SpikeSMOKE: 带有跨尺度门控编码的单目 3D 目标检测的尖峰神经网络

Xuemei Chen, Huamin Wang, Hangchi Shen, Shukai Duan *Member, IEEE,* Shiping Wen, *Senior Member, IEEE,* and Tingwen Huang, *Fellow, IEEE*

***Abstract*—由于在自动驾驶等领域的广泛应用导致的能耗增加，3D 目标检测的低能耗成为一个重要的研究领域。具有低功耗特性的脉冲神经网络（SNNs）可以为这一研究提供一种新颖的解决方案。因此，我们在本文中将 SNNs 应用于单目 3D 目标检测，并提出了 SpikeSMOKE 架构，这是一个针对低功耗单目 3D 目标检测的新尝试。众所周知，与人工神经网络（ANNs）相比，SNNs 的离散信号会导致信息损失，并限制其特征表达能力。为了解决这个问题，受到生物神经元突触过滤机制的启发，我们提出了一种跨尺度门控编码机制（CSGC），该机制通过结合注意力方法的跨尺度融合和门控过滤机制来增强特征表示能力。此外，为了减少计算并加快训练速度，我们提出了一种新颖的轻量化残差块，该块可以保持脉冲计算范式并达到最高可能的检测性能。与基础的 SpikeSMOKE 在 3D 目标检测中的表现相比，结合 CSGC 的 SpikeSMOKE 在 $AP|_{R_{11}}$ 的 KITTI 自动驾驶数据集上的 0.7 IoU 阈值下分别可以达到 11.78 (+2.82，简单)、10.69 (+3.2，中等) 和 10.48 (+3.17，困难)。需要注意的是，与 SMOKE 的结果相比，SpikeSMOKE 的结果可以显著减少能耗。例如，在困难类别中能耗可以减少 72.2 %，而检测性能仅下降 4 %。轻量化的 SpikeSMOKE-L 可以在参数数量上减少 3 倍、计算量上减少 10 倍。**

***Index Terms*—Spiking Neural Networks, Monocular 3D Object Detection, Gating Coding, Lightweighting.**

I. 介绍

尽管 人工神经网络在包括计算机视觉（CV）[?] 、语音识别（SR）[?] 、自然语言处理（NLP）[?] 等多个领域表现出色，但它们仍然面临着能耗急剧增加的问题，尤其是随着模型日益复杂以及数据体量不断增大。为了解决这个问题，一些研究人员开始转向脑启发的尖峰神经网络（SNNs），因为它们具有事件驱动计算、具有生物学可解释性 [?] 、以及异步时间处理等优秀特性。目前，已经有大量优秀的研究将尖峰神经网络应用于计算机视觉（CV）、自然语言处理（NLP）以及其他领域 [?] [?] [?] [?] [?] [?] [?] [?] 。例如，从 Spiking ResNet [?] 到 EMS-ResNet [?] 再到 MS-ResNet [?] 的进步引入了越来越高效的事件驱动网络架构。Spikformer [?] 是 Transformer 在 SNNs 中的初始应用，随后不断的改进，引入了 Spike-Driven Transformer v1/v2 [?] [?] 。这一创新提供了强大的序列建模能力和高效的并行处理效率以应对复杂任务。对于目标检测，目前的研究集中在利用 2D 图像数据进行 2D 目标检测或利用 3D 数据进行 3D 目标检测。例如，在 Spiking-PointNet [?] 中，利用点云（一种表示 3D 空间信息的数据）进行 3D 目标检测，实现了点云在 SNNs 中的首次成功应用。在 [?] 中，提出了用于 2D 图像数据的尖峰驱动目标检测框架 SpikeYOLO。然而，由于缺乏深度信息，利用 2D 图像数据进行 3D 目标检测的单目 3D 目标检测尚未通过 SNNs 进行探索，无法利用 SNNs 的低功耗特性。单目 3D 目标检测结合了 2D 图像信息和深度估计以实现 3D 目标的检测和定位，已广泛应用于自动驾驶领域 [?] [?] [?] [?] [?] 。目前，人工神经网络领域有许多优秀的模型。例如，MonoWAD [?] 引入了一种创新的用于单目 3D 目标检测的天气自适应扩散模型，LabelDistill [?] 提出了用于基于摄像头的 3D 目标检测的标签引导跨模态知识蒸馏，Mono3DVG [?] 通过视觉外观的详细描述提供了一种检测 3D 物体的新视角，而 MonoCD [?] 是一种基于辅助信息的网络架构。然而，当将单目 3D 目标检测应用于资源受限的边缘设备（如自动驾驶 [?] ）时，其实际部署将受到限制。因此，实现具有低功耗特性的单目 3D 目标检测网络模型非常重要。天然具有低功耗特性的脑模型如脉冲神经网络（SNNs），可以为我们提供解决该问题的新见解。基于上述分析，我们应用 SNNs 于单目 3D 目标检测以构建一种新颖的架构 SpikeSMOKE。为了缩小 SNNs 离散信号与 ANNs 连续信号在特征表达性上的差距，我们提出了一种受生物神经元突触过滤机制启发的跨尺度门控编码机制（CSGC），该机制可以通过融合多尺度注意力来减少信息损失并提高模型的检测性能，如图 1 所示。为了在保持脉冲计算范式的同时进一步降低模型的功耗，我们创新性地提出轻量级残差块。在自动驾驶 KITTI 数据集和 CIFAR10/100 数据集上进行了大量实验。本文的贡献可以总结如下。

