Computer Science and Application 计算机科学与应用, 2025, 15(1), 187-198
Published Online January 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.151019

Hans汉斯

基于深层特征高效提取的脉冲神经网络目标检测方法

黄永斌[1]，李　晨[2]，董文波[2]，刘顺莲[3*]

[1]湖南工业大学计算机学院，湖南　株洲
[2]株洲中车时代电气股份有限公司，湖南　株洲
[3]湖南工业大学理学院，湖南　株洲

收稿日期：2024年12月23日；录用日期：2025年1月20日；发布日期：2025年1月28日

摘　要

由于出色类脑机制和自身节能性，脉冲神经网络(SNN)受到广泛研究，且在目标分类任务中取得显著进展，但其在目标检测领域的研究尚处于初步阶段。针对现有SNN目标检测算法在复杂场景下检测精度低的问题，本文构建了一种深层特征高效、轻量提取的脉冲神经网络目标检测架构——ES-YOLO。该架构提出了高效特征提取模块(SDF-Module)，并结合多尺度特征提取的空间金字塔设计，显著提升了模型在复杂目标检测任务中的性能。此外，通过引入脉冲解耦检测头，进一步优化了模型的检测精度和实时性。实验结果表明，在VOC2012数据集上，ES-YOLO模型获得了60.5%的mAP@0.5和37.1%的mAP@0.5:0.95的性能指标，相比EMS-YOLO分别提升了4%和3.7%。该模型不仅缩小了与同等架构ANN模型的性能差距，而且整体能耗为同等架构ANN的1/5。为后续SNN在目标检测任务中的广泛应用提供支持。

关键词

多尺度特征，深层特征，高效提取，脉冲解耦检测头

Spiking Neural Network Target Detection Method Based on Efficient Deep Feature Extraction

Yongbin Huang[1], Chen Li[2], Wenbo Dong[2], Shunlian Liu[3*]

[1]School of Computer Science, Hunan University of Technology, Zhuzhou Hunan
[2]Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou Hunan
[3]School of Science, Hunan University of Technology, Zhuzhou Hunan

Received: Dec. 23rd, 2024; accepted: Jan. 20th, 2025; published: Jan. 28th, 2025

*通讯作者。

文章引用: 黄永斌, 李晨, 董文波, 刘顺莲. 基于深层特征高效提取的脉冲神经网络目标检测方法[J]. 计算机科学与应用, 2025, 15(1): 187-198. DOI: 10.12677/csa.2025.151019

Abstract

Due to the outstanding brain-inspired mechanisms and energy efficiency, Spiking Neural Networks (SNN) have garnered widespread attention and achieved significant progress in object classification tasks. However, research on SNNs in the field of object detection remains in its early stages. To address the issue of low detection accuracy of existing SNN-based object detection algorithms in complex scenarios, this paper proposes a novel, deep-feature-efficient and lightweight SNN-based object detection architecture—ES-YOLO. The architecture introduces an efficient feature extraction module (SDF-Module) and incorporates a spatial pyramid design for multi-scale feature extraction, significantly improving the model's performance in complex object detection tasks. Furthermore, by integrating a spiking decoupled detection head, the model's accuracy and real-time performance are further optimized. Experimental results on the VOC2012 dataset demonstrate that the ES-YOLO model achieves 60.5% mAP@0.5 and 37.1% mAP@0.5:0.95, representing improvements of 4% and 3.7% respectively compared to EMS-YOLO. The model not only reduces the performance gap between SNNs and equivalent ANN models but also achieves overall energy consumption that is only 1/5 that of the equivalent ANN architecture. This work provides support for the broader application of SNNs in object detection tasks in the future.

Keywords

Multi-Scale Features, Deep Features, Efficient Extraction, Spiking Decoupled Head

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

近年来，卷积神经网络(CNN) [1]在计算机视觉领域，特别是在目标检测等复杂任务中，展现出了卓越的性能。然而，随着网络规模的不断扩展，CNN 虽然能够实现高精度[2]-[4]，但也面临着计算复杂度和能耗急剧增加的问题，这对实时应用构成了重大挑战。随着对生物神经系统研究的不断深入，学术界逐步探索更加贴近生物神经元行为的计算模型。不同于人工神经网络(Artificial Neural Networks, ANN)依赖连续激活函数进行信息处理，脉冲神经网络(Spiking Neural Networks, SNN [5]-[7])通过模拟生物神经元的脉冲发放机制，以离散的脉冲信号进行计算。相比之下，SNN 展现出更低的能耗和更优的实时处理能力。因此，SNN 被广泛认为是 ANN 的自然延伸，推动了神经网络技术朝着更高效、更节能的方向发展。

SNN 作为神经网络发展的第三代标志性成果[8]，凭借其基于脉冲信号的稀疏计算特性和事件驱动的工作机制，在低功耗移动设备和边缘计算平台上展现出巨大的潜力。然而，SNN 中二值化脉冲信号的非连续性质给直接训练带来了显著挑战，尤其在处理如目标检测这样的复杂任务时，其性能往往受到限制[9]。针对这一问题，Spiking-YOLO [10]提出了一种创新方案，即通过将人工神经网络(Artificial Neural Networks, ANN)转换为具有数千时间步的 SNN，初步构建了基于 SNN 的目标检测框架。尽管如此，这种方法通常依赖于预先训练好的 ANN 模型，并且需要长时间的模拟来达到与 CNN 相媲美的性能，这不仅牺牲了实时性(表现为较低的帧率 FPS)，还引入了较高的延迟。Spike Calibration [11]力求将时间步数减少到几百个，但其性能受到原 ANN 模型检测效果的限制。为了促进 SNN 的深度直接训练，Hu [12]和 Fang [13]分别提出了适用于分类任务的 MS-ResNet 和 SEW-ResNet 架构，成功解决了梯度消失与爆炸的问题，

