Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 13, No. 1, 2025, 25-39.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Original paper

Improving the Efficiency of Semantic Segmentation Implemented in Spiking Neural Networks

Elahe yadolahi and Sheis Abolmaali*

Department of Electrical and Computer Engineering Semnan University, Semnan, Iran.

Article Info

Article History:
Received 23 September 2024
Revised 23 November 2024
Accepted 24 January 2025

DOI:10.22044/jadm.2025.15076.2612

Keywords:
Supervised Learning, Image Processing, Semantic Segmentation, Spiking Neural Networks, RMP-Loss.

**Corresponding author: shabolmaali@semnan.ac.ir (Sh. Abolmaali).*

Abstract

Semantic segmentation is a critical task in computer vision, focused on extracting and analyzing detailed visual information. Traditional artificial neural networks (ANNs) have made significant strides in this area, but spiking neural networks (SNNs) are gaining attention for their energy efficiency and biologically inspired time-based processing. However, existing SNN-based methods for semantic segmentation face challenges in achieving high accuracy due to limitations such as quantization errors and suboptimal membrane potential distribution.

This research introduces a novel spiking approach based on Spiking-DeepLab, incorporating a Regularized Membrane Potential Loss (RMP-Loss) to address these challenges. Built upon the DeepLabv3 architecture, the proposed model leverages RMP-Loss to enhance segmentation accuracy by optimizing the membrane potential distribution in SNNs. By optimizing the storage of membrane potentials, where values are stored only at the final time step, the model significantly reduces memory usage and processing time. This enhancement not only improves the computational efficiency but also boosts the accuracy of semantic segmentation, enabling more accurate temporal analysis of network behavior. The proposed model also demonstrates better robustness against noise, maintaining its accuracy under varying levels of Gaussian noise, which is common in real-world scenarios. The proposed approach demonstrates competitive performance on standard datasets, showcasing its potential for energy-efficient image processing applications

1. Introduction

In the field of image processing, understanding images and their contents at a higher level as well as getting quantitative information by semantic segmentation is greatly important in everyday use. This technology is significant in many areas, such as object recognition, pattern detection, the medical sector, robotics, autonomous car technology, and many others. The ability to learn from data and extract semantic features through the recent and advanced deep neural networks has seen the deep neural networks (DNNs) used in the complexity of image processing. These networks, however, have limitations that limit their utilization in some energy-sensitive applications and resource-constraint environments [1].

These incorporate the huge vitality required for preparing and utilizing DNNs, and they ought to utilize tall execution devices. There are numerous ranges where vitality and asset utilization may be an essential concern, such as within frameworks like implanted frameworks, convenient mechanical autonomy, and wearable hardware. This challenge gets more articulate amid real-time picture

handling rates where the pictures must be prepared dynamically [2].

SNNs have become well known as a more efficient, low-power solution to perform tasks that truly implement the working model of the human brain. These networks process information in the form of electrical pulses, and with less accuracy, they are more efficient using much lower power, faster, and can be integrated into low-energy systems. All these benefits make SNNs a potential low-energy solution for tasks in energy-constrained and resource-limited situations compared to DNNs [3]. While SNNs have many benefits, training them with such a complex task as semantic image segmentation faces certain challenges: the dynamic features of SNNs, different training techniques, and reaching sufficient accuracy compared to ANNs.

In this work, we explore different approaches to using SNNs for semantic image segmentation. In this paper, we are going to build upon the architecture of Spiking-DeepLab V3 that was introduced in the paper 'Beyond classification: directly training spiking neural networks for semantic segmentation,' [4] by enhancing it with the RMP-Loss function from the paper 'RMP-Loss: Regularizing Membrane Potential Distribution for Spiking Neural Networks' [5]. The proposed model attempts to improve the performance considering both accuracy and efficiency for the said task of semantic segmentation. Specifically, we compare our results with those from [4], which also addresses the challenge of semantic segmentation using SNNs. Our proposed model shows better accuracy and performance compared to the previous methods in the semantic image segmentation tasks. To verify our model, we used benchmark datasets such as MNIST_CIFAR10 and PASCAL VOC 2012.

In the context of this work, the key contribution is: developing Spiking-DeepLab with RMP-Loss for semantic segmentation of images using SNNs. It builds upon the architecture of Spiking-DeepLab and further uses RMP-Loss to enhance membrane potential in subsequent layers of the network.

The distribution of spikes was also rather dispersed among the layers. The dispersion was identified when analyzing obtained curves; to improve it, RMP-Loss was used. Our experimental results show that our proposed model significantly improves IoU accuracy compared to Spiking-DeepLab and other SNN-based methods for semantic image segmentation.

These results highlight that SNNs can be viewed as a novel, powerful approach to semantic image segmentation that can enhance accuracy and efficiency considerably while solving existing challenges in the domain.

- Through several experiments we performed, it was identified that the Spiking-DeepLab model faces various challenges that affect its overall performance. Quantization errors occur from the discrete conversion of continuous membrane potentials that make precise detection of object boundaries harder.
- By integration of RMP-Loss we reduced these problems by directing the membrane potential distribution to values 0 and 1.
- We also conducted experiments and observed that the execution times were significantly higher than in the previous case due to the new computational complexity introduced by the RMP-Loss function. To address the increased time during network training caused by the complexities of RMP-Loss, the original model stored membrane potentials at multiple time steps for each layer, which led to high computational costs and inefficient resource usage. In the optimized version, membrane potential values are only stored at the final time step for each layer, significantly reducing memory usage and improving the computational efficiency of the model.
- In the numerous experiments we conducted, the results showed that using the proposed model significantly improved the mIoU accuracy. This improvement in accuracy was clearly evident across different datasets, indicating enhanced model performance in the task of semantic image segmentation.

The paper is laid out in the following manner:

Section 2 gives an overview of the related work which encompasses semantic segmentation, spiking neural networks (SNNs), ways of preparing SNNs, and the creation of RMP-Loss.

Section 3 is devoted to the enhancement of semantic segmentation using spiking neural networks. Section 4 describes the proposed methodology, including the representation of inputs, the use of leaky neurons in the integrate-and-fire model, the Spiking-DeepLab architecture, and the training configuration. Section 5 presents the experiments where the reader can find information about the datasets used, the results obtained, and the robustness analysis. Last but not the least, Section 6 gives the conclusion of the paper along with the research implications.

2. Related work

Although there is a more sophisticated neural network architecture, which has brought major development in image processing, SNNs are some of the most promising models for low-power, real-time applications. Many works have been done to improve SNN training for tasks such as semantic segmentation. In this section, we discuss some important methods concerning the semantic segmentation task, SNN construction, training approaches, and the introduction of the RMP-loss function.

2.1. Semantic segmentation

Semantic segmentation is one of the image processing techniques that partitions the image into multiple segments and associates each pixel in the image with a certain label. These semantic labels depict the category of an object or region of space to which a pixel pertains. Semantic segmentation, on the other hand, aims to assign a tag relating to what is being illustrated to each pixel in the image. Transport properties are identified based on the image data and its corresponding description to enable their analysis using computer vision algorithms. [6] For instance, in an image of a street, an algorithm used in semantic segmentation can partition the whole image into human, car, tree, building, sky, and road.

With the development of deep learning, the focus on semantic segmentation has grown in great interest. Deep learning models have garnered attention in this domain due to their several benefits, such as the ability to learn on their own, modeling capacity, capacity to handle vast information, and versatility [7].

CNNs are among the most critical and widely used architectures in deep learning, especially for segmenting images. These networks find applications in several fields like healthcare, automotive, and robotics because they allow the extraction of complicated features from images efficiently. Several architectures have been proposed for this purpose, such as U-Net [8], SegNet [9], and DeepLab [10].