(1) 为了提高单目 3D 目标检测的能效，我们从以低功耗特性著称的大脑类 SNNs 中获得灵感，并利用 SMOKE 网络的简单性构建了一种名为 SpikeSMOKE 的新颖架构。

(2) 在这个 SpikeSMOKE 中，我们引入了基于注意力机制的并行跨尺度门控编码机制 CSGC，以提高模型的信息表示能力。同时，我们也在 SpikeSMOKE 模型上提出了一个轻量级的残差块。

(3) 与 SMOKE 相比，SpikeSMOKE 在 3D 物体检测中的能量消耗可显著降低。例如，在难度大的类别中，能量使用可以减少 72.2 %，而检测性能仅下降 4 %。

(4) 结合 CSGC 的 SpikeSMOKE 在 KITTI 数据集上可以取得比基础 SpikeSMOKE 更好的结果。此外，为了验证所提出的 CSGC 编码策略的通用性，我们还在 CIFAR-10/100 数据集上进行了实验来验证这一点。

本文的组织结构如下。在第 II 节中，我们描述了用于尖峰神经元的 LIF 动力学公式、单目 3D 物体检测的模型架构、编码机制和轻量化方法。在第 III 节中，我们描述了 SpikeSMOKE 的宏观架构、跨尺度门控编码 (CSGC) 的详细实现和轻量化的残差结构。在第 IV 节中，我们展示了我们方法的实验结果和消融实验的结果。在第 ?? 节中，我们总结了本文。

II. 相关工作

A. 尖峰神经元

脉冲神经元模拟生物神经元通过突触传递信号序列的行为，通过膜电压 U 传递时间信号和空间信息。脉冲神经网络中常用的神经元由积分-发火（IF）[?] 神经元和泄露积

分-发火（LIF）[?] 神经元组成。我们使用了 LIF 神经元，其神经动态公式描述如下：

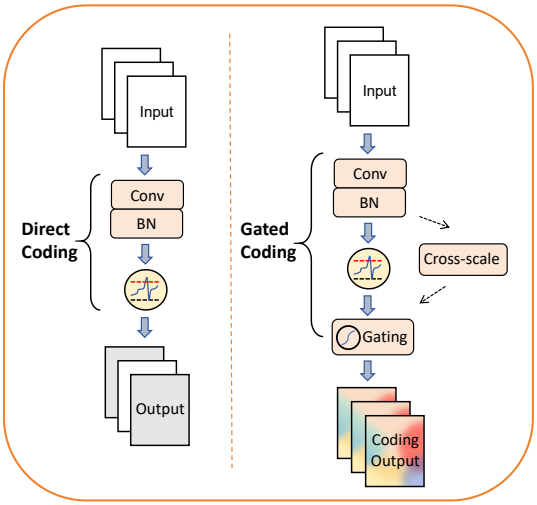

Fig. 1: 门控编码不同于直接编码。所提出的门控编码机制利用结合了一种门控单元的跨尺度融合模块（CSGC），以模拟生物神经元突触的过滤机制，从而对输入特征图产生过滤效果。

$$
\begin{aligned}
U_i^l[t] &= H_i^l[t-1] + \sum_j W_{ij} S_j^l[t] \\
S_i^l[t] &= Hea(U_i^l[t] - Uth) = \begin{cases} 1, U_i^l[t] > U_{th} \\ 0, U_i^l[t] < U_{th} \end{cases} \\
H_i^l[t] &= \tau U_i^l[t](1 - S_i^l[t]) + U_{reset} S_i^l[t]
\end{aligned} \tag{1}
$$

