Handwritten Digit Recognition Based On Spiking-Vovnet and Extended Application

Yonghui Guo1,*, Yuntao Li2, Yu Zhang3, Yonggang Li4 and Yonghua He5
1,2,3,4,5Space Engineering University, Beijing, China, 101416

Abstract. Spike Neural Network (SNN) can provide an efficient reasoning process. In this paper, by modifying the Vovnet network to use IF neurons to replace the traditional Vovnet RELU neurons, testing on the open handwritten digit mnist data set, and comparing the Vovnet network structure of CNN and SNN for comparative experiments, it is found that the Spiking-Vovnet network is guaranteed and traditional Under the premise of the accuracy of the Vovnet network, the speed has been greatly improved, and at the same time, the resource consumption is reduced. In order to expand the application, in the synthetic aperture radar (SAR) image target detection, the Spiking-Vovnet network is used as the FCOS ship target detection feature extraction network. For promote the prise degree of detection, the variability convolutional neural network is lead into design a SAR ship Feature extraction network for ship target detection.

Keywords: Spike neural network; Vovnet Network Structure; Spiking-Vovnet Network Structure; Ship Inspection; FCOS Network; Deformable Convolutional Neural Network.

1. Introduction

1.1. Formula Deduction

Pulse in an era of the convolution neural network, neural network as the third generation of artificial intelligence shows unique advantages, how to transform convolution neural network to impulse neural network has become the focus of current research, Cao and others for the first time, puts forward a transformation RELU layer mechanism, but the theory part of the work is very lack. Based on the analysis and interpretation of CAO[1], Rueckauer B[2] et al proposed a simple modification of the reset mechanism after impulse nerve firing, in which each impulse neural network neuron was converted into an unbigoted approximation of the objective function. In this paper, by learning and replicating the research of Rueckauer B[2] et al., the Spiking-Vovnet network structure is formed through modification, and the MNIST[3] data set of hand-written digital recognition is experimented and compared with the recognition accuracy of Spiking-Vovnet network structure and Vovnet network structure. And pave the way for the follow-up target interpretation of SAR image.

The basic principle of transforming SNN into ANN is the granting frequency of pulsed neurons ought to match the simulated activation value in ANN. By studying Rueckauer B[2] et al.'s paper, it is found that there is a one-to-one mapping between each SNN unit and ANN unit, even if it is possible for ANN to correspond to a set of pulsing neurons. For an L layer network, we let $W_l$, $l \in \{1, 2, 3, \cdots, L\}$ denote the weight matrix from the $l-1$ layer to the $l$ layer, biased as $b_l$. The quantity of neurons in
storey $l$ is $M_i$. Then the calculation of the continuous value of neuron $l$ in layer $i$ through RELU layer is as follows:

$$a_i^l := \max(0, \sum_{j=1}^{M^{l-1}} W_{ij} a_{j}^{l-1} + b_i^l)$$  

(1)

Consider $a^0 = x$, represents the normalized input where $x \in [0,1]$ the membrane of the first SNN neuron at time Potentiometric Indicates that the input potential for each time step is

$$z_i^l(t) = V_{thr} \left( \sum_{j=1}^{M^{l-1}} W_{ij} \Theta_{i,j}^{l-1} + b_i^l \right)$$  

(2)

In this case, $V_{thr}$ is the threshold location, and $\Theta_{i,j}^{l-1}$ is whether the pulse at time $t$ is generated or not:

$$\Theta_{i,j}^{l-1} := \Theta(V_i^l(t-1) + z_i^l(t) - V_{thr})$$  

(3)

Where $\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{other} \end{cases}$

The time window is $T$ and the step size of each time is $\Delta t \in R^+$, The maximum pulse emissivity is $r_{max} = 1/\Delta t$. The firing rate of the first layer input is related to the pixel value of the constant, or RGB channel, in which the emissivity of each SNN neuron $i$ is

$$r_i^l(t) := \frac{N_i^l(t)}{t}$$  

(4)

Here $N_i^l(t) := \sum_{r=1}^{t} \Theta_{r,i}^l$ is the quantity of pulses emitted up to time $t$ the number of pulses.