The U-Net model architecture is one of the most remarkable models used mainly in semantic image segmentation, particularly in the medical and biomedical domains [11]. Employing the encoder-decoder architecture, it is specially designed for the acknowledgment and recreation of the characterizing components of an image. In the encoding process, important and complicated elements are extracted from the input image, while in the decoding process, it helps in reconstructing the image and spotting the exact edge of an object. Another significant characteristic of U-Net is the ability to utilize “skip connections” that transmit high-resolution data from the encoder layers to the corresponding decoder layers. These connections assist in maintaining important information wherever the segmentation of data is concerned, which in one way or another enhances the model's accuracy. For this reason, U-Net performs exceptionally well in applications in which high precision and details are acceptable and must be retained [8].

SegNet may be a completely convolutional encoder-decoder design motivated by early thoughts from U-Net and created particularly for semantic picture division errands. This demonstrates the employment of complementary layers to exchange critical data from the encoder to the decoder, making a difference hold fundamental points of interest amid the division preparation. One of the standout highlights of SegNet is its utilization of max-pooling records within the encoder organization. At each max-pooling step, the area of the greatest value in each locale is put away. These records are at that point utilized within the decoder organization to recreate the highlight maps. In other words, these records permit SegNet to protect the exact areas of highlights amid upsampling, hence moving forward division exactness. This approach not only makes a difference in diminishing the number of parameters and computational complexity but also makes strides in proficiency and preparation speed. Because of these characteristics, SegNet is broadly utilized in applications requiring quick and efficient preparation, such as independent frameworks and video surveillance [9].

DeepLab is one of the foremost advanced models for semantic picture division. By utilizing Atrous convolution [12], it grows the responsive field of channels without expanding the number of parameters or computational fetched, driving to more precise pixel forecasts. DeepLab is recognized as one of the top-performing models in segmentation tasks and is connected in areas like independent driving and therapeutic picture analysis [10].

In this paper, we utilize the DeepLab v3 demonstration for semantic division assignments. This adaptation of DeepLab offers critical auxiliary enhancements, especially within the utilization of the Atrous Spatial Pyramid Pooling (ASPP) module [12], giving improved capabilities for handling complex images. DeepLab v3 is particularly planned to utilize Atrous convolution, which permits it to extricate highlights at different scales. This highlight leads to a more precise

identification of protest boundaries and fine, subtle elements. By altering the Atrous rate, this show can viably change the open field of channels, coming about in higher exactness in question discovery in complex pictures. A nitty-gritty view of ASPP and Atrous convolution is given in Figure 1 , which makes a difference way better than getting the structure and usefulness of these modules.

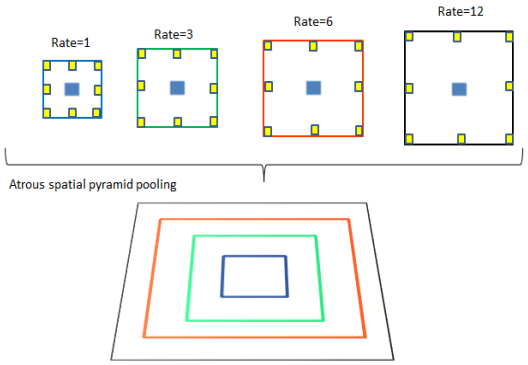

Figure 1. ASPP and Atrous Convolution Overview.

2.2. Spiking neural network

ANNs are complicated machine learning models derived from the formatting of the human brain. They are made of numerous small processing units called neurons, which are connected in such a manner that they can process as well as transmit information. In other words, for ANN-based problems, one can think of the task to be solved as being in the realm of classification, prediction, or control. However, because of the amount of mathematics executed in the network, the energy usage may be high in ANNs [13].

In the 1940s and 1950s, early studies on spiking neural networks (SNNs) began. SNNs have garnered attention due to their numerous advantages, such as low energy consumption, high processing speed, the ability to be implemented in low-power hardware, and adaptability to the human brain [3]. Unlike traditional ANNs, which use fixed numerical values to represent information, SNNs consider temporal dynamics and encode information through the timing and pattern of spikes [14].

SNNs consist of artificial neurons that simulate the neurons of the human brain. These neurons are linked and relay data in short electrical impulses known as spikes [15][16].

Two key mechanisms involved in the learning and functioning of SNNs are:

Spike-Timing-Dependent Plasticity (STDP): This mechanism is similar to how the human brain learns. STDP asserts that the efficiency of synapses, which is the relationship between neurons, alters depending on the comparison of the time of firing of the pre-and post-synapse neurons. Thus, it seems logical that when a post-synaptic spike begins just after the start of a post-synaptic spike, the strength of a synapse increases. On the other hand, if the postsynapse spike occurs in a slightly more delayed fashion than the presynapse spike, the strength of that synapse is weakened. SNNs can learn in STDP, depending on the input data, to update and modify their synapse connections in the form of spikes to obtain patterns from the data [17] [18].

Leaky Integrate-and-Fire (LIF): This model simulates the behavior of actual neurons in SNNs. Thus, the LIF model prescribes the integration of neuron membrane potential over synaptic inputs, ongoing continually. At a certain potential, the neuron propagates an impulse, which is termed a spike, and then returns to the initial state. With the help of LIF models, SNNs can reproduce intricate temporal profiles, which are important for many informational processes [14]. Thus, SNNs can substitute ANNs in many applications because of multiple benefits, including low energy demand, high calculating velocity, and good compatibility with low-power devices. However, there is a need for future work and studies for the learning and optimization of SNNs, unlike ANNs, since SNNs are a lot more complex than ANNs.

2.3 Methods for Prepare Spiking Neural Systems

The coordinated preparation of SNNs is more troublesome compared to ANNs. This trouble emerges because of the brokenness and non-differentiability of the spiking enactment work, which forbids the application of customary strategies of gradient-based preparing.

In the ANN-to-SNN conversion method, first, an ANN is trained for the given task. Then it transfers trained parameters to a similar-structure SNN. This can maintain acceptable accuracy for smaller SNNs, while for larger and deeper networks, accumulated approximation errors lead to reduced accuracy.

Spiking neural networks are designed to run primarily for many time steps to achieve a more accurate output in many applications. This could be at the cost of high computational time and energy consumption in some applications that need fast processing. With increasing numbers of time steps, there is a larger computational complexity entailing further resource requirements. A few new approaches, such as the retraining of the SNN parameters after conversion from ANN, have been considered lately to deal with such issues. These

methods achieve accuracy for converted SNNs similar to that of the original ANNs by re-optimizing the parameters. Another problem of direct training in SNNs is, of course, the nature of the spiking activation: it is discontinuous and non-differentiable, hence classical gradient methods cannot be applied. One remedy is the use of surrogate gradients.

The spiking activation function here is replaced by a continuous differentiable function that allows the computation of gradients during training. Several methods have been developed to design such surrogate functions, which have improved the accuracy and efficiency of the training of SNNs [19]. Surrogate gradients, in combination with other methods like temporal normalization, enable the direct training of more complex SNNs on harder tasks like semantic image segmentation, achieving performances close to ANNs.

In all, ANN-to-SNN conversion and surrogate gradient methods are two promising solutions to overcome the challenges of training SNNs and take full advantage of their benefits in energy-sensitive and resource-constrained applications. Another benefit that comes with using ANN-to-SNN conversion and surrogate gradients is a reduction in energy consumption, increasing the speed of processing, and the ability for implementation in low-power SNN hardware [20].