$U_i^l[t]$ 表示第 l 层中第 i 个神经元的膜电压，W_{ij} 表示第 l 层和第 l-1 层的连接权重，$S_j^l[t]$ 表示第 l-1 层中第 j 个神经元的脉冲，Hea 是 Heaviside 函数。在离散时间步 T，当神经元的膜电压大于脉冲神经元阈值，即 $U_i^l[t]>U_{th}$ 时，神经元会发出一个脉冲 [?] 。τ 表示衰减系数。发出脉冲 1 后，膜电压通常会重置为 0。

常见的单目 3D 目标检测方法包括两阶段方法（MonoLSS [?] ）、单阶段基于锚的方法（GrooMeD-NMS [?] ）和单阶段无锚的方法（SMOKE [?] 、MonoCD [?] 、MonoDGP [?] ）。与第一种容易引入二维噪声 [?] 的方法相比，以及第二种通常需要复杂的数据预处理和非最大值抑制 [?] [?] 的方法相比，第三种方法具有模型简单和计算效率高的优势。在本文中，我们使用单阶段无锚方法的 SMOKE [?] 架构作为基础架构，该架构既忽略了与二维检测框架相关的冗余，也不需要额外的数据。SMOKE 架构继续采用 centernet [?] 的关键点检测，抛弃了传统的二维检测模块，仅保留 3D 检测部分，通过多步解耦提高了参数收敛性和检测精度。3D 边界框是通过预测图像平面上物体的 3D 投影中心以及属性变量而获得的。3D 边界框的属性回归被解耦为一个 8 元组 $(\alpha_x, \alpha_y, \alpha_z, \alpha_l, \alpha_w, \alpha_h, sin\beta, cos\beta)$ 。这些参数通过损失函数进行优化，并转换以获得用于构建 3D 边界框的真实参数 $(x, y, z, w, h, l, \theta)$ 。

B. 编码机制

在尖峰神经网络中，首先需要将数据编码成尖峰形式，有几种常见的编码方式：率编码 [?] 基于神经元发射的平均速率编码信息，时相编码 [?] 通过尖峰相对于某个参考事件的时间位置表达，时间编码 [?] 、[?] 、[?] 基于尖峰的时间间隔和顺序表示信息，频率编码则涉及在特定时间段内神经元发射尖峰的频率。此外，还有各种复杂的编码方法（ [?] [?] [?] [?] ），选择合适的编码方法对于模型的表现力至关重要。为了减少尖峰神经网络中离散信号和连续信号之间的差距，我们提出了一种新颖的门控编码方法，并结合跨尺度融合单元。

轻量级模型指的是通过移除或转换部分模型架构和连接，来减少参数数量和计算量，并压缩模型训练时间。神经网络中有几种常见的轻量化模型方法：a) 剪枝：Chen 等人提出了一种基于神经元重要性的 SNN 剪枝再生机制，以去除网络中不必要的连接或神经元。SSNN 设计了一种通过随机初始化模型剪枝的稀疏 RSNN。b) 模型量化：Spiking-Diffusion 提出了基于矢量量化离散化的扩散模型。c) 模型蒸馏：TSSD 提出了一种时空蒸馏方法。d) 深度可分离卷积：首次在 MobileNets 中提出，后来在 ANN 领域得到了广泛的扩展和应用。深度可分离卷积的有效性已有大量实验或理论证据，我们提出的轻量级残差单元即源于这一思想。

如图 2 所示，

III. 方法

A. SpikeSMOKE

SpikeSMOKE 架构以单级物体检测网络 SMOKE 为基础设施，并保留其宏观架构，因为该架构可以忽略二维检测框架的冗余并且不需要额外的数据。

主干网络：由于 DLA34 可以通过深度收敛特征融合不同尺度的特征图，我们使用脉冲神经元的脉冲发放率来模拟其 ReLU 激活函数，将其转换为 Spike-DLA34 作为我们的主干网络。

颈部：为了确保模型完全由脉冲驱动，我们在 DLAup 结构中的每个普通卷积之前添加一个脉冲神经元，这可以利用可变形卷积来提高特征表示能力。

头：该组件由两个分支组成。一个是用于关键点分类的热图，另一个是 3D 边界框回归。这两个分支可以处理来自 Neck 网络的特征图，以获得 3D 对象检测结果。

对于 ANN2SNN，由于连续信号被转换为离散信号，信息损失是不可避免的。为了提高特征表达能力，我们提出了一个名为 CSGC 的跨尺度融合注意力编码单元。该方法可以利用不同的卷积核来融合跨尺度的上下文信息。

