文章编号:1671-4598(2026)03-0171-06 DOI:10.16526/j.cnki.11-4762/tp.2026.03.021 中图分类号:TP391.41 文献标识码:A

基于 SpikeYOLO 改进的视频目标检测算法

张昭然
(中国电子科技集团公司 第五十四研究所,石家庄 050081)

摘要:脉冲神经网络因其突出的生物可解释性和能效,在对能量消耗有极端限制的应用领域正在受到越来越多的关注,但目前已实现算法普遍存在识别能力相对传统算法较弱、未充分利用输入数据的时间相关性等问题;为此提出了一种基于 SpikeYOLO 的改进算法,该算法引入多尺度组扩张卷积模块与残差时移模块,通过将过去相邻帧的部分通道移入本帧并进行特征融合,在增加少量参数的情况下改善了算法的信息提取和时间相关性的利用能力;通过引入多尺度组扩张卷积和残差结构,使算法能够有效融合调整后的特征图,以及减小非本帧特征引入导致的算法表达能力减弱和训练不稳定;实验结果表明,该算法在选取人与车两种目标的 COCO2017 静态数据集上的 $mAP50$ 达到 72.0%,在类似的基于 KITTI 和 MOT15 数据集制作的动态数据集上的 $mAP50$ 也达到 80.1%,相比基本算法分别提升了 1.7%和 1%,而算法参数量仅增加 2%,有利于提升脉冲神经网络在视频相关目标识别任务上的应用前景。

关键词:关键词:脉冲神经网络;时移模块;视频目标识别;多尺度组扩张卷积;SpikeYOLO

An Improved Video Object Detection Method Based on SpikeYOLO

ZHANG Zhaoran
(The 54th Institute of China Electronic Technology Corporation, Shijiazhuang 050081, China)

Abstract: Spiking Neural Networks (SNNs) now attract significant attention in application fields with extreme energy constrains due to their exceptional bio-interpretability and energy efficiency. However, compared to conventional recognition algorithms, existing SNN-based algorithms often exhibit limited recognition capabilities and neglect the temporal correlations in input data. To address these issues, this paper proposes an enhanced algorithm based on SpikeYOLO, which introduces a multi-scale dilated group convolution (MSGDC) module and a residual temporal shift (TS) module. By shifting partial channels from adjacent previous frames to the current frame for feature fusion, the algorithm improves information extraction and temporal correlation utilization with only a minimal increase in parameters. Additionally, the MSGDC and residual architecture effectively integrate refined feature maps while mitigating performance degradation and training the instability caused by non-local feature incorporation. Experimental results show that the proposed method achieves the $mAP50$ of 72.0% on a random static dataset constructed from COCO 2017 and the $mAP50$ of 80.1% on temporal-related dataset derived from KITTI and MOT15, demonstrating an improvement of 1.7% and 1.0% over the baseline algorithm, respectively, with only 2% increase in parameter scale. The research can be beneficial for enhancing the application of SNNs in video-related object recognition tasks.

Keywords: SNN; temporal shift module; video target recognition; MSGDC; SpikeYOLO

0 引言

近年来随着“云端协同”等概念的提出,基于边缘设备的目标检测算法成为人们关注的热点之一。边缘设备通常存在能耗和算力限制,直接移植现有的目标识别算法不能达成最优的使用效果,算法性能仍有很大的优化空间,此外,这些应用所生成的图像数据大多为时间连续数据,存在着一定的时间相关性。脉冲神经网络(SNN, spiking neural network)由于其优异的能耗表现,正在受到广泛的关注和研究[1-2]。其通过模仿生物神经元的行为模式,仅在神经元膜电位超过阈值时发放脉冲,具有相对的能耗优势[3]。

目前基于随机静态图像的深度脉冲神经网络已有一些研究,例如文献[4]首次提出了具备深度脉冲神经

收稿日期:2025-12-30; **修回日期**:2026-01-27。

作者简介:张昭然(2000-),男,硕士研究生。

引用格式:张昭然.基于 SpikeYOLO 改进的视频目标检测算法[J].计算机测量与控制,2026,34(3):171-176,185.

