# Spiking Neural Network Enhanced Gesture Recognition Using Low-Cost Single-Photon Avalanche Diode Array

Zhenya Zang, Xingda Li, Xi Chen, David Day Uei Li

## Abstract

We present a compact spiking convolutional neural network (SCNN) and spiking multilayer perceptron (SMLP) to recognize stationary gestures in a completely dark environment, using a 9.6 USD single-photon avalanche diode (SPAD) array with 8×8 spatial resolution. Photon intensity data from the sensor is leveraged to train and test the network. A vanilla convolutional neural network (CNN) is also implemented to compare the performance of the SCNN, with the same network topologies and training strategies. The SCNN is trained from scratch instead of being converted from CNN. The result indicates that SCNN achieves comparable result than CNN and exhibits much lower computational complexity with only 8 timesteps. The code and dataset are available at

<https://github.com/zzy666666zzy/TinyLiDAR_NET_SNN>.

## Introduction

Spiking neural network (SNN), known as a typical implementation of neuromorphic computing, is an emerging algorithm for a wide range of applications in vision [Ref], due to its high computational energy efficiency and accuracy. Hardware communities also leverage the high efficiency to design neuromorphic chips, such as Tianjic [1], TrueNorth [2], Liohi [3], etc. SNN usually processes data from event-based cameras that only record moving objectives and generate event-sequences, represented by spike streams. In this work, while maintaining the accuracy, we aim to pursue a low-cost gesture recognition, in terms of product price and computing. The contribution of this work is three-fold:

We generate and opensource a dataset of gestures captured by a low-cost SPAD array. With the datasets, we train SCNN and SMLP from scratch instead of using conversion to guarantee the accuracy and computing.

We evaluate SCNN’s and SMLP’s accuracy of classification in training and inference phase.

We discuss the computational complexity of SCNN and SMLP and demonstrate the efficacy of hardware implementation.

## Related Work

### Machine learning and Low-cost SPAD

The series of low-cost, portable sensor from ST microelectronics (shorthanded ST hereafter) become increasingly popular in the research of computer vision and pattern recognitions, with the synergy of machine learning. A sensor embedding a VSCEL module and a SPAD array, VL53L1X, was used to classify five types of objects by using the on-chip histogram method to capture reflected photon from the objectives [4]. The same sensor was employed to capture low-resolution depth images and generate high-resolution and skeleton images using DL [5]. Also, five VL53L1X were mounted on a small drone to realize obstacle avoiding and maze-solving [6]. ST also developed its integrated software tool to realize dynamic gesture recognition using time-of-flight technology [7]. The sensor was also successfully integrated with a costume RISC-V processor on a drone to measure its motion and distance from the ground, thereby assisting detecting human poses using DL [8].

## Neural Network Details

The sensor was configured as intensity mode in the firmware to generate photon counts for each pixel. All the images in the datasets were collected in a dark room. The training datasets include 5,000 images of 10 kinds of gestures, where each kind of gestures have 500 images. With the same ratio, the test datasets have 1,000 images, with 100 images for each gesture. Networks in this work were implemented using PyTorch. We utilized Bicubic interpolation [9]method from OpenCV to enhance the spatial resolution from 4×4 to 25×25 yet try to keep high fidelity. During the training phase, normalized images in each batch were encoded as Poisson spike train with 8 timesteps, using rate encoding [10]. We used early stopping strategy with 20 epoch patients to avoid over-fitting. Adam is the optimizer. The learning rate is 10­-3. Cross-entropy is the loss function.