对于 CSGC，我们设计了一个平行架构，包括通道注意力、空间注意力和一个门控单元。输入 $x \in R^{B \times C \times H \times W}$ 首先在每个时间步上重复，以获得时间维度的信息。然后，在通道注意力部分中，我们使用一个线性-ReLU-线性结构，通过方程（2 ）来学习和更新通道维度中的权重：

$$
CA(x) = Linear\left(ReLU(Linear(x))\right). \tag{2}
$$

在空间注意力部分，我们通过不同的卷积核（方程（3 ））对相同的特征图进行特征提取。使用小卷积核来捕获小目标的局部细微特征，使得能够检测到小尺寸的物体。相反，较大的卷积核用于覆盖较大的区域以检测大物体，并有效地捕捉全局信息。我们采用三种不同大小的卷积核，并通过可学习参数 α 、β 和 γ 动态地为不同卷积处理后的特征图分配权重。同时，我们调用 [?] 的残差思想，将原始特征图与输出部分相连接。

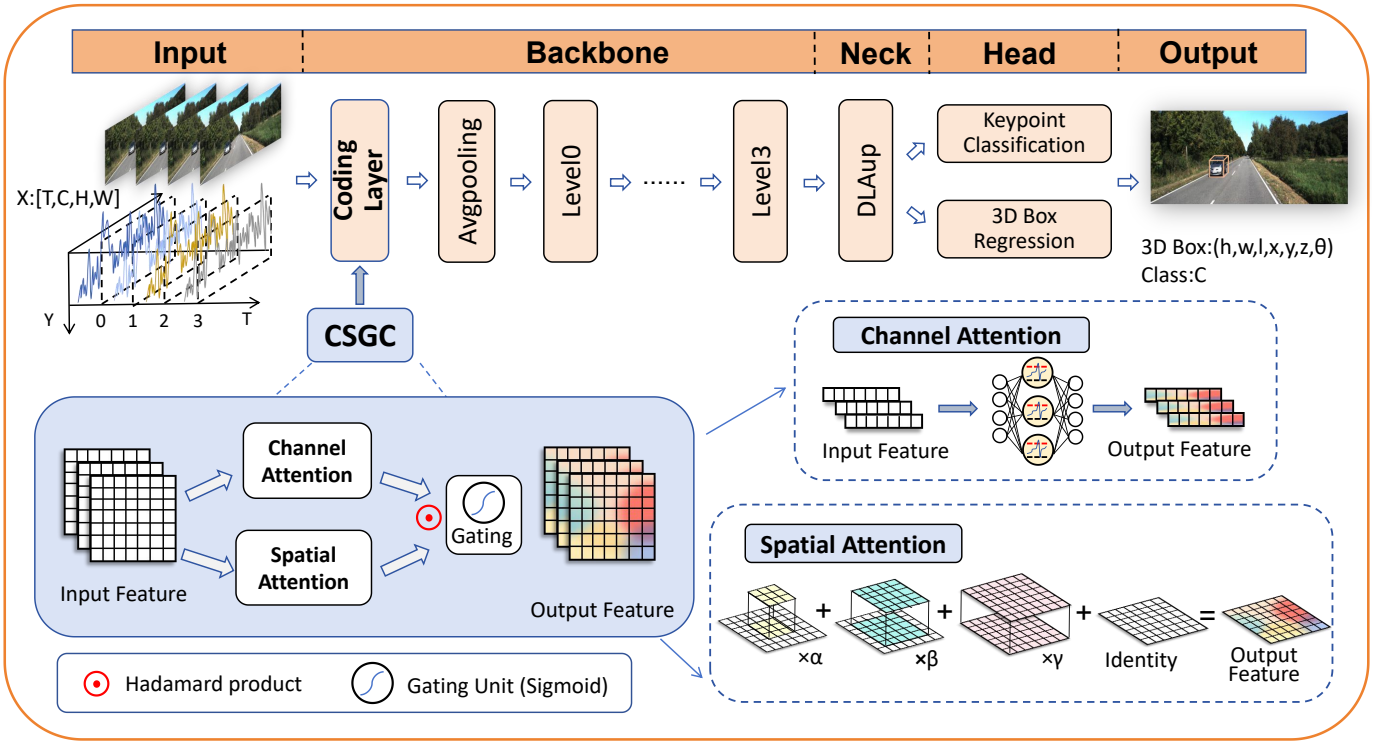

Fig. 2: SpikeSMOKE 整体网络架构。输入是一张图像，输出是物体检测的 3D 包围盒和物体类别。引入了一种基于跨尺度注意力融合的 CSGC 编码机制，以减少数据转换带来的性能损失并增强复杂特征的表示能力。