网络架构的目标检测算法 Spiking-YOLO。该算法基于 TinyYOLO[5]，优化了深度神经网络下脉冲发放率低下和计算低效的问题。文献［6］提出了首个通过直接训练方法构建的深度脉冲神经网络目标识别算法 EMS-YOLO。文献［7］则提出了一种利用当前较为流行的 Transformer 架构实现目标识别的算法。文献［8］所提出的改进算法 SpikeYOLO 是目前深度脉冲神经网络算法中识别性能较为优秀的，其 13.2 M 模型在 COCO2017 上实现了 59.2% 的 $mAP50$，并在以上提出的各种算法中在相同情况下有最低的能量消耗。

尽管如此，以上所述算法仅利用了数据帧内的空域信息，在视频目标识别的情境中缺乏对输入信息帧间相关性的利用。在传统神经网络目标识别领域，为了提升算法对输入数据时间相关性的提取能力，引入 3D CNN 是最直接的方法[9]。3D CNN 通过给卷积操作增加时间维度上的卷积，使得其能够实现时间维度上的特征提取。但当卷积核深度增加，其对应的卷积计算量也随之增加，不利于资源受限条件下的应用。为了解决这一问题，文献［10］提出了时移（TS，temporal shift）模块。不同于 3D CNN 的将参数在时间维上拓展的思想，时移模块则是将数据在时间维上拓展，能够在不增加计算量的情况下以存储和数据吞吐量为代价实现和 3D CNN 近似的效果。文献［11］进一步研究了时移操作在视频动作识别任务上的效果，结果表明时移操作能有效提取时间相关性并增强算法对时间相关数据的理解，且能够同时提升静态图像的识别能力。

借鉴这一思想，文献［12］提出了首个在深度脉冲神经网络上实现引入了时移模块的图像处理算法，能够在保留反向梯度流动的情况下有效融合不同帧的特征。但其方法主要应用在图像分类任务，算法实践基于 Resnet 算法改进，其改进方法对视频图像识别任务和更加复杂结构算法的效果仍有待探索。

为了解决以上提出的脉冲神经网络在视频目标识别任务上的问题，本文提出一种基于改进 SpikeYOLO 算法的视频目标识别算法，通过向算法结构中引入 TS 模块实现图像提取特征的跨帧传递。为了融合时移模块生成的混合特征图数据，改进算法在时移操作后加入了多尺度组扩张卷积（MSGDC，multiScale group dilated convolution)[13]，在尽可能控制计算量的情况下融合不同时间的特征，并引入了残差连接以保持算法对原帧数据特征的提取能力，从而减小了跨帧特征引入导致的算法表达能力减弱和训练不稳定，提升了 SpikeYOLO 算法的识别性能。

1 系统结构及原理架构网络

SpikeYOLO 是近年来表现最为优秀的深度脉冲神经网络目标识别算法之一，其在能耗和识别能力等方面都达到了目前已有工作的前沿。该算法主要的创新点在于优化了脉冲神经元在复杂深度网络上的部署架构。由于直接按人工神经网络架构实现脉冲神经网络容易导致深层神经元放电率降低的脉冲退化现象，算法的识别能力严重降低。该算法通过将 YOLOv8[14] 的宏观算法架构设计和 Meta-SpikeFormer 的简化的、利于脉冲神经网络实现的微观架构设计进行了结合，并进行了一系列优化，使得其目标识别能力相比先前的类似研究有了很大的进步。