实现了超过百层的深度 SNN 训练。但在目标检测任务中，当面对多尺度对象特征时，非脉冲卷积操作引起的能耗增加问题尤为明显。为此，EMS-YOLO [14]开发了一种全新的全脉冲残差块——EMS-ResNet，成为首个基于 YOLO 框架直接训练的 SNN 模型。完全避免了非脉冲卷积造成的多余乘积累加(Multiply-Accumulate, MAC)操作，相较于传统的转换或混合 SNN，展示了更好的性能和更低的能耗。最近，Meta-SpikeFormer [15]作为一种最新的 SOTA 模型，采用倒置残差结构结合脉冲驱动的 Transformer，首次以预训练和微调的方式处理目标检测任务，达到了 SNN 目标检测领域的最先进水平。不过，这些方法在脉冲神经网络目标检测方面，性能和效率仍与 ANN 存在一定的差距。

本研究旨在通过设计一种 Efficient Spiking YOLO (ES-YOLO)网络架构，将脉冲模块与经典的 CNN 特征金字塔及空间金字塔设计相结合，直接训练 SNN，以缩小其在目标检测任务上与 ANN 之间的性能差距。实验结果显示，所提出的模型能够在较短的时间步长内实现高精度和低能耗的目标检测，为 SNN 在实时视觉应用中的广泛部署提供了坚实的技术基础。

2. 相关工作

2.1. 脉冲神经网络

脉冲神经网络是一种受生物学启发的新一代人工神经网络模型。不同于传统卷积神经网络中信息以连续数值形式传递的方式，SNN 的信息处理方式更加贴近生物神经元的工作机制，即通过离散的脉冲事件来传递信息。在 SNN 中，信息的传输依赖于脉冲的生成，而脉冲的生成则基于描述生物过程的微分方程，特别是对神经元膜电位动态的模拟。当神经元[16]-[18]的膜电位累积至预设的阈值时，神经元将发射一个脉冲，并随后重置其膜电位。

在众多神经元模型中，Leaky Integrate-And-Fire (LIF) [19]模型因其简洁性和对生物神经元行为的良好模拟而被广泛采用。LIF 模型没有详细探讨膜电位变化中离子的具体动力学过程，而是集中于膜电位变化的核心机制。在这个模型中，神经元的细胞膜被视为一个电容器，能够累积外部刺激引起的电荷，从而导致膜电位的逐步升高。此外，细胞膜还被抽象为具有有限漏电电阻的组件，允许电荷随着时间的推移逐渐泄漏，这反映了神经元膜电位的自然衰减过程。与简单的 Integrate-And-Fire (IF) [20] [21]模型相比，LIF 模型能更好地捕捉到真实生物神经元的动态特性，因此在模拟生物神经系统以及处理复杂任务时，SNN 展现出更高的准确性和适应性。这种特性使得 SNN 在低功耗计算、实时处理等应用场景中具有显著优势。

2.2. Sandglass Block

在传统残差网络结构[22]中，输入和输出的通道数相对较多，而中间卷积层的通道数较少，形成所谓的“瓶颈”结构。这种结构通常包括两个 1 × 1 卷积层，用于降低和恢复维度，以及一个 3 × 3 卷积层用于特征提取，并通过一个跳跃连接将输入直接添加到经过两个卷积层后的输出上。为了解决 3 × 3 卷积层的庞大参数量这一问题，Depthwise Separable Convolutions 技术[23]被提出，该技术将标准卷积分解为 Depthwise 卷积和 Pointwise 卷积两部分，其中前者专注于单个通道上的特征提取，后者则负责跨通道的信息整合，从而有效减少了计算复杂度和参数量。

在 MobileNetV2 [24]中，引入了一种名为“倒置残差结构”(Inverted Residuals)的设计，其核心思想是在网络结构中先增加通道数，再减少，与传统瓶颈结构的设计理念相反。具体实现方式为：首先利用 1 × 1 的 Pointwise Convolution 扩展输入的通道数，接着通过 3 × 3 的 Depthwise Separable Convolution 进行特征提取，最后再用 1 × 1 的 Pointwise Convolution 将通道数压缩至原始大小。尽管倒置残差结构表现出色，但其中间扩展层所编码的特征图在经过通道压缩后可能会丧失部分有用信息。

基于以上观察，Zhou [25]提出了 Sandglass Block 设计方案，即通过反转两个 Pointwise Convolution 的顺序来优化模型，旨在进一步减少参数和计算成本。将 Depthwise Convolution 应用于高维特征图，并将其置于 Residual Path 的起始和结束位置，从而增强了对丰富空间信息的编码能力，提高了特征表示的质量。通过这种方式，Sandglass Block 不仅维持了较低的参数量和计算成本，而且确保了信息的有效传递和梯度的顺畅回传。

3. 构建 Efficient Spiking YOLO 架构

EMS-YOLO [14]作为第一个端到端的 SNN 目标检测模型，采用和 YOLO 设计相同的结构，将由 ReLU 激活函数组成的 ResNet 结构更换为由 LIF (脉冲神经网络激活单元)为激活函数组成的 EMS-ResNet 结构，并利用 STBP(时空反向传播)完成模型的训练。