$$\begin{aligned} SA(x) =& Conv_{3\times3}\left(ReLU\left(Conv_{3\times3}(x)\right)\right)\cdot\alpha + \\ & Conv_{5\times5}\left(ReLU\left(Conv_{5\times5}(x)\right)\right)\cdot\beta + \\ & Conv_{7\times7}\left(ReLU\left(Conv_{7\times7}(x)\right)\right)\cdot\gamma + \\ & identity. \end{aligned} \tag{3}$$

经过上述通道和空间注意力处理后的输出尖峰信息由一个 Sigmoid 门控单元控制和调整，该单元定义为方程（4）：

$$O(x) = Sig\left(SA(x)\odot CA(x)\right), \tag{4}$$

，其中 $\odot$ 表示 Hadamard 积。然后，它与脉冲神经元的输出进行 Hadamard 积运算，模拟突触的过滤功能，其定义由方程（5）：

$$\hat{O}(x) = O(x)\odot SNN(x). \tag{5}$$

B. 轻量级残差

引入了一种新颖的轻量级残差模块，如图 3 所示，通过结合深度可分离卷积和膜短路方法来减少模型参数和计算负载，从而加快其训练速度。具体而言，深度卷积用于对每个输入通道进行单独的卷积操作，这意味着可以显著减少冗余计算，因为每个通道仅进行一次卷积。然后，对深度卷积的输出施加逐点卷积。之后，再次施加另一个深度卷积。需要注意的是，采用了基于膜的捷径路径以保持身份映射。

对于常规的基于膜的残差块，计算量和参数量可以分别通过方程（**??**）和方程（**??**）计算：当引入上述轻量残差块时，计算量和参数量可以分别通过方程（6-1）和方程（6-2）计算：

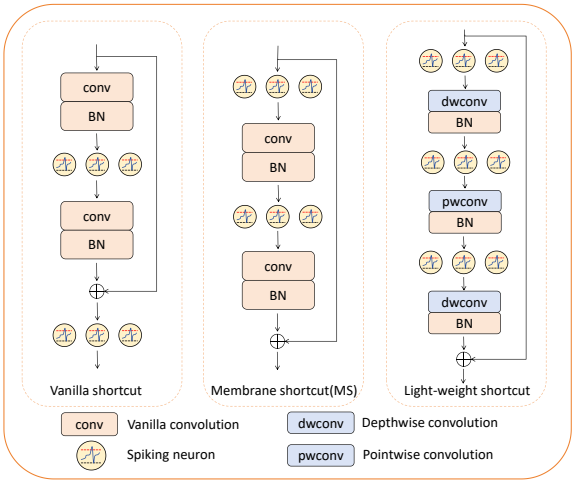

Fig. 3: 轻量级残差块。最右侧的图使用深度卷积和点卷积替代标准卷积。

$$k\cdot k\cdot Cin\cdot Wout\cdot Hout + Cin\cdot Cout\cdot Wout\cdot Hout, \tag{6-1}$$

$$k\cdot k\cdot Cin + Cin\cdot Cout. \tag{6-2}$$

对于使用轻量级残差块的 SpikeSMOKE，卷积核大小 k 为 3×3，$C_{\text{in}} = 2C_{\text{out}}$，$W_{\text{out}} = \frac{1}{2}W_{\text{in}}$，$H_{\text{out}} = \frac{1}{2}H_{\text{in}}$。

然后，可以通过方程（7 ）计算出计算工作的比例：

$$\frac{Cin \cdot Wout \cdot Hout(3k^2 + 2Cin)}{k^2 \cdot CinCout \cdot Wout \cdot Hout(1 + 4Cin)} = \frac{1}{2Cin} + \frac{1}{27}. \tag{7}$$

。同样，它们的参数量比率是方程（8 ）给出的：

$$\frac{k^2 \cdot Cin + Cin \cdot Caut + k^2Cout}{k^2Cin \cdot Cout + k^2Cout \cdot Cout} = \frac{1}{2Cin} + \frac{1}{27}. \tag{8}$$

。显然，与改进前的阶段相比，所提出的轻量级残差块在计算和参数量方面减少了大约 27 倍。

IV. 实验

A. 数据集

KITTI.