SpikeYOLO 算法的宏观结构和 YOLOv8 基本保持一致，包括主干网络（Backbone），颈部（Neck）和检测头（Head）三个部分，具体结构如图 1 所示。YOLOv8 是一种经典的单帧目标检测算法，其使用了无锚点预测头，将图像划分为众多网格，每个网格互相独立地预测其中的目标，提升了准确性和精度。其中，其主干网络使用了优化后的 CSPDarknet53[15] 架构，能够有效提取多层次目标特征。颈部网络为路径聚合＋特征金字塔结构，其有效结合了浅层位置信息和深层语义信息，使其能够适应各尺寸和形状的目标。检测头使用了无锚定设计，不依赖于锚框预测目标区域，并分别使用两个分支分别预测目标区域和目标类别。算法的可训练参数主要集中在骨干部分，包含两种特征提取模块和下采样模块。两种特征提取模块的主要区别在于其中的卷积操作是采用普通卷积还是重参数卷积。重参数卷积模块被应用到高层特征提取中以控制算法计算开销，增强对高层特征的相对提取能力，底层特征提取使用普通卷积模块。

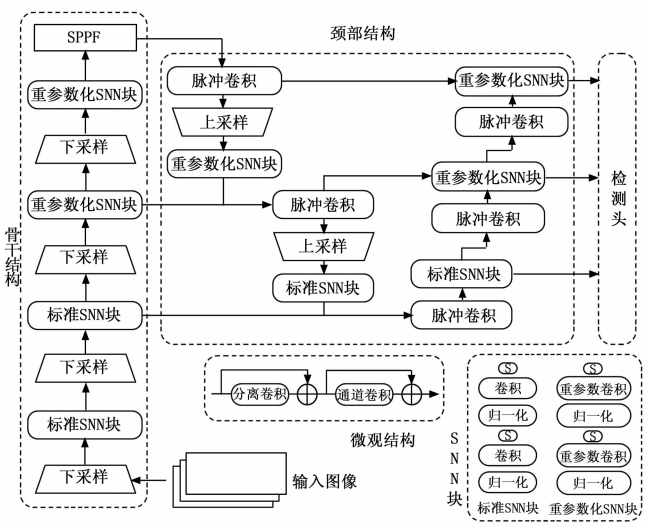

图 1 SpikeYOLO 结构图

在微观结构方面，由于直接利用原本 YOLOv8 的 C2f 模块结构并对其进行脉冲化改造经实验检验发现其

存在深层结构脉冲发放率下降、脉冲退化的问题。为此，SpikeYOLO 算法设计了基于 Meta-Spikeformer 微观结构的 SNN 块来替代原有的 C2f 模块。新的 SNN 块内部包括串联残差结构的分离卷积和通道卷积，模块结构相对简单，能够缓解深度网络中的脉冲退化现象。SNN 块有两种模块设计，用于低层特征提取的 SNN 块使用常规卷积，而用于高层特征提取的则使用了核大小为 3×3 的重参数化卷积，兼顾了提取性能和计算复杂度。

SpikeYOLO 的另一个创新点在于使用了新的脉冲神经元模型，优化了算法在训练阶段的精度。基本的脉冲神经元模型，常见的如泄露积分－发放（LIF，leaky integrate and fire）模型[16]，其推理和训练时都仅输出 0 和 1 两种值，其存在较大的量化误差。SpikeYOLO 算法提出了整数泄露积分发放模型（I-LIF），即在训练过程中，神经元允许表达有限的几个整数值，相比较原本仅能输出 0、1 两种数据的脉冲神经元极大地减少了训练时算法的误差。相对应地，其推理时仍然只输出 0 和 1，通过输出 1 的数量来表示其整数值，保持了其推理时的脉冲神经元特性。通过减小计算误差，I-LIF 神经元能有效增强深度神经网络的表达能力。