By reading the paper of Rueckauer B et al. it was found that IF neurons in SNN are unbiased estimators of the RELU activation function in time, and for the first layer of the neural network, i.e. the input layer, the relationship between the issuance rate $r$ of SNN neurons and the corresponding activation in ANN is discussed assuming a constant input of $z \in [0,1]$. For the IF neuron with subtractive resetting its membrane potential $V$ varies with time as

$$V_i = V_i - 1 + z - V_{\text{threshold}} \Theta_t$$  

(5)

Where $V_{\text{threshold}}$ is the firing threshold, usually set to 1.0 for the output pulse. The average firing rate in time step $T$ can be obtained by summing the membrane potentials.

$$\sum_{t=1}^{T} V_i = \sum_{t=1}^{T} V_{t-1} + zT - V_{\text{threshold}} \sum_{t=1}^{T} \Theta_t$$  

(6)

Shift all terms containing $V_i$ to the left and divide both sides simultaneously by $T$:

$$\frac{V_T - V_0}{T} = \frac{\sum_{t=1}^{T} \Theta_t}{T} = z - V_{\text{threshold}} \frac{N}{T}$$  

(7)

where $N$ is the quantity of pulses in $T$ time steps and $N/T$ is the release rate $r$. Using $z = V_{\text{threshold}} \alpha$ that is.
\[ r = a - \frac{V_T - V_0}{TV_{threshold}} \]  

Therefore, in the case of an infinitely long simulation time step \( T \).

\[ r = a, (a > 0) \]  

Similarly, for higher levels of the neural network, the interlayer release rate is fully sufficient.

\[ r^t = W^t r^{t-1} + b^t - \frac{V^t}{TV_{threshold}} \]  

1.2. Simulation Experiments

Python 3.6 was used as the experimental environment, and the experiment was carried out in the environment of pycharm2019. In order to verify the comparison capability between IF neuron and RULE function in the course of the experiment, the experiment adopted PyTorch1.6.0 framework for operation. The specific experimental process is as follows:

In this paper, we first observe the output pulse and the frequency of the pulse by giving constant input to the IF neuron. First, create a new IF neuron layer, determine the input and draw the input \( x_i \) of each IF neuron. The experimental results are shown in Figure 1:

![Figure 1. Input to IF neurons.](image)

Input is sent to the IF neuron layer, and \( T=128 \) steps are run to observe the pulses emitted by each neuron and the pulse emission frequency, as shown in Figure 2.

![Figure 2. IF neurons.](image)

It can be found that the frequency of the pulse emission is proportional to the size of the input \( x_i \) within a certain range. Draw the curve of IF neuron pulse emission frequency and input \( x_i \), and compare it with RELU ( \( x_i \)). Through simulation comparison, it is found that the curves of the two are
almost the same, and the pulse frequency cannot be greater than 1. Therefore, IF neuron cannot fit the situation that the input of RULE in ANN is greater than 1. The experiment is shown in Figure 3.

**Figure 3.** IF Neurons versus ReLU functions.

### 1.3. Handwritten Digit Recognition Experiment

Handwritten digital MNIST data set as a classic data set is widely used in network testing. In order to carry out experiments on the extended SAR image target interpretation, this paper adopts Vovnet network structure for hand-written digital recognition, uses IF neuron to replace RELU function, constructs Spiking-Vovnet network for hand-written digital recognition detection, and uses Backbone of subsequent FCOS[7] model as feature extraction network model.

In Vovnet, multiple 3x3 small size convolution series are used to replace the large-size convolution layer, which can not only use fewer parameters to have the same receptive field, but also increase the number of nonlinear operations, so as to reduce the complexity of the model and elevate the expression capacity of the model. In order to improve the detection speed, the SPIKING-Vovnet network model is constructed by modification on the basis of the Vovnet network. The specific structure of the Vovnet network model and the SPIKING-Vovnet network model are shown in Figure 4 and Figure 5.

**Figure 4.** Vovnet network model.

**Figure 5.** Spiking-Vovnet network structure.
Operating system Window10; Software: Deep learning based Pytorch and Python Open Loop PyCharm.CUDA10.0 was used to speed up calculations. By modifying the model, the commonly used handwritten digital data set is used to test the network structure. By comparing the VOVNET network model and Spiking-Vovnet network, it is found that the detection time is enormously decreased while the accuracy is not changed much. Fig. 6 is the result of the experiment.

![Figure 6. Test results.](image)

| Model         | Detection rate/% | Testing time/s |
|---------------|------------------|----------------|
| Vovnet39      | 98%              | 0.6            |
| Spiking-Vovnet39 | 96.54%         | 0.35           |

Through the experiment, it is found that the accuracy of Vovnet network model is 98%, and the detection time is 0.6s. The accuracy of SPIKING-Vovnet network is 96.54%, and the detection time is 0.35s. Under the premise of little change in precision, the detection time is greatly reduced.