KITTI 数据集广泛应用于自动驾驶和计算机视觉领域。单目 3D 目标检测利用单摄像机数据进行汽车、行人和自行车分类的立体目标检测，训练集中有 3,712 张图片，验证集中有 3,769 张图片，涵盖了城市道路和高速公路等多种场景。检测到的目标包括 3D 边界框、鸟瞰图边界框、2D 边界框，并根据检测复杂性分为简单、中等和困难类别。我们使用了一种名为 11 点插值平均精度的评估指标 [?] ，该指标通过近似每个难度级别上的精度/召回曲线来定义，具体定义为公式（9 ）：

$$AP|N = \frac{1}{|N|} \sum_{n \in N} f_{inter}(n) \tag{9}$$

其中 N 表示代表精确区间的召回水平的集合，此实验取 $R_{11} = \{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \cdots, 1\}$ 。$f_{inter}(n)$ 表示通过插值计算获得的精度值，不是取精度值的平均值，而是在计算精度时取大于或等于当前召回的最大精度。它被定义为方程式（ 10 ）：

$$f_{inter}(n) = \max_{n':n' \geq n} f(n') \tag{10}$$

CIFAR-10/100。

CIFAR-10 和 CIFAR-100 数据集常用于图像分类。CIFAR-10 数据集包含 10 个类别的彩色图像，每个类别包含 6000 张 32x32 像素大小的图像，总共 60,000 张图像。CIFAR-100 数据集是在 CIFAR-10 基础上扩展到 100 个类别，每个类别包含 600 张 32x32 像素大小的图像。总共包含 60,000 张 32x32 像素大小的图像。

我们使用随机水平翻转、随机缩放和移动作为数据增强，旨在增加训练数据的多样性。在网络设计中，我们将组归一化的组数设置为 32，对于少于 32 的情况，我们取 16。基于文献 [?] 的分析，我们设置输入图像的分辨率为 1280x384，经过多次下采样后，尺寸减少为原始尺寸的 32 倍。我们使用 4×4090 个 GPU 进行了 172 个 epoch 的训练，每批次的大小为 4，学习率设置为 1.25×10^{-4} 。学习率的衰减策略是在第 47 和第 90 个 epoch 将学习率降低到原来的 10 倍。我们使用 0.25 的阈值来过滤检测目标。

B. 主要结果

在 KITTI 数据集上的目标检测性能

我们在 KITTI 数据集的验证集和测试集上验证了三维目标检测和鸟瞰图性能，分为简单、中等和困难三个等级。* 表示我们在配置环境并基于 SMOKE 进行训练后获得的实验结果。我们提出的带有 CSGC 的 SpikeSMOKE (SpikeSMOKE-CSGC) 的 3D 目标检测在简单、中等和困难三个等级下可以分别达到 28.83/11.78(简单)、22.75/10.69(中等)、19.44/10.48（困难）；对于 BEV 检测，它们可以分别达到 36.23/15.67（简单）、26.88/13.68（中等）、25.75/11.83（困难），如表 I 所示，其指标按 0.5/0.7 IoU 阈值由 $AP|_{R_{11}}$ 评估。从这些实验结果中可以发现，虽然我们的检测性能与 SMOKE-ANN* 相比有些许差距，但能量消耗已显著降低。例如，我们计算了困难类别在 0.7 IoU 阈值下的能量消耗，发现它可以减少 72.2 %，而检测性能仅降低 4 %。计算能量消耗的公式如下：

$$\begin{aligned} EC_{SNN} &= Synapsed^{SNN}_{activated} \times 0.9, \\ EC_{ANN} &= FLOPS^{ANN} \times 4.6, \end{aligned} \tag{11}$$

其中 0.9 表示每次累加运算的消耗，4.6 表示乘法字运算的消耗。众所周知，单目 3D 物体检测通常用于资源受限的场景，例如边缘设备、嵌入式设备等，因此轻量化是研究的重要部分。基于这一点，我们进一步通过轻量化残差块来讨论 SpikeSMOKE-CSGC 的轻量化。实验显示，参数数量仅为 SMOKE-ANN* 的 1/3，而计算量仅为 1/10，如表 II 所示。因此，SpikeSMOKE 模型可能为降低单目 3D 物体检测的能耗并提高其能源效率提供了一种新的有效解决方案。与基准模型 SpikeSMOKE 相比，所提出的 SpikeSMOKE-CSGC 在 3D 物体检测和 BEV 检测方面均表现出显著提升，3D 物体检测的提升分别为 2.82、3.2 和 3.17，BEV 检测的提升分别为 3.6、3.73 和 2.51，如表 I 所示。经过轻量化处理的 SpikeSMOKE-CSGC (SpikeSMOKE-LCSGC) 的结果也显示出相比 SpikeSMOKE-L 显著的增强，如表 II 所示。因此，改进的 CSGC 方法对 2D/3D 物体检测具有显著影响。