2　基于改进 SpikeYOLO 的视频目标识别算法

视频目标识别相对于传统的目标识别具有特殊的数据特点，包括有实时性要求；帧间关系是单向，无法提前获取未获取的帧；数据帧间相关性强、帧间存在一定信息等特点。对于以上情境，改进算法设计了一个插入式的改进模块，引入时移（TS）操作，通过将过去帧的部分特征传递到本帧实现帧间目标特征的提取和利用，并使用残差结构保证算法仍能使用反向传播进行训练，避免特征图拼接造成的反向传播断裂问题；引入多尺度组扩张卷积（MSGDC），将输入通道分组分别进行扩张卷积从而减小卷积核大小，在卷积计算的参数量和计算量增长较少的情况下有效融合时移模块拼接出的特征图，提高了改进算法的目标识别性能。

2.1　时移模块

时移操作是一种低成本的提升视频目标识别算法性能的手段，它通过将特征跨帧地在时间维度上融合，充分利用帧间图像的时间相关性，加强算法的时序建模能力。相关实验数据证明，该模块的引入对静态目标的识别也有一定的加强能力。

时移操作源于对 3D CNN 参数量简化的思考。对一个简化的情况，设想一个卷积核大小为 3 的卷积操作：

$$Y_i = W_1 X_{i-1} + W_2 X_i + W_3 X_{i+1} \tag{1}$$

其中：W_i 为卷积权重，X_i 为时间 i 时待计算的数据。对卷积计算过程进行分析，每一个时间点上的计算实际是先将过去和未来的两部分数据移入当前数据，之后对其进行加权乘法。因此 3D CNN 卷积计算存在一种工程上的简化方法，即先将数据进行移动并放入本帧，之后在本帧使用常规 2D CNN 卷积进行进一步特征提取。已有的工作表明，只传递某时间点其中的一部分特征，算法的识别性能会有明显提高，且算法的计算延时和数据开销等方面将会有明显优化。这种时移操作通过将卷积的一个维在帧间维度上展开，减少了卷积计算本身的维度，在算法中以几乎不增加参数量的方式优化了目标识别和时序建模能力。

从先前的算法分析中可知，卷积操作既需要过去时间的数据也需要未来时间的数据。但对于视频目标识别任务，不可能获取未出现的数据。因此，不同于标准时移模块的双向数据移动方法，TS-SpikeYOLO 算法的时移模块被设计为单向的，即将过去的一部分通道数据移入到本帧。模块初始输入时，由于没有可以移入当前向量的数据，移入的数据将会补零，当识别最后一帧数据时，部分特征会被直接丢弃。具体操作示意如图 2 所示。

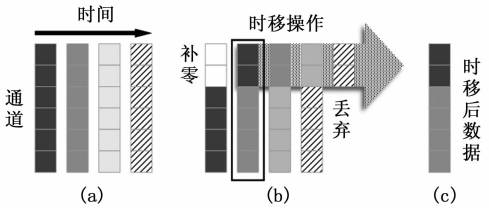

图 2　时移操作示意图

假设算法的数据向量为 $\boldsymbol{X} \in R^{T\times B\times C\times H\times W}$，其中，$T$ 为时间步，即为算法的时间步数，在研究中均设为 1，B 为训练输入图像的批大小，C 为通道数，H、W 为输入数据的高宽。将 C 个输入通道分为 C_f 组，则每组包含的通道数 C_g 为：

$$C_g = C/C_f \tag{2}$$

在按比例移动后，输出向量为 $\boldsymbol{Y} \in R^{T\times B\times C\times H\times W}$，其中 C 为如下描述的状态：

$$C = [C_{i-1}^{1}, \cdots, C_{i-1}^{C_s\times C_f\times\beta}, C_i^{(C_s\times C_f\times\beta)+1}, \cdots, C_i^{c}] \tag{3}$$

其中：β 为移动比例，i 为时间。由于时移操作融合了不同帧所属的特征，原本单帧之间的特征关系将被破坏，且计算上其移动数据的部分将从梯度图中断开，无法进行反向传播计算。为了解决该问题，模块引入了残差连接，将原有特征直接引入到输出并按比例将两者进行融合，时移模块将可以保留算法的梯度流动和空间特征提取能力，减轻当前后输入帧时间相关性较弱时引入的特征扰动及其造成的性能下降，并增强算法的学习稳定性。融合比例的经验值一般为 0.1 到 0.5。为了达成可能的最优性能，对该参数分别设为 0.25 和 0.5 的