针对 EMS-YOLO 在多尺度特征融合和复杂场景下的检测精度上存在不足的问题，本研究结合 YOLO 多尺度特征融合架构与 SNN 模块，设计了 ES-YOLO 模型架构(如图 1 所示)。具体而言，本文在 EMS-YOLO 的基础上，借鉴 YOLOX 的设计思路，构建了包含 Backbone、Neck 和 Head 完整架构。并将计算冗杂的 EMS-ResNet 基本结构替换成 SDF-ResNet 基本结构，使得模型精度不丢失的基础上更加轻量；其次增加 Neck [26] [27]部分，增强了多尺度特征的提取能力，从而更好地应对复杂场景下的对象检测任务；最后，借鉴了 ANN 在 head 部分解耦检设计思想，设计了脉冲解耦检测头，其用解耦方式处理不同类型的检测任务，并将检测头数量由两个增至三个，进一步提升了该模型的检测性能。

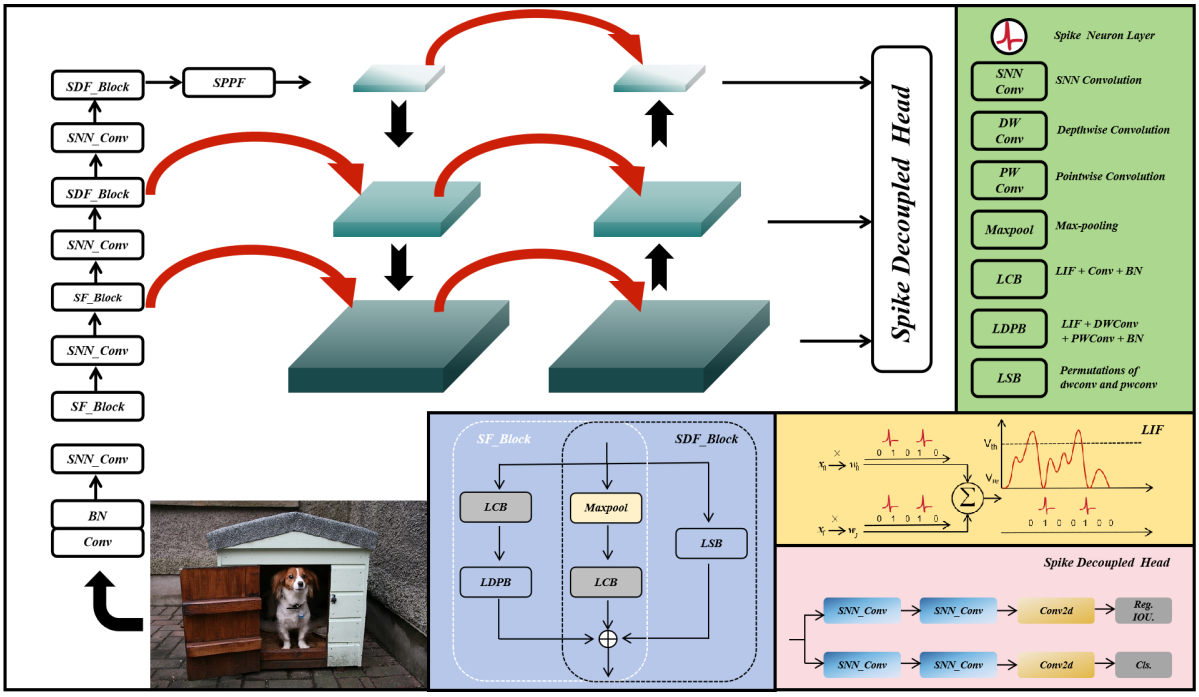


Figure 1. Overall architecture of ES-YOLO spikinng neural network
图 1. ES-YOLO 脉冲神经网络整体架构

3.1. SNN 数据输入

对于 SNN 的输入可以表示为 $X \in R^{T\times C\times H\times W}$，其中 T 是时间步长，C 是通道，$H\times W$ 表示空间分辨率。为了充分利用 SNN 的时空特性，通常第一个卷积层负责将输入转换为脉冲信号。在这一层中，输入图像

