\boldsymbol{W}^l\boldsymbol{s}))$，且 $v_i^l(0) < \theta^l$，那么在任意时间步，每个神经元在发放脉冲 $v_i^l(t)$ 后的膜电位都将小于 θ^l，其中 i 表示每个神经元的索引。

证明。我们用归纳法证明。对于 $t = 0$，容易看出 $v_i^l(0) < \theta^l$。对于 $t > 0$，我们假设 $v_i^l(t-1) < \theta^l$。由于我们已将阈值设为可能的最大输入，且 $x_i^{l-1}(t)$ 表示从第 $l-1$ 层到第 l 层第 i 个神经元的输入，因此对于任意 t，$x_i^{l-1}(t)$ 不会大于 θ^l。因此我们有

$$m_i^l(t) = v_i^l(t-1) + x_i^{l-1}(t) < \theta^l + \theta^l = 2\theta^l, \tag{S14}$$

$$s_i^l(t) = H(m_i^l(t) - \theta^l), \tag{S15}$$

$$v_i^l(t) = m_i^l(t) - s_i^l(t)\theta^l. \tag{S16}$$

如果 $\theta^l \leqslant m_i^l(t) < 2\theta^l$，那么我们有 $v_i^l(t) = m_i^l(t) - \theta^l < \theta^l$。如果 $m_i^l(t) < \theta_l$，那么 $v_i^l(t) = m_i^l(t) < \theta_l$。根据数学归纳法，$v_i^l(t) < \theta^l$ 对任意 $t \geqslant 0$ 均成立。 □

A.5 量化步数的影响

表 S2 报告了采用不同量化步数 L 和不同时间步 T的转换后脉冲神经网络的性能。对于 VGG-16，当量化步数为 $L = 2$时，我们在 CIFAR-10 数据集上取得了 86.53% 的准确率，在 CIFAR-100 数据集上取得了 61.41% 的准确率，且仅需 1 个时间步。当量化步数为 $L = 1$时，我们无法训练源人工神经网络。

A.6 与 CIFAR-10 数据集上最先进的监督训练方法进行比较

值得注意的是，我们的超低延迟性能可与其他最先进的监督训练方法相媲美。表 S3 报告了 CIFAR-10 上混合训练和反向传播方法的结果。反向传播方法需要足够多的时间步来传递判别信息。因此，所列方法至少需要 5 个时间步才能达到 ∼91% 的准确率。相比之下，我们的方法仅用 4 个时间步便可达到 94.73% 的准确率。此外，由于还需通过 STDB 进一步训练，混合训练方法需要 200 个时间步才能达到 92.02% 的准确率，而我们的方法仅用 4 个时间步便可达到 93.96% 的准确率。

A.7 CIFAR-100数据集上的比较

表 S4 报告了 CIFAR-100 上的结果。无论从高准确率还是超低延迟来看，我们的方法都优于其他方法。对于 VGG-16，当 $T = 32$ 时，所提方法的准确率比 SNNC-AP 高 3.46%，比 RTS 高 69.37%。即使时间步仅为 4，我们仍能达到 69.62% 的准确率。这些结果表明，我们的方法优于以往的转换方法。