情况进行了实验，结果证明设置在 0.25 左右性能最好，并设置算法使用 0.25 作为融合比例的固定值。

2.2 多尺度组扩张卷积（MSGDC）

时移操作不是一个分立的模块，直接单独插入时移模块将会导致目标识别能力的严重劣化，需要在时移模块后引入具有特征融合和提取能力的模块，一般为 2D CNN 卷积。相比于目前实现较多的 Resnet 网络结构，SpikeYOLO 的结构更加复杂，其在整体分为三种模块的情况下，三种模块之间还存在各不同特征层次的信息传递。例如，其颈部结构使用了多尺度特征传递和提取，存在较多的旁路路径，这使得残差模块的引入不再像 Resnet 一样简单插入到残差结构的卷积模块前，其配套的卷积模块也需要具有额外的通道融合能力和高层特征的提取能力才能实现有效性能提升。

根据以上思路，引入残差时移操作后，跨帧拼接的特征矢量需要重新进行融合。为了重建数据的空间相关性，改进算法引入多尺度组扩张卷积 MSGDC，以将不同帧的特征进行重新融合并尽可能捕获多尺度特征。相对于传统卷积或拼接，该模块引入了分组卷积、扩张卷积和多尺度思想，能够更好地保留并提取信息，在参数和计算量代价可控的情况下增强其算法的表达能力，提升目标识别精度。

传统卷积的卷积核在输入向量上连续滑动并逐个计算结果，最后相加。此时卷积操作的感受野尺寸和卷积核的尺寸大小相当。目标识别任务需要算法有良好的上下文关联能力，需要尽可能大的感受野。通过在卷积核中插入零元素，扩张卷积可在不增加算法有效参数的情况下拓展卷积操作的感受野，从而能够捕获距离变化更显著的特征。此外，识别目标物体大小也可能会有很大差异，通过并行进行多个不同感受野大小的卷积操作并在输出时进行融合，可以有效提升算法对不同大小目标的识别能力。

分组卷积是一种提升卷积计算效率的有效方法。其最早在 Alexnet[17] 中提出，通过将输入通道划分为多个组，每组分别独立进行卷积操作，并在最后进行通道融合。由于每个卷积核需要输入的计算通道数减少，能够显著降低计算量。单纯地进行分组操作在操作过深或未能将输出特征进行融合的情况下可能导致信息隔离，因此整体结合以上技术，多尺度组扩张卷积将输入通道划分为 3 组，如图 3 所示，每组分别对应一种尺度的扩张卷积，并在最后插入一个 1×1 卷积对输出特征进行融合。

整体上，MSGDC 模块串接到 TS 模块之后，主要作用是将经过时移后的数据进行融合和进一步特征提取。由于时移操作本身是进行跨帧融合，若将其插入的层次太低，当输入时间相关性不高的帧时会导致引入大量噪声，直接影响整个网络的特征提取能力。因此，为了增强网络对多种不同输入数据情况下的适应能力，改进算法的时移操作应被接入到 SPPF 模块后。改进算法宏观模块结构如图 4 所示。

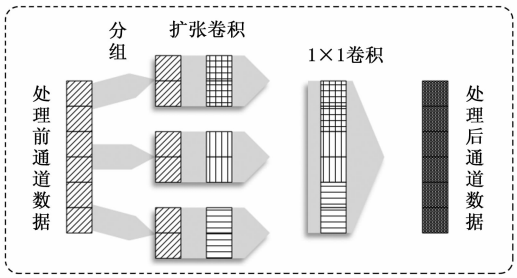

图 3 分组卷积示意图

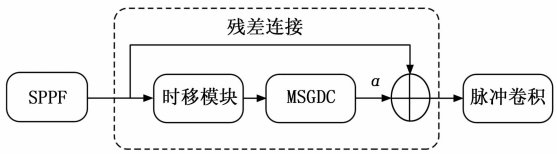

图 4 TS-MSGDC 模块结构示意图

3 实验与结果分析

为了验证改进算法的性能，设计了两种数据集用于算法的训练和验证，并对算法在两种数据集上分别进行了消融实验。通过实验，对算法中融合参数的选择和其性能表现进行了验证和分析，给出了相对最优的选择。