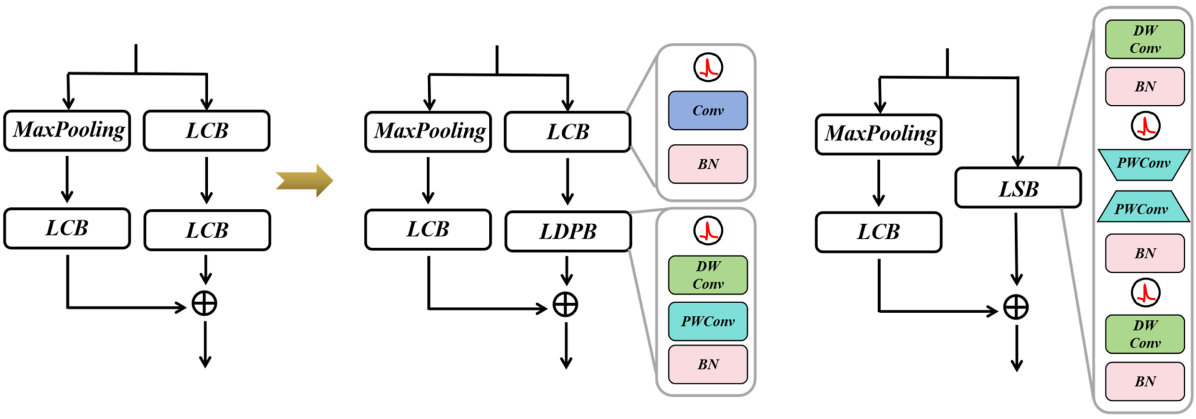
首先沿时间步长展开，形成五维张量，然后通过 LIF 神经元对加权输入进行整合。

当膜电位超过预设的激发阈值时，LIF 神经元会产生脉冲序列。这些脉冲序列作为网络的输出，传递到后续层进行进一步处理，这一过程称为直接输入编码[28] [29]。随后，通过 SNN_Conv 和 SDF-Module 从不同的维度和通道数量提取对象特征，从而有效增强网络的整体性能。SNN_Conv 层通过脉冲驱动的卷积操作提取时空特征，而 SDF-Module 则进一步处理和融合这些特征，进一步推动了网络在目标检测任务中的表现。

3.2. Slim Deep Feature Modulet 设计

在使用 EMS-ResNet 构建过深的脉冲神经网络时，发现存在计算效率低和模型架构大的问题。为了解决该问题，本文引入了 Sandglass Block 的设计思想，其核心在于将 Shortcut 连接置高维度特征之间，以更好地保留信息传递和梯度回传。且相比于标准卷积每个卷积核是同时操作输入图片的每个通道，Depthwise Conv 的一个卷积核负责一个通道，一个通道只被一个卷积核卷积，保证了在不显著增加计算成本的情况下，对高维度特征进行有效的特征提取。

结合 EMS-Module 和 Sandglass Block 的思想，本文设计了两种轻量级的 SDF-Module，分别为 SF Block 和 SDF Block，如图 2 SDF-Module 设计图所示。SF Block 主要用于浅层特征提取，而 SDF Block 则用于深层特征提取。这两种模块均采用了深度可分离卷积和逐点卷积替代部分原有的 LCB 模块，以进一步减少模型的参数量和计算复杂度，同时保持模型的性能。



(a) EMS-Module　　(b) SF Block　　(c) SDF Block

Figure 2. SDF-Module design diagram
图 2. SDF-Module 设计图

SF Block 数学表达式可以表示为：

$$X = \mathrm{BN}\left(\mathrm{Conv}\left(\mathrm{LIF}\left(\mathrm{MaxPool}\left(x\right)\right)\right)\right) \tag{1}$$

$$W' = \mathrm{BN}\left(\mathrm{Conv}\left(\mathrm{LIF}\left(x\right)\right)\right) \tag{2}$$

$$W'' = \mathrm{BN}\left(\mathrm{PWConv}\left(\mathrm{DWConv}\left(\mathrm{LIF}\left(W'\right)\right)\right)\right) \tag{3}$$

$$\mathrm{SF\ Block}\left(X\right) = X + W'' \tag{4}$$

DWConv 表示 3 × 3 的深度卷积操作，而 PWConv 表示 1 × 1 的逐点卷积操作。LIF 表示脉冲神经网络中的脉冲发放神经元激活函数，它模拟生物神经元的脉冲发放过程。通过这种模块设计，SF Block 能

够在减少计算量的前提下保持对输入数据的特征提取能力。深度卷积有效捕捉了空间特征，而逐点卷积则通过减少通道间冗余，进一步压缩参数量。通过脉冲神经元激活函数的引入，模型也能更有效地模拟时间维度上的特征变化，这为目标检测中的时序信息提取提供了可能性。

SDF Block 数学表达式可以表示为：

$$S' = \mathrm{PWConv}\left(\mathrm{PWConv}\left(\mathrm{LIF}\left(\mathrm{BN}\left(\mathrm{DPConv}\left(x\right)\right)\right)\right)\right) \tag{5}$$

$$S'' = \mathrm{BN}\left(\mathrm{DWConv}\left(\mathrm{LIF}\left(S'\right)\right)\right) \tag{6}$$

$$\mathrm{SDF\ Block}\left(X\right) = X + S'' \tag{7}$$

其中第一个 DWConv 仍然是 3 × 3 的深度卷积操作，用于提取空间特征。第一个 PWConv 用于缩小输入通道数，以减少计算量和参数规模。接着第二个 PWConv 用于扩大通道数，同时编码通道间的特征。然而，由于逐点卷积无法有效捕捉空间信息，最后在结构中增加了一个 3 × 3 的深度卷积层，以学习表达空间上下文特征。

通过这种结构，SDF Block 实现了在保持模型轻量化的同时，深层特征的有效提取。深度卷积可以捕获局部空间特征，而逐点卷积在压缩通道的同时维持了跨通道的信息传递。加上最后的深度卷积，保证了网络可以处理更加复杂的空间关系。同时，这种模块设计也确保了信息流的有效传递，防止梯度在深层网络中的消失，从而提升了脉冲神经网络在目标检测或识别任务中的效率和性能。

3.3. Neck 设计

本文在模型架构中引入 Neck 设计，是位于骨干网络(backbone)和头部网络(head)之间的模块，它负责增强或调整骨干网络提取的特征，以便更好地服务于最终的目标检测任务。该部分主要涉及两个概念，特征金字塔是一种用于多尺度目标检测的技术，旨在解决目标检测中的尺度变化问题。传统的单尺度特征图难以同时处理大目标和小目标，而 FPN 通过构建一个多尺度的特征金字塔来解决这个问题。空间金字塔 SPPF 主要用于加速 SPP 的计算过程，同时保持其多尺度特征提取的能力。