2.4. RMP-Loss

RMP-Loss is the acronym for 'Regularizing Membrane Potential Loss', which compares to regularizing the conveyance of the layer possibilities in SNNs. This imaginative strategy proposes to play down quantization mistakes from SNNs, which are accepted to have a lower esteem of quantization blunder the closer the layer potential is to or 1. Therefore, RMP-Loss could be a straightforward and productive strategy for moving forward with exact SNNs without expanding its complexity or the number of parameters [5] .

The RMP-loss usefulness includes a few key stages. First, the layer potential for each neuron is calculated, speaking to the voltage of the neuron's film and its part in neuron actuation. Next, the, by and large, conveyance of film possibilities and spike conveyance are computed to decide how numerous neurons and spikes exist at each potential level. Finally, RMP-Loss utilizes misfortuned work that compares the layer potential dissemination with the target spike dispersion.

This work is intended to bring the membrane potential near 0 and 1 to avoid issues of quantization errors. Quantization error in Spiking Neural Networks (SNNs) arises from the quantization of the membrane potential and is usually assumed to equal 0 and 1 for spiking excitatory neurons. One of the key effects is that when the membrane potentials are spread with a wide range, it results in increased noise and instability in the timing of spikes hence arising at quantization error. These errors deteriorate the accuracy of spiking generated and this is unsuitable for tasks like the semantic segmentation where particular attention is paid to features and boundaries detection.

RMP-Loss rectifies this circumstance by making membrane potentials more concentrated in a range around 0 and 1. It directly minimizes quantization errors by directing attention to distribution and improvement of the membrane potentials. Therefore, RMP-Loss lowers the quantization error compared to the previous standard deviation strategy to emerge the membrane potential distribution and reduce its variance enabling the correct spike generation without instabilities. This improvement in spike generation precision helps in feature identification and boundary segmentation, which is crucial for handling complex tasks.

Moreover, depending on the quantization step, the quantization noise decreases; thus, the timed spike synchronicity enhances, giving more correct segmentation. Consequently, the semantic segmentation of the model is improved and this increases the accuracy of the proposed model.

When looking at the conveyance of SPIKEs over distinctive layers, we watched critical scattering. Based on the gotten bends, we recognized this issue and rectified the scattering utilizing RMP-Loss. This adjustment has made a difference, made strides in the layer's potential dissemination, and expanded the model's exactness in different assignments.

Figure 2 ,Figure 3 and Figure 4 shows the membrane potential distribution across three layers of the spike neural network before applying RMP-loss, while Figure 5,Figure 6 and Figure 7 displays the membrane potential distribution after applying RMP-Loss.

- Layer 1:
  Before applying RMP-Loss: Initially, in the first layer, the membrane potential distribution exhibits a high degree of variability. Within this layer, the layer potential dispersion is more scattered, with possibilities crossing a more extensive extent. The top of the dissemination is less concentrated, which can demonstrate higher commotion and lower exactness in

film potential calculations, possibly driving poorer execution in creating neural spikes (Figure 2).
After Applying RMP-Loss: After applying RMP-Loss, the film potential dissemination gets much more concentrated, and the run of variety diminishes. This concentration shows decreased commotion and progressed precision in layer potential calculations, which may reflect way better arrangement execution in this layer (Figure 5).

- Layer 3:
  Before applying RMP-Loss: Within the third layer, the layer potential conveyance is more extensive, with more recognizable scattering. This may propose higher clamor and decreased precision in film potential calculations, likely diminishing the network's forecast exactness (Figure 3).
  After Applying RMP-Loss: After applying RMP-Loss, the layer potential conveyance gets to be more centered, with a diminished range of variety. This enhancement reflects decreased noise and expanded exactness in potential calculations, which may improve the network's execution in this layer (Figure 6).
- Layer 8:
  Before applying RMP-Loss: Within the eighth layer, the layer potential conveyance is more scattered, and the crest of the dissemination is moderately moo. This scattering and moo crest stature shows more commotion and decreased precision in layer potential calculations, which might lead to diminished arrange effectiveness in this layer (Figure 4).
  After Applying RMP-Loss: After applying RMP-Loss, the membrane potential dissemination within the eighth layer gets exceedingly concentrated, and the crest stature significantly increases. This rise in crest tallness shows a tall concentration and negligible commotion. The precision in deciding layer possibilities has significantly made strides, helping in more exact data preparation and an ideal neural spike era in this layer (Figure 7).

The RMP-Loss formula is defined as follows:

$$RMPLoss = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{T}\sum_{t=1}^{T}(p_i - p_{i,\hat{k}})^2\right) \quad (1)$$

In this equation, M speaks to the full number of neurons within the neural organize, and T is the entire number of time steps (or, in other words, the full number of conceivable spikes) for each neuron. The variable p_i signifies the steady-state film potential dispersion for neuron i, while $p_{i,\hat{k}}$ speaks to the layer potential dispersion for neuron i at time step t.
Here, t is used as an index to refer to a specific time step, ranging from 1 to T. Additionally, T is a constant that determines the total number of possible time steps for each neuron. This structure allows us to examine membrane potential changes over time for each neuron and perform a more detailed analysis of neuron behavior in spiking neural networks.
RMP-Loss can be integrated into DeepLab loss functions to enhance accuracy and reduce quantization error. For instance, the overall loss function formula can be expressed as:

$$L_{total} = L_{CE} + L_{RMP} * \lambda(n) \quad (2)$$

Where L_{CE} represents cross-entropy loss, L_{RMP} denotes RMP-Loss, and λ(n) acts as a dynamic balancing coefficient that changes throughout training. This coefficient allows the network to focus more on learning fundamental features in the early stages and, later, on optimizing membrane potential distribution and reducing quantization error.

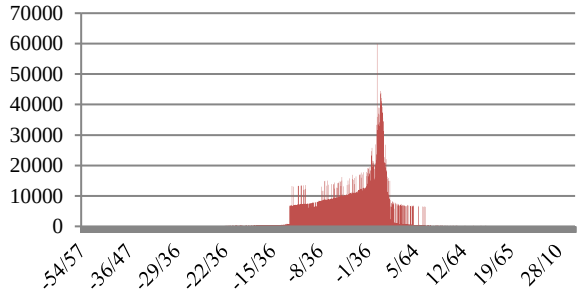

Figure 2. Membrane Potential distribution without RMP-Loss (Layer 1).

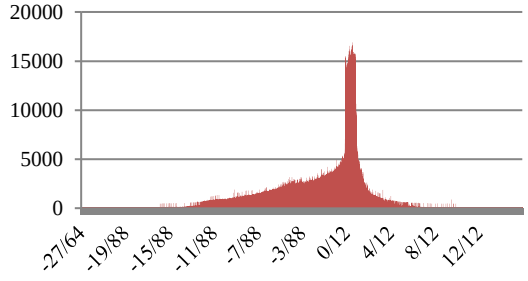

Figure 3. Membrane Potential distribution without RMP-Loss (layer 3).

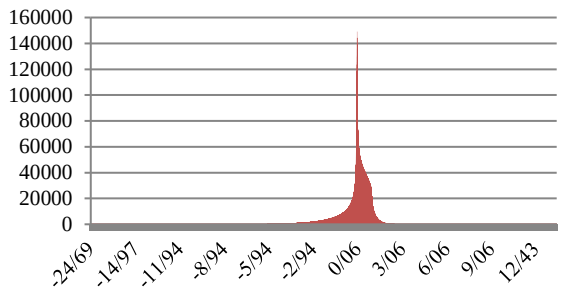

Figure 4. Membrane Potential distribution without RMP-Loss (Layer 8).