3.1 数据集与方法

算法训练和测试的数据集使用了分别基于 COCO2017[18] 制作的静态图像数据集和基于 KITTI[19] 与 MOT15[20] 两种数据集制作的动态图像数据集进行算法的训练和测试。为了训练算法对人和小汽车这两种目标的识别能力，制作的两种数据集都进行了过滤，只保留了原本数据集结果中人和车两类目标的结果，并剔除了不包含这两类目标任何一个的图像和数据。

COCO2017 数据集是微软 2017 年发布的大型图像数据集，广泛用于计算机视觉任务。经过提取后的静态数据集包括 67 847 张训练图片和 2 869 张测试图像。基于 KITTI 和 MOT15 制作的动态数据集则包括了 6 273 张训练图像和 4 132 张测试图像，具体分布如表 1 所示。基于 COCO2017 制作的数据集主要包含了大量不存在时间相关性的数据，用于训练算法的基本识别能力，而基于 KITTI 与 MOT15 制作的数据集则是由视频片段切分出的图像帧，存在较强的时间相关性，用于训练算法对时间相关数据的提取和适应能力。为了缓解数据集不平衡的问题，训练中使用了相同的数据增强方法，以提升算法的泛化能力。

表 1　数据集分布统计　　个

数据项	静态数据集	动态数据集
训练集 Person 目标数	262 465	30 040
训练集 Car 目标数	43 867	9 349
训练集图像数	67 847	6 273
验证集图像数	2 869	4 132

算法的训练使用多阶段方式进行。为了有效对比算法在不同配置下的性能，算法的训练使用如下策略：首先在静态数据集上训练，总训练轮数设为 200 轮。之后将该基本权重分别应用在基本算法和改进算法上进一步训练，两种训练途径总训练轮数都为 100 轮。训练使用了 4 块 2080Ti 显卡，batch size 均设为 48，lrf 均为 0.01。算法的时间步参数配置均设为 $T=1$ 以适应在线目标检测的预设环境。训练时首先使用 SGD 优化器构建基本权重，在微调阶段使用 AdamW 优化器进行微调。

为了评估算法的性能，使用平均精度（mAP，mean Average Precision）、精确度 P 和召回率 R 作为参数进行评判。其中 mAP 作为直接表现了算法识别能力的参数，是评判算法性能的主要标准。mAP 参数有多种不同使用条件，在此使用 $mAP50$，其是目标检测算法中用于评估性能的核心指标之一，计算的是所有类别交并比阈值为 0.5 这一特定条件下的 AP 平均值，它主要评估算法在定位要求相对宽松（即预测框与真实框有 50%的重叠）时的检测准确性。

召回率（Recall）是评估算法“查全能力”的指标，它衡量的是在所有真实存在的正样本（即真实目标）中，被算法正确检测出来的比例。其计算公式为：

$$R = TP/(TP + FN) \tag{4}$$

其中：TP（True Positive）代表正确检测到的目标数量，FN（False Negative）代表被算法漏检的真实目标数量。召回率越高，说明算法漏检的情况越少，对真实目标的覆盖率越好。

精确率（P）是评估算法“查准能力”的指标，它衡量的是在所有被算法预测为正样本（即算法认为存在目标）的检测结果中，真正是正样本（即预测正确）的比例。其计算公式为：

$$P = TP/(TP + FP) \tag{5}$$

其中：TP（True Positive）含义同上，FP（False Positive）代表误检（即算法预测存在但实际并不存在的目标，或预测类别错误）的数量。精确率越高，说明算法的检测结果越可靠，误检的情况越少。