表 S2：不同量化步数的影响。

量化步数	T=1	T=2	T=4	T=8	T=16	T=32	T=64	T=128
CIFAR-10 上的 VGG-16								
L=2	86.53%	91.98%	93.00%	93.95%	94.18%	94.22%	94.18%	94.14%
L=4	88.41%	91.18%	93.96%	94.95%	95.40%	95.54%	95.55%	95.59%
L=8	62.89%	83.93%	91.77%	94.45%	95.22%	95.56%	95.74%	95.79%
L=16	61.48%	76.76%	89.61%	93.03%	93.95%	94.24%	94.25%	94.22%
L=32	13.05%	73.33%	89.67%	94.13%	95.31%	95.66%	95.73%	95.77%
CIFAR-10 上的 ResNet-20								
L=2	77.54%	82.12%	85.77%	88.04%	88.64%	88.79%	88.85%	88.76%
L=4	62.43%	73.2%	83.75%	89.55%	91.62%	92.24%	92.35%	92.35%
L=8	46.19%	58.67%	75.70%	87.79%	92.14%	93.04%	93.34%	93.24%
L=16	30.96%	39.87%	57.04%	79.5%	90.87%	93.25%	93.44%	93.48%
L=32	22.15%	27.83%	43.56%	70.15%	88.81%	92.97%	93.48%	93.48%
CIFAR-100 上的 VGG-16								
L=2	61.41%	64.96%	68.0%	70.72%	71.87%	72.28%	72.35%	72.4%
L=4	57.5%	63.79%	69.62%	73.96%	76.24%	77.01%	77.1%	77.05%
L=8	44.98%	52.46%	62.09%	70.71%	74.83%	76.41%	76.73%	76.73%
L=16	33.12%	41.71%	53.38%	65.76%	72.80%	75.6%	76.37%	76.36%
L=32	15.18%	21.41%	32.21%	50.46%	67.32%	74.6%	76.18%	76.24%
CIFAR-100 上的 ResNet-20								
L=2	38.65%	47.35%	55.23%	59.69%	61.29%	61.5%	61.03%	60.81%
L=4	25.62%	36.33%	51.55%	63.14%	66.70%	67.47%	67.47%	67.41%
L=8	13.19%	19.96%	34.14%	55.37%	67.33%	69.82%	70.49%	70.55%
L=16	6.09%	9.25%	17.48%	38.22%	60.92%	68.70%	70.15%	70.20%
L=32	5.44%	7.41%	13.36%	31.66%	58.68%	68.12%	70.12%	70.27%

A.8 能量消耗分析

我们在 CIFAR-100 数据集上评估了我们的方法和对比方法（Li 等，2021；Deng 和 Gu, 2020）的能耗。这里使用相同的 VGG-16 网络结构。根据 Merolla 等人（2014）的分析，我们使用脉冲神经网络（SNN）的突触操作（SOP）来表示分类一张图像所需的基本操作数。我们采用 77fJ/SOP 作为 SNN 的功耗基准，并采用 12.5pJ/浮点运算作为人工神经网络（ANN）的功耗基准，这些数据来自 ROLLS 神经形态处理器（Qiao 等人，2015）。请注意，本研究未考虑存储器访问能耗，因为它取决于硬件。如表 S5 所示，当时间步相同时，我们的方法能耗约为 SNNC-AP 的两倍。然而，要达到相同的 73.55% 准确率，我们的方法所需能耗更低。

A.9 整体转换算法的伪代码

本节通过算法 1 总结了整个转换过程，包括从头开始训练人工神经网络，以及将人工神经网络转换为脉冲神经网络。伪代码中的 QCFS 表示所提出的量化-截断-取整-移位函数。

表 S3：与 CIFAR-10 数据集上最先进的监督训练方法进行比较

模型	方法	架构	脉冲神经网络准确率	时间步
		CIFAR-10		
HC	混合	VGG-16	92.02	200
STBP	反向传播	CIFARNet	90.53	12
DT	反向传播	CIFARNet	90.98	8
TSSL	反向传播	CIFARNet	91.41	5
动态阈值[1] 人工神经网络-脉冲神经网络		cNet	77.10	256
Ours	人工神经网络-脉冲神经网络	VGG-16	93.96	4
Ours	人工神经网络-脉冲神经网络	CIFARNet[2]	94.73	4