CSGC 的效率和广泛性验证。

我们在分类数据集 CIFAR-10/100 上验证了所提议的 CSGC 编码策略在 MS-ResNet18 上的分类结果，如表 IV 和表 V 所示。我们注意到，使用 CSGC 编码的 MS-ResNet18 在 CIFAR-10 上的准确率比直接编码方法高 1.06 %。在 CIFAR-100 上，所提议的方法 MS-ResNet18-SNN 结合 CSGC 编码在 6 个时间步长内可以达到 79.58 %，比 MS-ResNet18-SNN 高 3.17 %。因此，我们所提出的 CSGC 编码方法是有效且具有广泛适应性的。

各种注意机制效果的消融研究。

我们针对 CSGC 中的空间注意力 (SA) 和通道注意力 (CA) 在时间步长为 4、6 和 8 的情况下进行了二维目标检测的消融实验，如图 5 所示。我们观察到，当使用 SA 或 CA 时，检测性能优于基线，这意味着 CSGC 的每个模块都是有效的。显然，当 CA 和 SA 一起使用时，它们的检测结果更好。此外，随着时间步长的增加，检测性能会更好。

单个神经元阈值的消融研究。

由于尖峰神经元的阈值对模型的检测性能有很大影响，我们针对不同的神经元阈值在 2D 检测上进行了一些消融实验，以便找到最合适的 LIF 神经元阈值。基于实验结果，我们得知当神经元阈值 vth=0.75 时获得最佳性能，这也是本文中使用的阈值，如表 VI 所示。

定性结果。

我们提供了在 KITTI 数据集上单目 3D 物体检测的可视化结果，如图 4 所示。通过这些结果，我们可以直观地观察模型在复杂道路场景中准确识别和定位 3D 物体的能

TABLE I: 3D 物体检测和基于鸟瞰图（BEV）的检测结果是在 KITTI 数据集的验证集上针对汽车类别进行的。

Methods	Parameters (M)	Power (pJ)	3D Object Detection			Birds' Eye View		
			Easy	Moderate	Hard	Easy	Moderate	Hard
SMOKE-ANN*(0.5)	19.51	2.17E+11	29.16	25.22	21.47	32.52	29.59	25.86
SpikeSMOKE(0.5)	19.51	5.97E+10	20.87	19.72	16.89	27.64	22.47	21.65
SpikeSMOKE-CSGC(0.5)	19.56	6.04E+10	28.83	22.75	19.44	36.23	26.88	25.75
SMOKE-ANN(0.7) [?]	19.51	2.17E+11	14.76	12.85	11.5	19.99	15.61	15.28
SMOKE-ANN*(0.7)	19.51	2.17E+11	12.03	11.14	10.92	17.97	13.08	12.06
SpikeSMOKE(0.7)	19.51	5.97E+10	8.96	7.49	7.31	12.07	9.95	9.32
SpikeSMOKE-CSGC(0.7)	19.56	6.04E+10	11.78	10.69	10.48	15.67	13.68	11.83

(0.5/0.7) indicates that the metrics are evaluated by $AP|_{R_{11}}$ at the 0.5/0.7 IoU thresholds.

TABLE II: 的三维物体检测和鸟瞰图检测结果是在 KITTI 数据集的汽车类别的测试集上进行的。

Methods	Parameters (M)	Power (pJ)	3D Object Detection			Birds' Eye View		
			Easy	Moderate	Hard	Easy	Moderate	Hard
SMOKE-ANN*	19.51	2.17E+11	28.97	23.57	20.21	30.48	27.76	25.81
SpikeSMOKE	19.51	5.97E+10	21.80	15.23	14.99	25.99	20.69	17.92
SpikeSMOKE-L	6.32	2.24E+10	19.46	10.52	10.23	24.01	15.74	15.37
SpikeSMOKE -LCSGC	6.37	2.29E+10	20.64	15.32	12.93	25.15	18.98	18.08

The table is evaluated using the $AP|_{R_{11}}$ with a 0.5 IoU threshold.