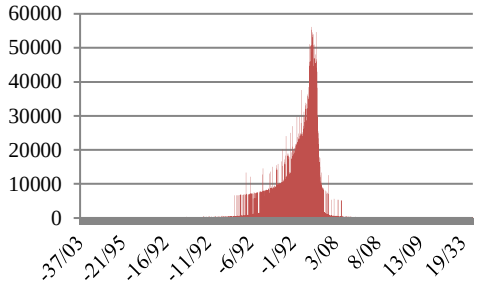

Figure 5. Membrane Potential distribution with RMP-Loss (Layer 1).

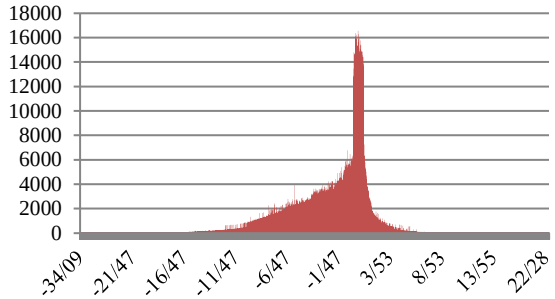

Figure 6. Membrane Potential distribution with RMP-Loss (Layer 3).

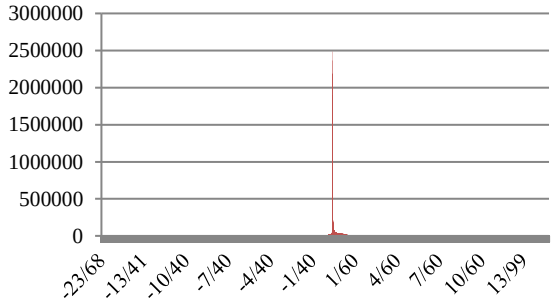

Figure 7. Membrane Potential distribution with RMP-Loss (layer 8).

3. Spiking semantic segmentation

The emerging and sophisticated techniques of rising semantic segmentation in the areas of computer vision and deep learning are based on the use of spiking neural networks (SNNs), intended to provide an even closer representation of the natural structure and operation of the human brain [21]. Instead of using long waveforms, they implement digital short pulses of 0 and 1 to simulate neuronal firing to demarcate images into different zones, each with a different meaning. This comprises feature extraction from an image, spike-based segmentation of the image, and the generation of semantic labels for each part of the image at the end [4].

In this step, the semantic features of the input image are acquired by the processing of the spiking neural networks. Semantic features can represent color, texture, and the shape of objects in an image. Feature extraction is a major stage for preparing the image to be processed further, which improves different stages of processing. Further, image segmentation is done by segmenting the image into several regions. By this time, the network labels each pixel for its appropriate class, whether it is sky, tree, building, or other categories. It also includes segmenting the image into regions, thus enabling the network to make better sense of what each of those regions means. These labels provide the class of each region and can help machine vision systems gain a much better understanding of their surroundings. It holds high suitability for autonomous vehicles, mobile robotics, and virtual reality due to the advantage of its low energy consumption and higher processing capability than that of ANNs.

However, there are a few problems in this field. Spiking neural network-based approaches are less accurate than traditional ones [22]. Furthermore, these networks require more advanced training techniques to improve their performance and accuracy. As a result, additional research is being conducted to optimize and improve the efficiency of spiking neural networks to fully utilize the technology's potential.

4. Methodology

Semantic picture division is one of the foremost imperative assignments within the field of computer vision and has ended up omnipresent over different regions such as mechanical autonomy, medication, expanded reality, etc. Be that as it may, issues like non-differentiability, quantization or data misfortune, and destitute transformation procedures from ANNs to SNNs pose troubles in performing semantic division utilizing Spiking Neural Systems.

In this paper, an unused approach to semantic picture division utilizing SNNs is proposed. This approach utilizes the DeepLab design for actualizing SNNs and utilizes the RMP-LOSS function for preparing the organization. DeepLab may be a kind of deep neural network (DNN)

engineering planned for semantic picture division. On the other hand, RMP-LOSS may be an unused proposed misfortune for the preparation of SNNs to utilize to a degree that will manage the dispersion of neuronal layer possibilities.

The combination of DeepLab and RMP-LOSS makes a difference in overcoming the challenges of utilizing SNNs for semantic picture division and accomplishes progress that comes about in this region. RMP-LOSS directly administers the conveyance of neuronal layer possibilities amid preparation, making a difference in the arrangement to learn fitting spike-like actions that are fundamental for the right working of SNNs in picture division errands. This strategy yields great results in semantic picture division utilizing SNNs and can contribute to the development of made strides strategies in this field.

When the RMP-Loss function was added to the Spiking-DeepLab model, the execution time first increased because of the extra computational overhead, but this processing time increase was offset by an important innovation in the optimization of membrane potential storage: whereas the original model stored membrane potentials at multiple time steps for each layer, which resulted in a high computational cost, our model would only compute the final output, which resulted in inefficient resource utilization and long execution times; in the optimized version, membrane potential values are now only stored at the final time step for each layer, which not only significantly lowers memory consumption but also greatly improves the computational efficiency of the model. Consequently, the overall execution time was significantly decreased, and the precision of picture segmentation and the accuracy of the temporal analysis of network behavior were further enhanced.

4.1. Leaky Neurons in the Integrate-and-Fire Model

In this paper, we use the Leaky Integrate-and-Fire (LIF) neuron for spiking neural systems, a simple but effective demonstration inspired by genuine neural cells that analyze each neuron's layer potential to determine its activation or deactivation and the era of spikes [23]. The LIF show offers a few points of interest due to its scientific effortlessness and its capacity to reenact complex worldly elements of neurons. These focal points incorporate brain recreation for examining brain work and instruments of learning and memory, as well as data preparation in areas such as discourse acknowledgment, picture preparation, and mechanical autonomy.

When the input flag I(t) is connected to the LIF neuron, the film potential changes. Since voltage and current are persistent factors, the differential condition is communicated as follows:

$$\tau m = \frac{dV_m}{dt} = -V_m + RI(t) \tag{3}$$

$$v_i^t = \lambda * v_i^{t-1} + \frac{1}{2}\sum_j w_{ij} * o_j^t \tag{4}$$

Symbols λ and w_{ij} represent the spillage calculation and the weight association between the presynaptic neuron j and the postsynaptic neuron i, individually. When the neuron actuates, fast voltage changes happen over the cell layer, driving to the era of an electrical beat. The number and recurrence of spikes contain critical data. When u_i^t surpasses the terminating limit (θ), neuron i produces a spike yield (o_i^t):

$$o_i^t = \begin{cases} 1 & if \quad v_i^t > \theta \\ 0 & else \end{cases} \tag{5}$$

During a specific refractory period following a spike, the neuron is unable to produce a new spike, which prevents random and explosive firing of the neuron.

4.2. Representation of Inputs

In this consideration, we make utilize of three distinctive datasets, MNIST_CIFAR10 and PASCAL VOC 2012 and ADE20K, to prepare and gather spiking neural systems (SNNs) for semantic picture division. Since SNNs work based on spikes, converting inactive pictures into spike trains is basic. To attain this, we utilize the rate-coding strategy.

In rate coding, each pixel within the picture is spoken to by a particular escalated run, including the least and greatest conceivable values for that pixel. At each minute, the genuine concentration of each pixel is compared to an arbitrarily produced number inside that run. In case the arbitrary number is more noteworthy than the pixel concentrated, a spike is produced; something else, no spike is delivered.

Over time, as the produced spikes gather at each minute, the spike design steadily reflects the key highlights of the picture. In this way, Spiking Neural Systems (SNNs) can prepare the highlights inside the pictures as a stream of spikes and perform the assignment of semantic division.