3.2　静态数据集实验结果

为了观察改进算法在静态数据集上的表现，分别对基线算法、完整改进算法以及分别去除了 TS 模块和 MSGDC 模块的算法进行了测试。各种情况的测试均使用了相同的基本预训练权重，并训练了相同的轮数。测试的结果如表 2 所示。

表 2　静态数据集实验结果　　%

算法	P	R	$mAP50$
SpikeYOLO	77.3	60.8	70.3
SpikeYOLO＋TS	78.0	63.4	71.7
SpikeYOLO＋MSGDC	77.8	62.4	71.2
SpikeYOLO＋TS-MSGDC	78.1	63.5	72.0

由表 2 可知，改进的 SpikeYOLO 算法相较于基线 SpikeYOLO 算法，其 $mAP50$ 提升了 1.7%，P 与 R 也分别提升了 0.8%和 2.7%。对算法实验结果分析，在静态数据集上，改进的算法实现了预期效果。只引入 TS 模块的算法相比只引入了 MSGDC 模块的算法，性能提升更加显著。这可能由于 TS 算法引入的额外扰动延缓了算法的过拟合，以及部分特征可能在输入数据中重复出现，提高了这些重复出现图像的特征提取结果。在训练过程中还对 TS-MSGDC 模块输出数据与原数据的融合比例设置进行了比较，当设为 0.1 与 0.5，相比设为 0.25 时，识别性能下降了 0.21%和 0.14%，因此融合比例设为 0.25 是较为合适的选项。

在静态数据集上训练两种算法的损失函数值如图 5 所示。从图中可以发现改进算法初始的损失值较高，这是由于训练开始时引入了随机初始化的 MSGDC 模块和时移模块后导致了其后模型传递的空域特征的混乱。随着训练的进行，原算法的损失已无法下降，达到最优，而改进算法的损失仍可下降，最终小于原算法的最小值，证明了改进算法相对具有更优的识别能力。

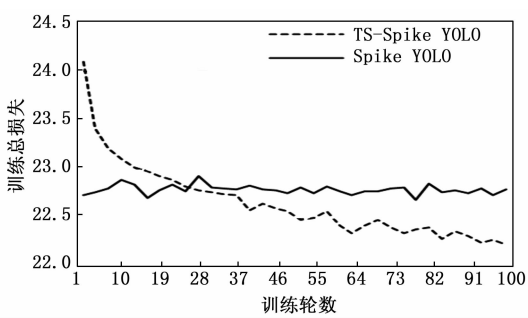

图 5　静态数据训练损失曲线图

3.3　动态数据集实验结果

在动态数据集上，算法的实验流程与静态数据集类似，分别对基线算法、改进算法和分别去除改进模块的算法进行了测试。所有实验都基于相同的静态数据集下训练的基础权重，并训练了相同的轮数。实验结果如表 3 所示。

表 3 动态数据集实验结果 %

算法	*P*	*R*	*mAP*50
SpikeYOLO	86.3	68.6	79.1
SpikeYOLO+TS	86.6	69.2	79.7
SpikeYOLO+MSGDC	86.5	68.7	79.3
SpikeYOLO+TS-MSGDC	87.0	70.5	80.1

由表 3 可知，相比于基本算法，改进算法在动态数据集上也拥有更好的表现。其机理可能为对于受到遮挡或模糊的目标，其部分特征将会被跨帧传递，生成的新特征图可能包含了原图被遮挡或模糊的部分特征，使得算法能够减少因为图像受到污染或破坏导致的特征信息丢失。分析数据集特征，发现构建的动态数据集相较于静态数据集，其数据特征更加恶劣，存在大量的小目标和遮挡目标等，可能是算法表现相对较差的原因。