3.4. 脉冲解耦检测头(Spike Decoupled Head)设计

检测头部分是网络中专门用于目标检测的组件，接收来自 Neck 模块的多尺度特征图，并输出最终的检测结果。在本文中提出的脉冲解耦检测头(Spike Decoupled Head)概念，其设计灵感来自于经典目标检测架构中的任务分离思想，核心在于通过解耦的方式处理不同类型的检测任务。每个子网络专注于特定任务，分类子网络负责学习区分不同类别的特征，而回归子网络则专注于学习物体边界框的精确位置。这种设计不仅提高了模型的灵活性和可扩展性，还能够在保持高性能的同时，有效减少计算资源的消耗和功耗。

在卷积神经网络解耦检测头的结构中，特征通道数根据线性下降策略逐步减少，以减小计算量并提高模型的推理效率。通过使用线性插值函数对通道数进行调整，确保在卷积操作中保持特征提取的有效性与计算效率的平衡。为了适应脉冲神经网络的动态时序特性，本文在解耦检测头中引入了 SNN 卷积层(Snn_Conv2d)，以替换传统的卷积层，对分类任务和定位任务进行了有效的解耦处理。在具体实现中(如图 3 所示)，初始的 1 × 1 卷积将输入特征图的通道数进行压缩，随后在检测分支通过两层 3 × 3 卷积逐步提取特征，最后通过 1 × 1 脉冲卷积输出物体的边界框参数和置信度。分类分支同样经过两层 3 × 3 卷积后，通过 1 × 1 脉冲卷积输出类别概率分布。脉冲卷积有效地将时间信息编码进神经元的发放频率和时序，提升了时序依赖任务的表现。同时，解耦的卷积结构允许分类与定位任务独立优化，将分类和回归

任务分别交由独立的子网络处理，可以更精确地优化每个任务的性能，同时显著减少任务间的干扰。此外，利用 SNN 的时间累积处理特性，检测头能够在时间维度上对信息进行有效的累积和整合，从而进一步提高检测的准确性。

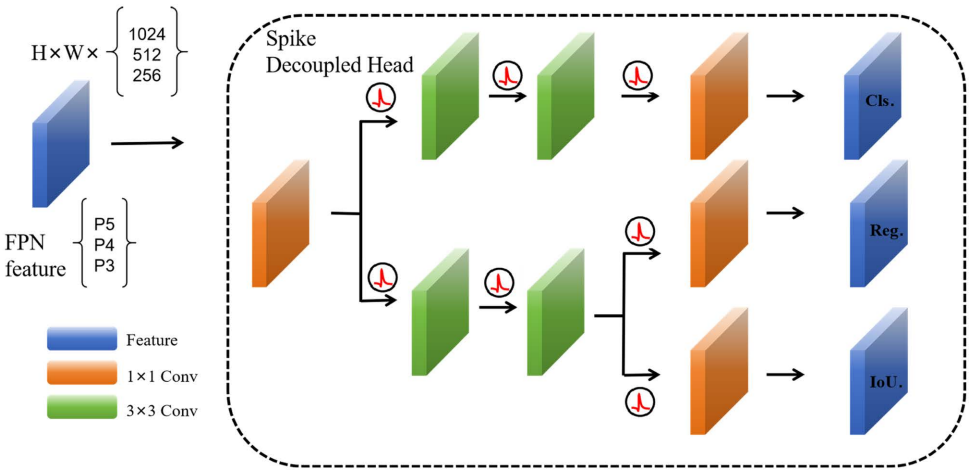


Figure 3. Spike decoupled detection head design diagram
图 3. 脉冲解耦检测头设计图

4. 实验结果与分析

本文方法在 VOC2012 数据集上进行了详细的消融实验，实验分为两个主要部分：

1. 模型对比实验：将本文提出的方法与现有的多种经典模型进行对比，以评估其在目标检测任务中的综合性能。通过比较不同模型的检测精度和能耗等指标，验证本文方法的优势。但由于 SNN 目标检测算法处于兴起阶段，现有的 SNN 目标检测算法最大的优势在于其节能性，检测精度与实效性上与最先进的 ANN 目标检测算法仍有部分差距，故现有 SNN 目标检测算法[10] [14] [15] [30]进行对比实验分析时，选择的 ANN 算法均为 ResNet34 或 YOLOv3-tiny 进行对比，因此，本文保留这一约定俗称的规则，选择 YOLOv3-tiny 进行对比，并在此基础上与 YOLOv5 进行对比进一步凸显本文精度上的提升。而对比的 SNN 方法分为两种，一种是 ANN 转 SNN 的方法，代表作是 Spikingyolo；另一种是直接训练的方法，代表作是 EMS-YOLO，本文都进行了对比分析。

2. 模型架构消融实验：评估不同网络架构对 ES-YOLO 模型性能的影响，本文主要包括 SDF-Model、Neck 和脉冲解耦检测头的设计。

4.1. 实验配置与评价指标

实验的硬件平台采用了配备 24GB 显存的 NVIDIA RTX3090 显卡进行训练与测试。为了确保实验的公平性和有效性，所有实验中均严格控制了数据预处理和训练参数的设置。具体参数配置如下：LIF 神经元的重置值 V_{reset} 被设置为 0，膜时间常数 τ 被设置为 0.25，阈值 V_{th} 被设置为 0.5，系数 α 被设置为 1，并采用 SGD 优化器，初始学习率设置为 1e−2。