4.3. Spiking-DeepLab

Not at all like image classification errands, division systems classify each pixel based on the two-dimensional input picture. Hence, division designs

ought to keep up tall spatial determination for including maps at the conclusion of the organization. Also, having a huge responsive field makes a difference the systems reveal connections between objects in a scene, driving to progressed execution. For this reason, DeepLab presents a profound neural arrange engineering for semantic picture division and proposes an Atrous Convolution to meet this prerequisite Encourage, these layers are taken after by a three-layer classifier, which relegates each pixel in a picture to a name from predefined classes concurring to a learned design. In other words, this layer characterizes to which category or protest a pixel has a place. The utilization of expanded convolutional layers permits covering expansive responsive areas without expanding parameters [10].

Unlike standard convolution, Atrous convolution carves out information from larger areas of the image without a reduction in spatial resolution and hence improves semantic segmentation accuracy, especially for high-resolution images. The key factor during Atrous Convolution is a dilation rate that determines how many pixels need to be inserted between each convolution filter element. While increasing the dilation rate, the filter effectively expands and can cover a greater area of interest in the image.

The next step is to place several layers of convolution in parallel, each with different dilation rates. Objects of different scales can then be captured because the dilation rate of the convolution layer varies. The output then passes through a pixel-wise classifier. The resulting tensor has channels equal to the number of classes [12].

Now, the output layer is retrieved by performing an average pooling operator on this tensor. This serves to fuse information by averaging out the features extracted from different layers; also, the computational load is drastically reduced. Therefore, every pixel will give a numerical value representing the class of that pixel. It enhances the accuracy of semantic segmentation and object boundary detection since useful information from different layers is effectively combined into a final layer. This enables the network to highlight more complex features, raising the bar for overall model performance (Figure 8).

In this paper, we utilize three datasets, MNIST_CIFAR10 and PASCAL VOC 2012 and ADE20K, to prepare and gather spiking neural systems (SNNs) for semantic picture division. Since SNNs work based on spikes, changing inactive pictures into spike trains is basic. To attain this, we utilize the rate-coding strategy.

In rate coding, each pixel within the picture is spoken to by a particular escalated run, including the least and greatest conceivable values for that pixel. At each minute, the genuine concentration of each pixel is compared to an arbitrarily produced number inside that run. In case the arbitrary number is more noteworthy than the pixel concentrated, a spike is produced; something else, no spike is delivered.

Over time, as the produced spikes gather at each minute, the spike design steadily reflects the key highlights of the picture. In this way, Spiking Neural Systems (SNNs) can prepare the highlights inside the pictures as a stream of spikes and perform the assignment of semantic division [4].

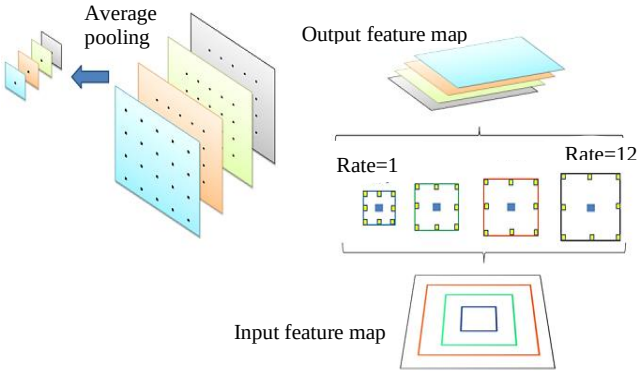

Figure 8. ASPP model with Average Pooling.

4.4 Architecture and Training Configuration of the Spiking-DeepLab Model

In this work, the Spiking-DeepLab model is being used to perform semantic image segmentation. It is inspired by an eight-layer base network. This particular number of layers is used to alleviate some of the challenges with deep neural networks: the mismatch between the real and surrogate gradients. One of the key benefits of spiking neural networks is their vitality proficiency compared to normal neural systems, which is particularly basic for AI applications on edge gadgets.

The essential structure of the Spiking-DeepLab demonstrates coordinating two expanded convolutional layers. These layers are then taken after by a three-layer classifier, which allows each pixel in a picture to a name from predefined classes agreeing to a learned design. In other words, this layer characterizes to which category or question a pixel has a place. The utilization of expanded convolutional layers permits covering expansive responsive areas without expanding parameters.

In the training of the Spiking-DeepLab model, together with the loss concerning the main task, the RMP-Loss function is also involved.

In the underlying variant of the model, film possibilities were not put away during the different time ventures for each layer, and just the last result of the organization was determined as the model's outcome. This approach commonly centered exclusively on the last result of the organization and the general picture grouping results. In any case, in the refreshed rendition, the film likely qualities at the last timestep for each layer are put away. This change permits us to notice further elements of the information handling process at each layer and direct more exact investigations of the organization's conduct over the long haul.

In the first version of the model, the membrane potentials of the neurons were not saved during the various time steps in each of the layers, but only the network output was computed as the model's output. This approach was mainly concerned with only the final output of the network as well as the overall results of the image semantic segmentation. However, in the enhanced version, only the membrane potential values at the final timestep for each layer are stored.

This change enables us to study deeper features of the data processing at each layer and perform more accurate analyses of the network dynamics over the time.

It is especially beneficial for the model change in applications of semantic segmentation and its combination with RMP-Loss. Since RMP-Loss is used to minimize the distribution of membrane potentials to minimize quantization errors and enhance the model's ability to detect various areas of an image, storing the membrane potential at the final time-step helps in providing better information on how the membrane potentials correspond to the different aspects of an image at the last stage of processing. This improvement not only helps in the improved analysis and understanding of the behavior of the network but can also result in improved efficiency in the process of correct classification of the image and reducing the errors in the segments.

Additionally, this enhancement has led to a reduction in execution time, as storing only the membrane potentials at the final timestep minimizes redundant computations and optimizes memory usage, further enhancing the model's overall performance and applicability.

This loss function, by its definition, minimizes the conversion error of membrane potentials into a discrete format. During training, RMP-Loss would shift the distribution of membrane potentials towards binary values of 0 and 1.

With a learning rate of 3×10^{-3}, a clump measure of 16, a learning rate diminishment by a figure of 10 at 50% of the overall number of ages, a add up to 60 pages, 20-time steps, a spillage figure of 0.99, and a layer potential limit of 1.0.

The proposed Spiking-DeepLab model uses an encoder and decoder structure. The encoder is composed of a few early layers of the base network, which are convolutional and pooling layers for input feature extraction, while the decoder consists of the last few layers of dilated convolutional and classifier layers. The decoder then needs to generate the final output by predicting a semantic label for each pixel in an image based on extracted features.

5. Experiments

In this section, we will introduce the performance of the proposed methods using two publicly available datasets: PASCAL VOC 2012 and MNIST_CIFAR10. Both have their specific features and challenges w.r.t. semantic segmentation. The PASCAL VOC 2012 dataset is a challenging dataset that, within computer vision and machine learning, provides real-world images with accurate labels. Besides, the MNIST_CIFAR10 datasets have been selected to form the basis for assessing our algorithms because they are some of the most diverse and widespread in training deep learning models. Further details about each of these datasets are given in the following sections.

5.1. Dataset

In this paper, we assess our strategies on two challenging and assorted semantic division datasets: PASCAL VOC 2012 and MNIST_CIFAR10 and ADE20K.

PASCAL VOC 2012 [24] has been very prevalent and broadly utilized as a standard dataset within the field of computer vision as well as machine learning, particularly for protest location and division in pictures. This dataset incorporates 20 protest classes of real-world pictures, counting individuals, cars, creatures, and other objects, etc. Each picture is depicted, and the portions containing each protest are checked as well (in Figure 9). PASCAL VOC 2012 incorporates 22,951 pictures separated into three primary subsets: case, preparing, approval, and testing sets.