Fig. 4: 单目 3D 物体检测在 KITTI 数据集上的可视化结果清晰直观地展示了检测算法在识别和定位数据集中复杂道路场景中的 3D 物体时的性能和准确性。

TABLE III: 车类的二维目标检测结果是在 KITTI 数据集的验证集上进行的。

Methods	2D Object Detection		
	Easy	Moderate	Hard
SMOKE-ANN*	80.49	72.01	68.76
SpikeSMOKE	73.32	62.85	55.67
SpikeSMOKE-CSGC	75.58	65.49	64.37
SpikeSMOKE-L	51.59	44.01	42.71
SpikeSMOKE-LCSGC	52.33	49.04	43.04

TABLE IV: 在分类任务
数据集 CIFAR-10 上的分类结果。

Architecture	Coding Schemes	Time Steps	CIFAR10 Acc. (%)
ResNet-19 [?]	Phase Coding	8	91.40
VGG-16 [?]	Temporal Coding	100	92.68
ResNet-19 [?]	Rate Coding	6	93.16
MS-ResNet-18	Direct Coding	6	94.92
MS-ResNet-18	CSGC Coding	6	95.98(+1.06)

力，从而证实该技术在自动驾驶等领域具有显著的潜在应用。

V. 结论

随着 3D 物体检测在自动驾驶等应用中的广泛使用，低能耗问题越来越受到关注。众所周知，低功耗是仿脑 SNNs 的一个关键特征，为能效的 3D 物体检测提供了一个潜在的新解决方案。基于此，我们构建了一种用于单目 3D 物体检测的新颖 SpikeSMOKE 架构。由于 SNNs 的离散信号特性可能导致信息丢失及其对特征表示能力的限制，我们借鉴生物神经元中的突触滤波过程，提出了一种 CSGC 机制。此外，我们还提出了一种轻量化残差块，以减少计算量的同时保持脉冲计算范式。在 KITTI 数据集上的实验结果表明，SpikeSMOKE 相比 SMOKE 具有更高的能效，例如在 Hard 类别上能耗降低了 72.2%，而检测性能仅下降了 4%。此外，实验结果还显示，基于 CSGC 的 SpikeSMOKE 相比基线 SpikeSMOKE 取得了显著提升。SpikeSMOKE-L（轻量版）相比 SMOKE 可进一步减少 3 倍参数量和 10 倍计算量。在 CIFAR-10/100 分类任务中，CSGC 编码策略分别提高了 1.06% 和 3.17% 的准确性，验证了其通用性。总体而言，SpikeSMOKE 架构和 CSGC 机制可以为低功耗单目 3D 物体检测提供一种高效且可行的解决方案。

TABLE V: 在分类任务
数据集 CIFAR-100 上的分类结果。

Methods	Architecture	Spike	Params (M)	Time Steps	CIFAR100 Acc. (%)
ANN [?]	MS-ResNet-18	×	12.54	N/A	80.67
MS-ResNet-SNN [?]	MS-ResNet-18	✓	12.54	6	76.41
CSGC-SNN	MS-ResNet-18	✓	12.72	6	79.58
	MS-ResNet-18	✓	12.72	4	77.97

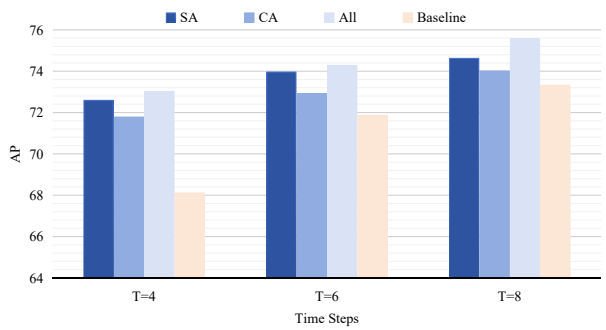

Fig. 5: 在不同时间步骤的 CSGC 中进行的不同注意力模块的消融实验。

References

TABLE VI: 针对脉冲神经元阈值的消融实验。

Methods	vth=0.25	vth=0.5	vth=0.75	vth=1.0	2D Object Detection		
					Easy	Moderate	Hard
SpikeSMOKE-CSGC	✓				64.31	53.84	51.62
		✓			72.06	61.27	60.67
			✓		75.58	65.49	64.37
				✓	54.10	45.88	44.98