本文采用平均检测精度(mean Average Precision, mAP)作为评价指标，即通过计算在不同阈值条件下，模型预测结果与真实标签的匹配程度来衡量模型的整体性能，计算公式如下：

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{10}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \tag{11}$$

其中 TP (True Positive)：真正例，表示正确预测正样本分类的数量。FP (False Positive)：假正例，表示错误地将负样本分类的数量。FN (False Negative)：假反例，表示错误地将正样本分类为负样本的数量。TN (True Negative)：真反例，表示模型正确地将负样本分类为负样本的数量。P 表示精确率(Precision)，是正确的正样本预测数量与所有正样本预测数量的比值。R 表示召回率(Recall)，是正确的正样本预测数量与所有实际正样本数量的比值。平均精确度(Average Precision, AP)，即精度－召回率曲线。N 表示检测数据集的类别数。

4.2. 网络耗能计算

SNN 相较于传统 ANN 的一个显著优势在于其更低的能耗。在神经网络的运算过程中，能量消耗通常通过浮点运算次数(Floating-point Operations, FLOPs)进行衡量。相较于 ANN 每次计算涉及浮点数的乘法和加法运算，SNN 的神经元仅在脉冲触发的时刻参与累加运算(Accumulation Calculation, AC)，在神经元静默时不消耗能量。对于单次卷积操作，SNN 的能量消耗计算如下：

$$\mathrm{FLOPs} = k^2 \times O^2 \times C_{in} \times C_{out} \tag{12}$$

$$E_{SNN} = \sum_{l=0}^{n} T \times r \times \mathrm{FLOPs}(l) \times E_{AC} \tag{13}$$

上式中，k 是卷积核大小，O 是输出特征图大小，C_{in} 和 C_{out} 分别为输入和输出维度，T 表示时间步长；r 表示脉冲激活率，n 表示卷积次数，E_{AC} 表示一次 AC 运算所消耗的能量。

在本文中，耗能计算依托于文献[31]中提到的 45 nm CMOS 神经形态芯片作为硬件平台，进行整体能耗推算。这种硬件设备在现有的 SNN [32] [33]算法的能耗计算中常被用作参考。在该硬件上，$\mathrm{E_{AC}}$ 的值为 0.9 pJ。

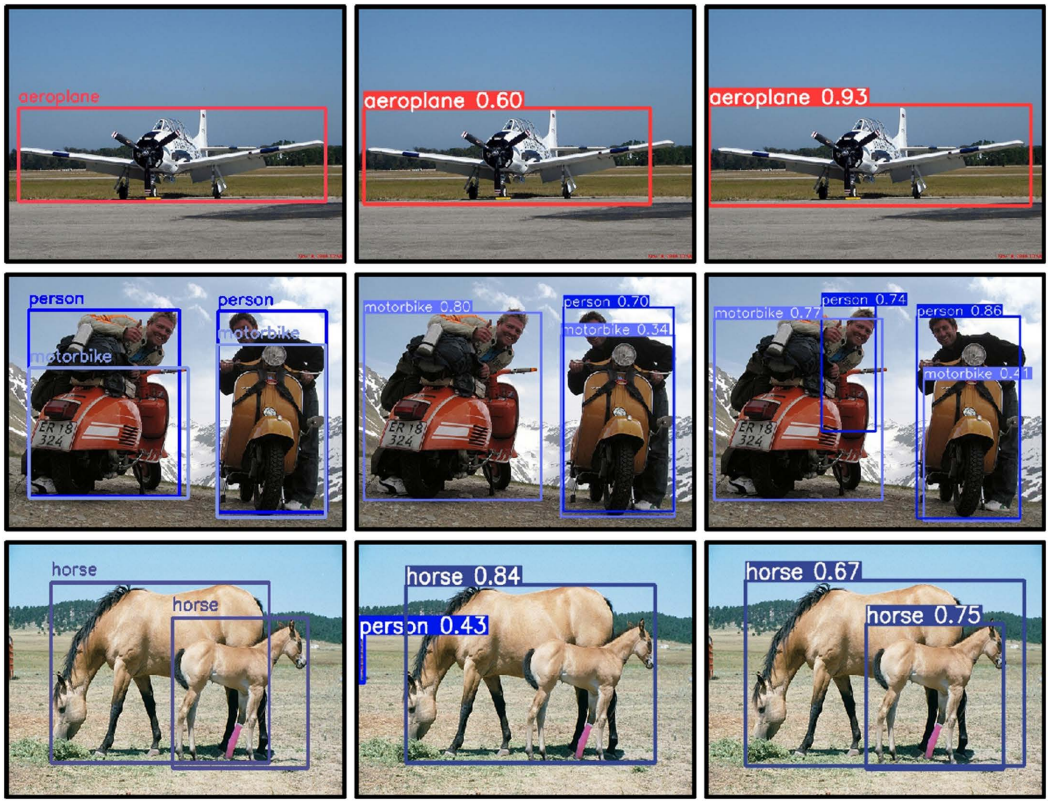
4.3. 模型对比实验

通过模型对比实验(如表 1 所示)，对比了不同模型在 ANN 和 SNN 架构下的性能表现，涵盖了时间步长、参数规模、功耗及精度(mAP@0.5 和 mAP@0.95)，且由于 ANN 网络处理的是由 ReLU 等激活函数处理的连续数值，不涉及时间步长概念，因此在表中标为“-”。在 ANN 模型中，YOLOv5 的 mAP@0.5 达到了 61.7%，虽检测性能较好，但其较高的能耗限制了实际应用的效率，相比之下，SNN 能够很好解决。通过 ANN 转 SNN 的 Spiking-YOLO 模型，在较高的时间步长下，将 mAP@0.5 提升至 51.8%，但性能仍未达到理想水平。而直接训练的 SNN 模型，EMS-YOLO 在低功耗条件下取得了较好的精度表现，当时间步长增加到 4 时，mAP@0.5 提升至 56.5%，进一步提升检测性能与能效方面的优势。