The MNIST [25] is one of the most often used datasets in machine learning and computer vision. It consists of 70,000 photos, with each image representing a handwritten numeral spanning from 0 to 9. All of the images are black and white and are 28x28 pixels. These photos are purposefully created to feature noise and variances in

handwriting style, allowing models to generalize more well in real-world digit identification tasks.

The CIFAR-10 [26] dataset could be a widely used picture dataset in profound learning and computer vision. It comprises a little picture of distinctive colors with a picture resolution of 32x32 pixels, and the pictures are assembled into 10 classes. Some of these classes incorporate planes, cars, winged creatures, cats, deer, pooches, frogs, horses, transport, and trucks. The pictures show tall differing qualities in lighting conditions, points, and foundations, which empowers models prepared on this information to generalize well when experiencing real-world pictures.

To measure the effectiveness of the suggested approaches, we combined the images from the MNIST and CIFAR-10 datasets to test the model in practical settings. Since the MNIST pictures are black and white pictures of 28x28 pixels and the CIFAR-10 pictures are colored 32x32 pixels pictures, the MNIST images are resized to 32x32 pixels. It is done by having the CIFAR-10 images as the background and randomly placing the MNIST images on top as the foreground. This allows us to create a new set of composite images of handwritten digits and other objects.

This design helps us effectively focus on evaluating the models' performance and their ability to recognize handwritten digits in complex scenarios. The combined dataset presents a challenge for evaluating deep learning algorithms' ability to identify digits against diverse backgrounds. This dataset includes 10,000 images for training and 2,500 images for testing (as shown in Figure 10).

ADE20K [27] is a generally involved benchmark in the field of semantic division, covering an expansive scope of genuine scenes, from indoor spaces to outside conditions [28]. It contains more than 20,000 pictures explained at the pixel level with 150 article and stuff classifications, empowering point by point division [29]. One critical component of ADE20K explanations is the utilization of grayscale pictures for semantic marks. This choice is driven by contemplations of effortlessness, computational proficiency, and consistency. Every pixel in the grayscale mark picture relates to a particular class, with changing shades of dark addressing different item or stuff classes. This strategy improves on the handling pipeline by eliminating the requirement for multi-channel variety portrayals, thus diminishing computational overhead during preparing and assessment. Besides, the grayscale design is especially appropriate for joining with profound learning models, as it permits the model to zero in additional really on the spatial connections and dissemination examples of article classes in the picture, without being impacted by pointless color variations. (as shown in Figure 11).

Figure 9.Sample image and segmentation label from the PASCAL VOC 2012 dataset.

Figure 10.Sample image and segmentation label from the MNIST-CIFAR10 dataset.

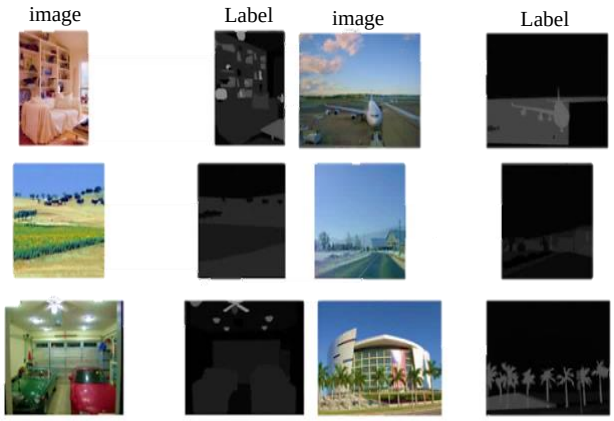

Figure 11.Sample image and segmentation label from the ADE20K dataset.

5.2. Results

The exploratory findings from Table 1 and Table 2 and Table 3 demonstrate that our proposed model, which combines two effective image-processing structures, specifically Spiking-DeepLab and RMP-Loss, outperforms the original Spiking-DeepLab [4] in semantic segmentation tasks.

Table 1 shows the result where our Spiking-DeepLab model with 20 time-steps has achieved a Mean IoU (MIOU) of 48.48% on the PASCAL VOC 2012 dataset, while the original model by [4] recorded 22.3%. This suggests an even better segmentation as the enhanced version of our model has a higher MIOU of 49.67%.

Moreover, the six time-steps Hybrid Spiking model [21] proposed for the suggested paper achieved a Mean IoU of 39.6%. This work has shown that both RMP-Loss and the enhancement of storing the membrane potential at the last time-step have led to significant improvements in segmentation accuracy compared to the Hybrid Spiking model, despite the fact that the Hybrid Spiking model is slightly worse than our model. These structural changes have made our model more accurate than hybrid models.

Similarly, in Table 2, our Spiking-DeepLab model with 20 time-steps achieved a Mean IoU of 80.52% on the combined MNIST-CIFAR10 dataset, which shows a slight improvement of 0.7% compared to the original model's 79.82%. The enhanced version of our model further enhanced this performance, reaching a Mean IoU of 81.14%, demonstrating continued progress in accuracy.

Table 3 uncovers that on the ADE20K dataset, the first Spiking-DeepLab model without RMP-Misfortune accomplished a standard mIoU of 5.84%. By consolidating RMP-Misfortune, the mIoU expanded to 7.32%, exhibiting a striking upgrade. Besides, the misfortune esteem diminished from 1.9320 to 0.7182, affirming the productivity of RMP-Misfortune in further developing the growing experience.

These outcomes highlight the adaptability of our proposed model, which accomplishes prevalent execution not just on easier datasets like MNIST-CIFAR10, yet in addition on more perplexing datasets like VOC 2012 and ADE20K. The model's capacity to maintain or further develop precision across datasets of differing intricacy features its strength and versatility.

These come about to recommend that our proposed engineering for Spiking-DeepLab can give more prominent exactness in semantic division assignments while keeping up a comparable number of time steps. This change is shown to be due to the more optimized plan of arranging layers, the utilization of progressed learning methods, and parameter optimization.

Table 1. Mean IoU (%) for Spiking-DeepLab on the PASCAL VOC 2012 dataset.

Method	Time-steps	%MIoU
Spiking-DeepLab[4]	20	22.3
Hybrid Spiking model[21]	6	39.6
Spiking-DeepLab (ours)	20	48.48
Spiking-DeepLab (ours-enhanced)	20	49.67

Table 2. Mean IoU (%)for Spiking-DeepLab on the MNIST-CIFAR10 dataset.

Method	Time-steps	%MIoU
Spiking-DeepLab[4]	20	٧٩.٨٢
Spiking-DeepLab (ours)	20	80.5٢
Spiking-DeepLab (ours-enhanced)	20	81.14

Table 3. Mean IoU (%)for Spiking-DeepLab on the ADE20K dataset.

Method	Time-steps	%MIoU
Spiking-DeepLab[4]	20	5.84
Spiking-DeepLab (ours-enhanced)	20	٧.٣2

5.3 Analysis of Robustness

Within the genuine world, cameras capture pictures that will not have the specified quality due to variables such as commotion, obscuring, or changes in surrounding lighting [30]. These issues can altogether affect the exactness of picture division models. To simulate these real-world conditions, Gaussian clamor has been utilized in this ponder. Gaussian commotion may be a sort of irregular commotion commonly watched in the picture information and effectively speaks to numerous sorts of normal clamor occurring in imaging frameworks. By shifting the standard deviation (0) of Gaussian clamor, we are ready to reenact diverse forces of commotion, permitting us to assess the model's execution beneath different loud conditions [31].