[1] 在 Loihi 神经形态处理器上实现 [2] 对于 CIFARNet，我们采用与 Wu 等人（2018）相同的架构。

表 S4：所提方法与以往工作在 CIFAR-100 数据集上的比较。

架构	方法	ANN	T=2	T=4	T=8	T=16	T=32	T=64	T>512
VGG-16	RMP	71.22%	-	-	-	-	-	-	70.93%
	TSC	71.22%	-	-	-	-	-	-	70.97%
	RTS	77.89%	-	-	-	-	7.64%	21.84%	77.71%
	SNNC-AP	77.89%	-	-	-	-	73.55%	76.64%	77.87%
	Ours	76.28%	63.79%	69.62%	73.96%	76.24%	77.01%	77.10%	77.08%
ResNet-20	RMP	68.72%	-	-	-	-	27.64%	46.91%	67.82%
	TSC	68.72%	-	-	-	-	-	-	68.18%
	Ours	69.94%	19.96%	34.14%	55.37%	67.33%	69.82%	70.49%	70.50%
ResNet-18	RTS	77.16%	-	-	-	-	51.27%	70.12%	77.19%
	SNNC-AP	77.16%	-	-	-	-	76.32%	77.29%	77.25%
	Ours	78.80%	70.79%	75.67%	78.48%	79.48%	79.62%	79.54%	79.61%

[1] RTS 和 SNNC-AP 使用改造后的 ResNet-18，而我们的方法使用标准的 ResNet-18。

表 S5：与以往工作的能耗比较

方法		ANN	T=2	T=4	T=8	T=16	T=32	T=64
RTS	准确率	77.89%	-	-	-	-	7.64%	21.84%
	操作数 (GFLOP/GSOP)	0.628	-	-	-	-	0.508	0.681
	能量 (毫焦耳)	7.85	-	-	-	-	0.039	0.052
SNNC-AP	准确率	77.89%	-	-	-	-	73.55%	76.64%
	操作数 (GFLOP/GSOP)	0.628	-	-	-	-	0.857	1.22
	能量 (毫焦耳)	7.85	-	-	-	-	0.660	0.094
Ours	准确率	76.28%	63.79%	69.62%	73.96%	76.24%	77.01%	77.10%
	操作数 (GFLOP/GSOP)	0.628	0.094	0.185	0.364	0.724	1.444	2.884
	能量（毫焦耳）	7.85	0.007	0.014	0.028	0.056	0.111	0.222

算法 1 人工神经网络-脉冲神经网络转换算法。

输入：初始权重为 $\boldsymbol{W}$ 的人工神经网络模型 $M_{\text{ANN}}(\boldsymbol{x};\ \boldsymbol{W})$；数据集 D；量化步长 L；初始动态阈值 $\boldsymbol{\lambda}$；学习率 ϵ。
输出：$M_{\text{SNN}}(\boldsymbol{x};\ \hat{\boldsymbol{W}})$

对于 $l=1$ 到 $M_{\text{ANN}}.$层，执行
    如果是 ReLU 激活，则
        将 ReLU($\boldsymbol{x}$) 替换为 QCFS($\boldsymbol{x}$; L, λ^l)
    结束 if
    如果是最大池化层，则
        将最大池化层替换为平均池化层
    结束 if
结束循环
对 $e=1$ 至训练轮次执行循环
    对数据集的长度 D 执行循环
        从 D 中采样小批量 ($\boldsymbol{x}^0, \boldsymbol{y}$)
        对第 $l=1$ 层至第 $M_{\text{ANN}}.$层执行循环
            $\boldsymbol{x}^l = \text{QCFS}(\boldsymbol{W}^l\boldsymbol{x}^{l-1};\ L, \lambda^l)$
        结束循环
        损失 = 交叉熵($\boldsymbol{x}^l, \boldsymbol{y}$)
        对于 $l=1$ 到 $M_{\text{ANN}}.$层，执行
            $\boldsymbol{W}^l \leftarrow \boldsymbol{W}^l - \epsilon\frac{\partial Loss}{\partial \boldsymbol{W}^l}$
            $\lambda^l \leftarrow \lambda^l - \epsilon\frac{\partial Loss}{\partial \lambda^l}$
        结束循环
    结束循环
结束循环
对于 $l=1$ 到 $M_{\text{ANN}}.$层执行
    $M_{\text{SNN}}.\hat{\boldsymbol{W}}^l \leftarrow M_{\text{ANN}}.\boldsymbol{W}^l$
    $M_{\text{SNN}}.\theta^l \leftarrow M_{\text{ANN}}.\lambda^l$
    $M_{\text{SNN}}.\boldsymbol{v}^l(0) \leftarrow M_{\text{SNN}}.\theta^l/2$
结束循环
返回 M_{SNN}