本文提出的 ES-YOLO 在直接训练 SNN 架构下，通过进一步优化网络结构，显著提高了模型精度。在时间步长为 4 时，mAP@0.5 达到了 60.5%，相比于 EMS-YOLO 提升了 4%，且功耗降低了 5.2 倍的基础上，检测效果如图 4 所示。这些结果表明，本文提出的方法在保持高性能的同时，并提高了能效，为 SNN 在实时目标检测应用中的广泛部署提供了有力支持。

Table 1. System resulting data of standard experiment
表 1. 标准试验系统结果数据

网络结构	模型名称	Epoch	时间步长	Param (M)	Power (mJ)	mAP@0.5/%	mAP@0.95/%
ANN	YOLOv3-tiny	-	-	8.85	-	52.3	-
	YOLOv5	150	-	21.2	112.5	**61.7**	**39**
ANN2SNN	Spiking-YOLO [10]	-	3500	10.2	-	51.8	-
Directly-trained SNN	EMS-YOLO [14]	150	1	33.9	7.25	49.7	27.2
			4	33.9	29	56.5	33.4
	本文方法		1	**27.8**	**5.4**	56.6	33.4
			4	27.8	21.6	60.5	37.1



(a) 真实值 (b) EMS-YOLO (T = 4) (c) ES-YOLO (T = 4)

Figure 4. VOC2012 data detection chart
图 4. VOC2012 数据检测图

4.4. 模型架构消融实验

本文在 VOC2012 数据集上进行了详细的模型架构消融实验。实验主要关注主干网络、Neck 和检测头的设计，通过比较不同配置下的模型参数量、mAP@0.5 和 mAP@0.95 指标，如表 2 所示。

Table 2. System resulting data of standard experiment
表 2. 标准试验系统结果数据

Model					Param (M)	mAP@0.5/%	mAP@0.95/%
EMS-Model	SDF-Module	Neck	Anchor-based	Spike Decoupled Head			
√			√		33	49.7	27.2
√				√	35.4	51.1	29.7
√		√	√		111.2	54.9	31.6
	√	√	√		25.4	54.7	30.1
	√	√		√	**27.8**	**56.6**	**33.4**

Neck 设计：在主干网络基础上加入了空间金字塔池化模块(SPPF)与特征金字塔。加入 Neck 部分后，大大提升了模型的检测性能，但也因此带来了模型参数量上的显著增加。

SDF-Module 设计：该模块设计是以轻量化为主导，将 EMS-Model 替换为 SDF-Module，使得模型在保持检测精度不丢失的基础上，使得模型参数减少到之前的四分之一。

脉冲解耦检测头设计：相比于传统单一检测头存在的类别预测和位置预测合并在一个头中导致的误差影响。本研究在 EMS-Model 和 SDF-Module+neck 的基础上，进一步引入了脉冲解耦检测头。这一设计通过解耦分类和回归任务，显著提升了模型的检测精度。实验结果表明，该模型在 mAP@0.5 和 mAP@0.95 指标上均达到了最佳性能，展示了其在目标检测任务中的优越性。

5. 结语

本研究通过结合 YOLO 的多尺度融合架构与 SNN 模块，有效解决了 EMS-YOLO 在多尺度特征融合和复杂目标检测上的不足。该架构通过引入特征金字塔与空间金字塔的设计理念，并结合 SDF-Module 及脉冲解耦检测头，使得模型在 $T=1$ 和 $T=4$ 时，mAP@0.5 分别提高了 6.9%和 4%。研究结果表明，该模型在性能上接近同等 ANN 架构，并在功耗方面表现出显著优势。实验结果验证了这种新型 SNN 架构能够在保持高性能的同时，显著降低计算资源消耗，为 SNN 在实时视觉应用中的广泛应用奠定了坚实的基础。

基金项目

新一代人工智能国家科技重大专项(2021ZD0109805)和湖南省教育厅优秀青年项目(NO. 23B0569)。

参考文献

[1] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al*. (2016) SSD: Single Shot Multibox Detector. In: *Lecture Notes in Computer Science*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[2] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[3] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 6517-6525. https://doi.org/10.1109/cvpr.2017.690

[4] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., et al. (2022) Swin Transformer V2: Scaling up Capacity and Resolution.