Figure 12 displays the exactness debasement of the Spiking-DeepLab show and our proposed show against distinctive levels of Gaussian commotion. As portrayed within the figure, the Spiking-DeepLab demonstration encounters noteworthy exactness corruption with expanding commotion escalated. In differentiation, our proposed demonstration, which employs the RMP-Loss work near the Spiking-DeepLab misfortune work, illustrates less exactness diminishment beneath loud conditions, in this way giving superior execution. This demonstrates the advantage of utilizing the RMP-Loss work in conjunction with the Spiking-DeepLab misfortune work in loud conditions, upgrading the solidity of our proposed demonstration against clamor. RMP-Loss diminishes the model's affectability to noise by overseeing the conveyance of membrane potential. Spiking neural networks (SNNs) display more noteworthy soundness against clamor compared to counterfeit neural networks [3] typically due to the operational characteristics of SNNs. The Poisson

spike generator in SNNs changes over pixel values into time spikes with a random dispersion. This includes making the show less touchy to the commotion and guarantees great execution in shifting lighting conditions, counting moo light, over-the-top light, and undesirable reflections that can influence picture quality.

To assess the stability of the Spiking-DeepLab, the model's exactness in both non-noisy conditions and loud conditions is computed utilizing the Mean IoU (MIoU) metric. The precision corruption is decided by calculating the percentage decrease in MIoU between non-noisy tests and noisy tests. Within the non-noisy condition, the model's exactness is computed through MIoU, characterized as follows:

$$IoU(c) = \frac{overlap}{prediction + ground \quad truth - overlap} \quad (6)$$

$$MIOU = mean\left(IOU(c)\ for \quad all \quad classes \quad c\right) \quad (7)$$

In the noisy condition, we begin with calculating the model's MIoU utilizing silent tests (Clean MIoU). At that point, Gaussian clamor with zero cruel and shifting standard deviation (σ) is included in the model's inputs to make boisterous inputs. At this arrangement, the model's MIoU in this condition is additionally calculated and alluded to as Clamor MIoU. At last, the relative exactness of corruption is obtained by calculating $\frac{(CleanMIOU - NoiseMIOU)}{CleanMIOU} * 100$.

A lower esteem demonstrates more noteworthy steadiness against commotion. This exploration makes a difference in assessing the versatility of spike-based semantic division models beneath real-world conditions and within the nearness of clamor.

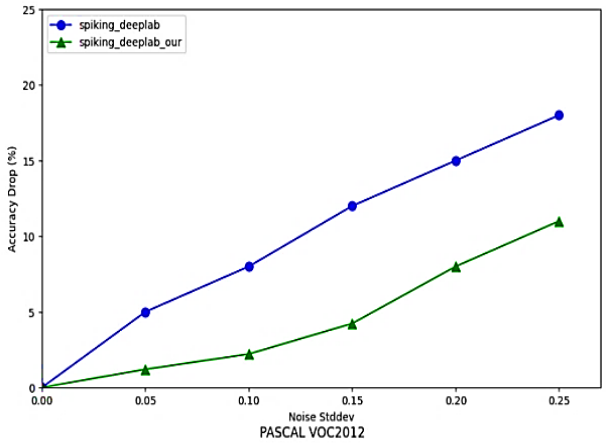

Figure 12. Comparison of Model Accuracy Drop against Gaussian Noise.

6. Conclusion

In this regard, the present study focuses on exploring and integrating two well-performing architectures in the image processing domain, namely Spiking-DeepLab and RMP-Loss. The proposed research is aimed at improving the effectiveness of semantic image segmentation through the application of spiking neural networks. Based on these merits, SNNs are regarded as some potential candidates for other image-processing tasks due to their low energy consumption as well as temporal dynamics.

The Spiking-DeepLab model based on DeepLabv3 managed the difficulties in the training of SNNs well by using the RMP-Loss function. This effectively improves the performance of the model by optimizing the MP distribution and reducing the quantization error. This model was evaluated with standard datasets, MNIST-CIFAR10, PASCAL VOC 2012 and ADE20K; this has resulted in very much improved semantic segmentation accuracy compared to other previously proposed approaches.

By observing the results, Spiking-DeepLab with RMP-Loss yielded an accuracy of 49.67% with a loss of 0.171 for PASCAL VOC 2012, while Spiking-DeepLab could only achieve an accuracy of 22.3% with a loss of 0.653. The proposed model also gave an accuracy of 81.14% on the MNIST-CIFAR10 dataset against an accuracy of 79.82% on the baseline model.

Additionally, when tested on the ADE20K dataset, the baseline Spiking-DeepLab model achieved a mean Intersection over Union (mIoU) of 5.84%. However, when RMP-Loss was incorporated, the mIoU enhanced to 6.22%, with the loss decreasing from 1.9320 to 1.7156. This demonstrates that our approach not only enhances performance on simpler datasets like MNIST-CIFAR10 and PASCAL VOC 2012 but also shows significant improvement on more complex datasets such as ADE20K, which includes a greater variety of object categories and scene complexities.

This demonstrates that our approach achieves greater accuracy improvements on complex datasets such as VOC 2012, which includes 20 object classes, compared to the MNIST-CIFAR10 dataset, which contains only 2 classes. Furthermore, the evaluation of the model on the ADE20K dataset, comprising 150 object classes, highlights the exceptional performance of the proposed model in processing real-world and complex scenes. The utilization of these three diverse datasets, ranging from the simplest to the most complex, has enabled a comprehensive assessment of the model's performance across

various domains and ensured its generalizability to different types of data.

Notice that this research is an enhancement of two architectures; the basic innovation here is the integration of RMP-Loss with Spiking-DeepLab. Thereby, this will enable us to take advantage of both architectures for better performance in semantic image segmentation.

Finally, considering continuous technique improvement in image processing and machine learning, it is necessary to design efficient, versatile models that could solve newly arisen problems in this domain. The approach described above will help not only to optimize the SNNs themselves but possibly to find better solutions to a lot of problems in different areas of medicine, robotization, and self-driving cars.

References

[1] H. Gholamalinejad and H. Khosravi, "Whitened gradient descent, a new updating method for optimizers in deep neural networks," *Technol. J. Artif. Intell. Data Min.*, vol. 10, no. 4, pp. 467–477, 2022, doi: 10.22044/jadm.2022.11325.2291.

[2] Q. Sun, C. Bai, H. Geng, and B. Yu, "Deep Neural Network Hardware Deployment Optimization via Advanced Active Learning," in *Proceedings -Design, Automation and Test in Europe, DATE*, 2021. doi: 10.23919/DATE51398.2021.9474100.

[3] L. Deng *et al.*, "Rethinking the performance comparison between SNNS and ANNS," *Neural Networks*, vol. 121, 2020, doi: 10.1016/j.neunet.2019.09.005.

[4] Y. Kim, J. Chough, and P. Panda, "Beyond classification: directly training spiking neural networks for semantic segmentation," *Neuromorphic Comput. Eng.*, vol. 2, no. 4, 2022, doi: 10.1088/2634-4386/ac9b86.

[5] Y. Guo *et al.*, "RMP-Loss: Regularizing Membrane Potential Distribution for Spiking Neural Networks," pp. 17391–17401, 2023, [Online]. Available: http://arxiv.org/abs/2308.06787.

[6] I. Ulku and E. Akagündüz, "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images," *Applied Artificial Intelligence*, vol. 36, no. 1. Taylor and Francis Ltd., 2022. doi: 10.1080/08839514.2022.2032924.

[7] M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov, "On Regularized Losses for Weakly-supervised CNN Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. doi: 10.1007/978-3-030-01270-0_31.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. doi: 10.1007/978-3-319-24574-4_28.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, 2017, doi: 10.1109/TPAMI.2016.2644615.

[10] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, 2018, doi: 10.1109/TPAMI.2017.2699184.

[11] S. Bukhori, M. Almas Bariiqy, W. Eka, and J. A. Putra, "Segmentation of Breast Cancer using Convolutional Neural Network and U-Net Architecture," *Technol. J. Artif. Intell. Data Min.*, vol. 11, no. 3, pp. 477–485, 2023, doi: 10.22044/jadm.2023.12676.2419.