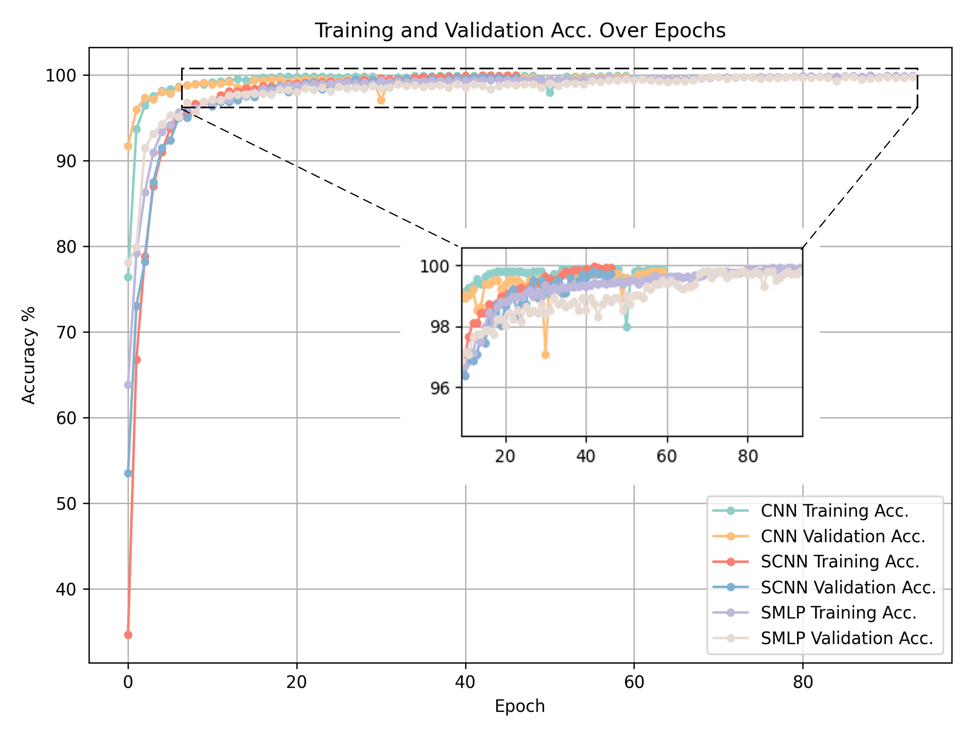


Figure . Training and validation accuracy over epochs. CNN, SCNN, and SMLP terminate training at 58, 48, and 93 epochs, respectively.

Two types of SNNs, SCNN and SMLP were constructed to investigate their differentiation in term of performance. In SCNN, batch normalization follows each convolutional layer to accelerate the training convergence. Leaky Integrate-and-Fire (LIF) neuro was adopted as the activation function. Besides, SMLP includes two fully connected layer followed by drop-off modules to alleviate overfitting. Details of the architectures are depicted in Fig. We also designed a vanilla CNN with the same training strategies and network topology to compare the training performance versus the SNNs. As shown in Fig., the vanilla CNN shows faster convergence than SCNN and SMLP. And the final accuracy of three networks is nearly same.

## Evaluation on Test Datasets

We used confusion matrix to evaluate the accuracy of classification of three networks. As shown in Fig, CNN.

## Conclusion

Reference

[1] J. Pei *et al.*, ‘Towards artificial general intelligence with hybrid Tianjic chip architecture’, *Nature*, vol. 572, no. 7767, pp. 106–111, 2019.

[2] F. Akopyan *et al.*, ‘Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip’, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 34, no. 10, pp. 1537–1557, 2015.

[3] M. Davies *et al.*, ‘Loihi: A neuromorphic manycore processor with on-chip learning’, *Ieee Micro*, vol. 38, no. 1, pp. 82–99, 2018.

[4] C. Callenberg, Z. Shi, F. Heide, and M. B. Hullin, ‘Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging’, *ACM Trans. Graph. TOG*, vol. 40, no. 4, pp. 1–12, 2021.

[5] A. Ruget *et al.*, ‘Pixels2Pose: Super-resolution time-of-flight imaging for 3D pose estimation’, *Sci. Adv.*, vol. 8, no. 48, p. eade0123, 2022.

[6] S. Pikalov, E. Azaria, S. Sonnenberg, B. Ben-Moshe, and A. Azaria, ‘Vision-Less Sensing for Autonomous Micro-Drones’, *Sensors*, vol. 21, no. 16, p. 5293, 2021.

[7] STMicroelectronics, ‘Turnkey gesture recognition solution based on VL53L5CX, VL53L7CX and VL53L8CX multizone Time-of-Flight ranging sensors’. STMicroelectronics, 2023. [Online]. Available: https://www.st.com/en/embedded-software/stsw-img035.html#documentation

[8] D. Palossi *et al.*, ‘Fully onboard ai-powered human-drone pose estimation on ultralow-power autonomous flying nano-uavs’, *IEEE Internet Things J.*, vol. 9, no. 3, pp. 1913–1929, 2021.

[9] ‘OpenCV: Geometric Image Transformations’. [Online]. Available: https://docs.opencv.org/4.x/da/d6e/tutorial\_py\_geometric\_transformations.html

[10] E. D. Adrian and Y. Zotterman, ‘The impulses produced by sensory nerve endings: Part 3. Impulses set up by Touch and Pressure’, *J. Physiol.*, vol. 61, no. 4, p. 465, 1926.