Since the research direction is SAR image target interpretation, the SPIKING-Vovner network as feature extraction network of FCOS model to detect the target in SAR image.

2. Target Detection Based on Improved FCOS Model

2.1. FCOS Target Detection Process

FCOS is an peer-to-peer target detection technique on account of deep convolutional neural network put forward by Zhi. And the method takes target detection as the solution of regression problem, and predicts the distance of each pixel to the real bounding box up, down, left and right by making a direct prediction for each pixel. By introducing the feature pyramid FPN structure, different layers are used to deal with different target boxes to avoid the overlap of targets and the inability to accurately judge the category of pixels. By introducing the center degree layer to find the center point of the target, that is, the near to the center of the target, the greater the value of the center degree, vice versa, to reduce the false detection of the target detection. In the model prediction stage, centrality value was multiplied by the classification output value, and the product of the two values was used as the final confidence score. Non-maximum suppression (NMS) was used to filter the false detection frame and improve the detection accuracy. FCOS does not need to extract the candidate region and avoids the manual design of the prior box, which enormously improves the velocity of target detection.

The structure of FCOS network is shown in Fig. 7. FCOS consists of two sub-networks: full convolution (Vovnet) neural network feature extraction and FPN multi-layer feature fusion. Among them P3, P4 and P5 of the FPN network are generated by the top down connection of 1x1 convolution of C3, C4 and C5 in the Full Convolution (Vovnet) neural network, and are predicted by
fusing the features of the five layers. Finally, by sharing the head among different layers, not only the parameters are reduced, but also the accuracy is improved.

2.2. Target Detection Based on Improved FCOS Model

This paper presents an improved FCOS model network for end-to-end training. Improve the basic network, and form the Spiking-Vovnet feature extraction network model. In order to increase the docking target survey, introduce the deformable convolution god into the network to form Spiking-Vovnet DCN\textsuperscript{[10]}. Compared with the traditional Vovnet network, in this paper, the Spiking-Vovnet DCN module is formed by introducing the deformable convolution god into the OSA5 module of the spiking-vovnet network, and the characteristic gold tower of FCOS is retained. The module Spiking-Vovnet DCN\textsuperscript{[11]} is shown in figure 8. The deformable convolution neural network is pulled into Spiking-Vovnet draw characteristic of the network, which improves the accuracy of SAR image target detection.

3. Simulation Analysis

In order to prove the performance of the proposed method, the improved SAR target detection feature extraction network is trained by constructing data sets and adopting reasonable training strategies. At the same time, for the generalization ability of the network, the trained model is used to detect the SAR image of gaofen-3 satellite. The ship target detection results were objectively compared with the traditional FCOS model and FCOS-DCN-VOVNET39 model.

3.1. Data Set Construction

For the detection of ship targets in this paper, the data sets used in the experiment are mainly from the open source Gaofen-3 satellite measured image data set on the sea surface, the SSDD ship detection open data set, and the TerraSAR-X measured SAR image data set collected by the network. The details are as follows:

Data set first\textsuperscript{[12]}: Target slices were made from the sea surface observation images using Tenrrasar-X, as shown in Fig. 10. The dataset contains 218 416×416 pixel target slice images, a total of 427 ship targets, 183 of which work in the hyperfine stripe mode, resolution is 3m, polarization mode is DH.
polarization, and the remaining 35 work in the beam mode. The resolution is 1m, and the polarization mode is HH polarization and VV polarization.

![Raw SAR image](image1.png) ![Ship target slice](image2.png)

**Figure 9.** SAR sea surface observations raw images and target slices.

Data set two[^13^]: Public SSDD data set is adopted. At present, most domestic researches on ship targets in SAR images take this data set. The SSDD data set mainly includes RADARSAT-2, TanrrASAR-X and Sentinel-1 sensors, and four polarization modes of HH, HV, VV and VH. The resolution is 1m-15m, and ships are targeted in large areas of ocean and near shore. In the SSDD data set, there are a total of 1160 images and 2456 ships, with an average of 2.12 ships per image.

Data set three: selects 7 measured Gaofen-3 images collected by the network as the experimental data set to evaluate the generalization ability of the final model. This data set consists of images with different irradiation angles, different imaging modes, different polarization modes and different resolutions.