2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 11999-12009. https://doi.org/10.1109/cvpr52688.2022.01170

[5] Furber, S.B., Galluppi, F., Temple, S. and Plana, L.A. (2014) The Spinnaker Project. *Proceedings of the IEEE*, **102**, 652-665. https://doi.org/10.1109/jproc.2014.2304638

[6] Benjamin, B.V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A.R., Bussat, J., *et al.* (2014) Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations. *Proceedings of the IEEE*, **102**, 699-716. https://doi.org/10.1109/jproc.2014.2313565

[7] Shen, J., Ma, D., Gu, Z., Zhang, M., Zhu, X., Xu, X., *et al.* (2015) Darwin: A Neuromorphic Hardware Co-Processor Based on Spiking Neural Networks. *Science China Information Sciences*, **59**, 1-5. https://doi.org/10.1007/s11432-015-5511-7

[8] Maass, W. (1997) Networks of Spiking Neurons: The Third Generation of Neural Network Models. *Neural Networks*, **10**, 1659-1671. https://doi.org/10.1016/s0893-6080(97)00011-7

[9] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representations by Back-Propagating Errors. *Nature*, **323**, 533-536. https://doi.org/10.1038/323533a0

[10] Kim, S., Park, S., Na, B. and Yoon, S. (2020) Spiking-Yolo: Spiking Neural Network for Energy-Efficient Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 11270-11277. https://doi.org/10.1609/aaai.v34i07.6787

[11] Li, Y., He, X., Dong, Y.T., Kong, Q.Q. and Zeng, Y. (2022) Spike Calibration: Fast and Accurate Conversion of Spiking Neural Network for Object Detection and Segmentation.

[12] Hu, Y.F., Deng, L., Wu, Y.J., Yao, M. and Li, G.Q. (2021) Advancing Spiking Neural Networks towards Deep Residual Learning.

[13] Fang, W., Yu, Z.F., Chen, Y.Q., Huang, T.J., *et al.* (2021) Deep Residual Learning in Spiking Neural Networks. *Advances in Neural Information Processing Systems*, **34**, 21056-21069.

[14] Su, Q., Chou, Y., Hu, Y., Li, J., Mei, S., Zhang, Z., *et al.* (2023) Deep Directly-Trained Spiking Neural Networks for Object Detection. 2023 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Paris, 1-6 October 2023, 6532-6542. https://doi.org/10.1109/iccv51070.2023.00603

[15] Yao, M. (2024) Spike-Driven Transformer V2: Meta Spiking Neural Network Architecture Inspiring the Design of Next-generation Neuromorphic Chips.

[16] Dayan, P. and Abbott, L. (2001) Computational Neuroscience: Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, 162-166.f

[17] Brunel, N. and Latham, P.E. (2003) Firing Rate of the Noisy Quadratic Integrate-and-Fire Neuron. *Neural Computation*, **15**, 2281-2306. https://doi.org/10.1162/089976603322362365

[18] Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C. and Brunel, N. (2003) How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs. *The Journal of Neuroscience*, **23**, 11628-11640. https://doi.org/10.1523/jneurosci.23-37-11628.2003

[19] Gerstner, W. and Kistler, W.M. (2002) Spiking Neuron Models. Cambridge University Press. https://doi.org/10.1017/cbo9780511815706

[20] Abbott, L.F. (1999) Lapicque's Introduction of the Integrate-and-Fire Model Neuron (1907). *Brain [Research Bulletin*, **50**, 303-304. https://doi.org/10.1016/s0361-9230(99)00161-6

[21] Burkitt, A.N. (2006) A Review of the Integrate-And-Fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics*, **95**, 1-19. https://doi.org/10.1007/s00422-006-0068-6

[22] Wu, Z.F., Shen, C.H. and Van Den Hengel, A. (2016) Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. *Pattern Recognition*, **90**, 119-133.

[23] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/cvpr.2017.195

[24] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/cvpr.2018.00474

[25] Zhou, D., Hou, Q., Chen, Y., Feng, J. and Yan, S. (2020) Rethinking Bottleneck Structure for Efficient Mobile Network Design. In: *Lecture Notes in Computer Science*, Springer, 680-697. https://doi.org/10.1007/978-3-030-58580-8_40

[26] Wang, A. (2024) YOLOv10: Real-Time End-to-End Object Detection.

[27] Zhu, X., Lyu, S., Wang, X. and Zhao, Q. (2021) Tph-Yolov5: Improved Yolov5 Based on Transformer Prediction Head

for Object Detection on Drone-Captured Scenarios. 2021 *IEEE/CVF International Conference on Computer Vision Workshops* (*ICCVW*), Montreal, 11-17 October 2021, 2778-2788. https://doi.org/10.1109/iccvw54120.2021.00312

[28] Sengupta, A., Ye, Y., Wang, R., Liu, C. and Roy, K. (2019) Going Deeper in Spiking Neural Networks: VGG and Residual Architectures. *Frontiers in Neuroscience*, **13**, Article 95. https://doi.org/10.3389/fnins.2019.00095

[29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., *et al*. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **2017**, 5998-6008.

[30] Ali, M.H. (2023) Advanced Efficient Strategy for Detection of Dark Objects Based on Spiking Network with Multi-Box Detection.

[31] Horowitz, M. (2014) Computing's Energy Problem (and What We Can Do about It). 2014 *IEEE International Solid-State Circuits Conference Digest of Technical Papers* (*ISSCC*), San Francisco, 9-13 February 2014, 10-14.

[32] Li, Y., Guo, Y., Zhang, S., *et al*. (2021) Differentiable Spike: Rethinking Gradient-Descent for Training Spiking Neural Networks. https://proceedings.neurips.cc/paper/2021/file/c4ca4238a0b923820dcc509a6f75849b-Paper.pdf

[33] Kim, Y., Chough, J. and Panda, P. (2022) Beyond Classification: Directly Training Spiking Neural Networks for Semantic Segmentation. *Neuromorphic Computing and Engineering*, **2**, Article 044015. https://doi.org/10.1088/2634-4386/ac9b86