从训练结果分析，改进算法引入的时延和参数量增加规模都较为有限。改进算法相比原有 13.2 M 算法的参数规模增加了 0.3 M，约 2%，推理时间增加了约 1 ms，在资源消耗相对可控的条件下，改进算法实现了性能的有效提升，能够适应视频目标识别的任务环境。

对两种不同数据集上算法的表现进行分析，改进算法使用的 TS-MSGDC 模块能够在静态、动态两种数据集上实现相对基线算法更优的识别能力。通过在高层将提取特征进行跨帧传递，改进算法能够利用输入数据的时间相关性，减小目标过小、目标模糊等对算法识别造成的影响，提升算法的总体目标识别能力。

4 结束语

本文提出了一种用于视频目标识别的 SpikeYOLO 改进算法，该算法在基本算法的骨干特征提取网络后引入了时移模块对输入数据特征进行了跨帧传递，有效利用了输入数据在时间维度上的信息，提升了算法对目标的识别能力；引入了多尺度组扩张卷积模块，在增加参数量较少且尽可能扩展卷积感受野的情况下实现了对跨帧融合的特征数据进行融合增强。实验表明，本文提出的改进算法在构建的静态数据集和动态数据集上分别达到了 72.0%和 80.1%的 *mAP*50，相比基本算法性能提高了 1.7%和 1%，参数量提升了 2%，提升了算法在目标检测任务上的检测能力。

参考文献：

[1] ZHANG G，FENG L，ZHOU F，et al. Spiking neural networks in intelligent edge computing [J]. Consumer Electronics Magazine，2025，14 (4)：66－75.

[2] PASUPULETI M K. Spiking neural networks for energy-efficient edge intelligence [J]. International Journal of Academic and Industrial Research Innovations，2025，5 (5)：404－414.

[3] FERREIRA P，WANG S，GAO Y，et al. A comparative review of deep and spiking neural networks for edge AI neuromorphic circuits [J]. Frontiers in Neuroscience，2025，19.

[4] KIM S，PARK S，NA B，et al. Spiking-YOLO：spiking neural network for Energy-Efficient object detection [C] //Proceedings of the AAAI Conference on Artificial Intelligence，2020：11270－11277.

[5] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：unified，Real-Time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：779－788.

[6] SU Q，CHOU Y，HU Y，et al. Deep Directly-Trained spiking neural networks for object detection [EB/OL]. ArXiv，2023：2307.11411. DOI：10.48550/arXiv.2307.11411.[2025-12-30]. https://arxiv.org/abs/2307.11411.

[7] YAO M，HU J，HU T，et al. Spike-driven transformer V2：meta spiking neural network architecture inspiring the design of Next-generation neuromorphic chips [C] //International Conference on Representation Learning，2024：52885－52907.

[8] LUO X，YAO M，CHOU Y，et al. Integer-Valued training and Spike-Driven inference spiking neural network for High-Performance and Energy-Efficient object detection [C] //Proceedings of the International Conference on Computer Vision. Lecture Notes in Computer Science，2025：253－272.

[9] TRAN D，BOURDEV L，FERGUS R，et al. Learning spatiotemporal features with 3D convolutional networks [C] //Proceedings of the IEEE International Conference on Computer Vision，2015：510－517.

[10] WU B，WAN A，YUE X，et al. Shift：a zero FLOP，zero parameter alternative to spatial convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：951－960.

[11] LIN J，GAN C，HAN S. TSM：temporal shift module for efficient video understanding [C] //Proceedings of the IEEE International Conference on Computer Vision，2019：718－727.

[12] YU K，ZHANG T Q，XU Q，et al. TS-SNN：temporal shift module for spiking neural networks [EB/OL]. (2025-05-07)[2025-12-30]. https://arxiv.org/abs/2505.04165.

[13] GAO T，ZHANG Z，ZHANG Y，et al. BHViT：binarized hybrid vision transformer [C] //2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，San Diego，CA，2025.

（下转第 185 页）