### 3.2. Analysis of Ship Target Detection Results

In order to more accurately evaluate the target detection performance of the proposed network in complex scenes, three indexes of false alarm rate $P_f$, detection rate $P_d$ and quality factor FOM were used for quantitative analysis:

$$
\begin{align*}
    P_f &= \frac{N_f}{N_f + N_d} \\
    P_d &= \frac{N_d}{N_f} \\
    FOM &= \frac{N_d}{N_f + N_d}
\end{align*}
$$

The experiment is aimed at the method in this paper, the traditional FCOS model and FCOS-DCN-Vovnet39 model target detection method obtains target detection results for objective comparison. The FCOS model target detection method adopts Vovnet39 as the Backbone. The above method is used to train on PASCAL VOC data set, and the comparison results are shown in Table 2.

| Model                  | Detection rate | Figure of Merit | Detection time |
|------------------------|----------------|-----------------|----------------|
| FCOS-Vovnet39          | 71.43          | 63.83           | 1.484          |
| FCOS-DCN-Vovnet39      | 92.86          | 86.67           | 1.624          |
| The design model of this article | 82.56          | 76.63           | 0.942          |

[^13^]: Public SSDD data set
This paper improves fcos model and applies it to ship target detection. Without changing the target detection rate, the detection time is greatly reduced and the detection quality factor is improved. However, compared with fcos-dcn-vovnet39 model, the proposed method greatly reduces the detection time, but sacrifices the detection probability and quality factor.

4. Conclusion

This paper propose a SPIKING-Vovnet network model, which verifies that the speed of SPIKING-Vovnet network is faster than that of traditional Vovnet network to a certain extent, through the verification of handwritten digit recognition and ship target detection. In the SAR image target detection experiment, the SAR images from different sources are used to construct the sample data set of ship detection and carry out network training. The experimental results show that, compared with the traditional model detection methods of FCOS and FCOS-DCN-Vovnet 39, the Spiking-Vovnet network designed in this paper improves the detection time without significantly reducing the detection rate and quality factor of ship detection, which proves the effectiveness of the proposed method.

References

[1] Cao Y, Chen Y, Khosla D. Spiking deep convolutional neural networks for energy-efficient object recognition[J]. International Journal of Computer Vision, 2015, 113(1): 54-66.
[2] B. Rueckauer, I.-A. Lungu, Y. Hu, M. Pfeiffer, and S.-C. Liu. Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. Frontiers in neuroscience, 11:682, 2017.
[3] Lecun Y , Bottou L . Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
[4] Hu C H,Chen C,He C,Pei H,Zhang J X. “SAR detection for small target ship based on deep convolutional neural network.Journal of Chinese Inertial Technology”.1005-6734(2019)03-0397-09.
[5] Yue B Z,Han S. “A SAR Detection Method Based on Improved Faster R-CNN.Ji Suan Ji Yu Xi Dai Hua.”1006-2475(2019)09-0090-06.
[6] Kang M,Leng X,Lin Z,et al . “A modified faster R-CNN based on CFAR algor-Ithm for SAR ship detection”[C]/IEEE International Workshop on Remote Sensing with Intwlligent Processin,2017.
[7] Tian, Zhi and Shen, et, al. “Fully Convolutional One-Stage Object Detection”.Proc.Int.Conf. Computer Vision (ICCV),2019.
[8] Youngwan Lee, Joong-won Hwang, et, al.“An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection”.Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.
[9] Lin, Tsung-Yi and Dollar, et, al.“Feature Pyramid Networks for Object Detection”.Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017.
[10] Zhu, Xizhou and Hu, et, al“Deformable ConvNets V2: More Deformable, Better Results”.Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2019.
[11] Fan H C, Li W Z, Zhang C Q.“Anchor-free traffic sign detection”[J].Journal of Geo-information Science, 2020,22（1）: 88-99.DOI: 10.12082/dqxkx.2020.190424.
[12] SUN Xian, WANG Zhirui, SUN Yuanrui, et al. “AIR-SARShip-1.0: High-resolution SAR ship detection dataset[J]”. Journal of Radars, 2019, 8(6): 852–862. DOI: 10.12000/JR19097
[13] Li J , Qu C , Shao J . “Ship detection in SAR images based on an improved faster R-CNN”[C]/Sar in Big Data Era: Models, Methods & Applications. IEEE, 2017. DOI: 10.1109/BIGSARDATA.2017.8124934.