[12] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Rethinking Atrous Convolution for Semantic Image Segmentation Liang-Chieh," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, 2018.

[13] N. H. Thinh, T. Hoang Tung, and L. V. Ha, "Depth-aware salient object segmentation," *VNU J. Sci. Comput. Sci. Commun. Eng.*, vol. 36, no. 2, 2020, doi: 10.25073/2588-1086/vnucsce.217.

[14] J. K. Eshraghian *et al.*, "Training Spiking Neural Networks Using Lessons from Deep Learning," *Proc. IEEE*, vol. 111, no. 9, 2023, doi: 10.1109/JPROC.2023.3308088.

[15] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks*, vol. 10, no. 9, 1997, doi: 10.1016/S0893-6080(97)00011-7.

[16] H. Aghabarar, K. Kiani, and P. Keshavarzi, "Digit Recognition in Spiking Neural Networks using Wavelet Transform," *Technol. J. Artif. Intell. Data Min.*, vol. 11, no. 2, pp. 247–257, 2023, doi: 10.22044/jadm.2023.12613.2415.

[17] S. Schmidgall, J. Ashkanazy, W. Lawson, and J. Hays, "SpikePropamine: Differentiable Plasticity in Spiking Neural Networks," *Front. Neurorobot.*, vol. 15, 2021, doi: 10.3389/fnbot.2021.629210.

[18] S. A. Lobov, A. N. Mikhaylov, M. Shamshin, V. A. Makarov, and V. B. Kazantsev, "Spatial Properties of STDP in a Self-Learning Spiking Neural Network Enable Controlling a Mobile Robot," *Front. Neurosci.*, vol. 14, 2020, doi: 10.3389/fnins.2020.00088.

[19] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-based optimization to

spiking neural networks," *IEEE Signal Process. Mag.*, vol. 36, no. 6, 2019, doi: 10.1109/MSP.2019.2931595.

[20] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Networks*, vol. 111. 2019. doi: 10.1016/j.neunet.2018.12.002.

[21] T. Zhang, S. Xiang, W. Liu, Y. Han, X. Guo, and Y. Hao, "Hybrid Spiking Fully Convolutional Neural Network for Semantic Segmentation," *Electron.*, vol. 12, no. 17, 2023, doi: 10.3390/electronics12173565.

[22] C. Zhou, L. Ye, H. Peng, Z. Liu, J. Wang, and A. Ramírez-De-Arellano, "A Parallel Convolutional Network Based on Spiking Neural Systems," *Int. J. Neural Syst.*, vol. 34, no. 5, 2024, doi: 10.1142/S0129065724500229.

[23] D. Zipser, B. Kehoe, G. Littlewort, and J. Fuster, "A spiking network model of short-term active memory," *J. Neurosci.*, vol. 13, no. 8, 1993, doi: 10.1523/jneurosci.13-08-03406.1993.

[24] B. Quan, B. Liu, D. Fu, H. Chen, and X. Liu, "Improved deeplabv3 for better road segmentation in remote sensing images," in *Proceedings - 2021 International Conference on Computer Engineering and Artificial Intelligence, ICCEAI 2021*, 2021. doi: 10.1109/ICCEAI52939.2021.00066.

[25] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2003. doi: 10.1109/ICDAR.2003.1227801.

[26] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," *... Sci. Dep. Univ. Toronto, Tech. ...*, 2009, doi: 10.1.1.222.9220.

[27] B. Zhou *et al.*, "Semantic Understanding of Scenes Through the ADE20K Dataset," *Int. J. Comput. Vis.*, vol. 127, no. 3, 2019, doi: 10.1007/s11263-018-1140-0.

[28] Y. Kawano and Y. Aoki, "TAG: Guidance-free Open-Vocabulary Semantic Segmentation," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.11197.

[29] Y.Liu, C. Liu, K. Han, Q. Tang, and Z. Qin, "Boostin Semantic Segmentation from the Perspective of Explicit Class Embeddings," in *Proceedings of the IEEE International Conference on Computer Vision*, 2023. doi: 10.1109/ICCV51070.2023.00082.

[30] "Nisy images edge detection: Ant colony optimizaion algorithm," *J. Artif. Intell. Data Min.*, vol. 4, no. 1, 016, doi: 10.5829/idosi.jaidm.2016.04.01.09.

[31] A. . Boyat and B. K. Joshi, "A Review Paper : Noise Models in Digital Image Processing," *Signal Image Process. An Int. J.*, vol. 6, no. 2, 2015, doi: 10.5121/sipij.2015.6206.

بهبود کارایی تقسیم‌بندی معنایی در شبکه‌های عصبی اسپایکی

الهه یدالهی و شیث ابوالمعالی*

گروه مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران.

ارسال ۲۰۲۴/۰۹/۲۳؛ بازنگری ۲۰۲۴/۱۱/۲۳؛ پذیرش ۲۰۲۵/۰۱/۲۴

چکیده:

تقسیم‌بندی معنایی یک وظیفه حیاتی در بینایی کامپیوتر است که بر استخراج و تحلیل اطلاعات بصری دقیق تمرکز دارد. شبکه‌های عصبی مصنوعی سنتی (ANNs) پیشرفت‌های قابل‌توجهی در این زمینه داشته‌اند، اما شبکه‌های عصبی اسپایکی (SNNs) به دلیل بهره‌وری انرژی و پردازش زمانی الهام‌گرفته از سیستم‌های زیستی، توجه زیادی را به خود جلب کرده‌اند. با این حال، روش‌های موجود مبتنی بر SNN برای تقسیم‌بندی معنایی با چالش‌هایی در دستیابی به دقت بالا مواجه هستند که ناشی از محدودیت‌هایی مانند خطاهای ناشی از کوانتیزاسیون و توزیع نامطلوب پتانسیل غشایی است. این پژوهش یک رویکرد اسپایکی نوین مبتنی بر Spiking-DeepLab معرفی می‌کند که از تابع هزینه تنظیم‌شده برای توزیع پتانسیل غشایی (RMP-Loss) برای مقابله با این چالش‌ها بهره می‌گیرد.مدل پیشنهادی که بر اساس معماری DeepLabv3 ساخته شده است، از RMP-Loss برای بهبود دقت تقسیم‌بندی از طریق بهینه‌سازی توزیع پتانسیل غشایی در SNNها استفاده می‌کند. با بهینه‌سازی ذخیره‌سازی پتانسیل‌های غشایی، جایی که مقادیر فقط در گام زمانی نهایی ذخیره می‌شوند، مدل به‌طور قابل‌توجهی استفاده از حافظه و زمان پردازش را کاهش می‌دهد. این بهبود نه‌تنها بهره‌وری محاسباتی را افزایش می‌دهد، بلکه دقت تقسیم‌بندی معنایی را نیز بهبود بخشیده و تحلیل دقیق‌تری از رفتار زمانی شبکه را امکان‌پذیر می‌سازد.مدل پیشنهادی همچنین در برابر نویز مقاومت بهتری از خود نشان می‌دهد و دقت خود را تحت سطوح مختلف نویز گوسی که در سناریوهای دنیای واقعی معمول است، حفظ می‌کند. رویکرد پیشنهادی عملکرد رقابتی را در مجموعه داده‌های استاندارد نشان می‌دهد و پتانسیل آن را برای کاربردهای پردازش تصویر با بهره‌وری انرژی بالا به نمایش می‌گذارد.

کلمات کلیدی: یادگیری نظارت‌شده، پردازش تصویر، بخش‌بندی معنایی، شبکه‌های عصبی اسپایکی، RMP-Loss.