

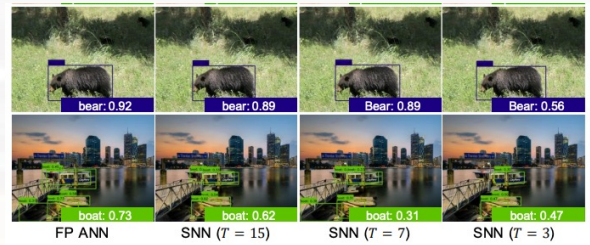
FEEL-SNN: Robust Spiking Neural Networks with Frequency Encoding and Evolutionary Leak Factor

Mengting Xu^{1,2}, De Ma^{1,2}, Huajin Tang^{1,2},
Qian Zheng^{1,2*}, Gang Pan^{1,2*}

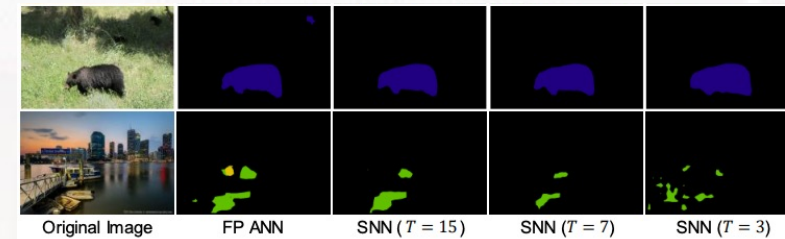
¹ The State Key Lab of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China

² College of Computer Science and Technology, Zhejiang University, Hangzhou, China

- Brain-inspired spiking neural networks (SNNs) have been increasingly prominent recent years



object detection



semantic segmentation

Image source: fast-snn [TAPAMI 2023]

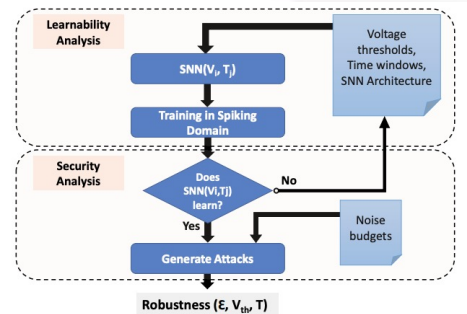
- SNNs are still vulnerable to adversarial noise



Image source: Intriguing properties of neural networks [ICLR2014]

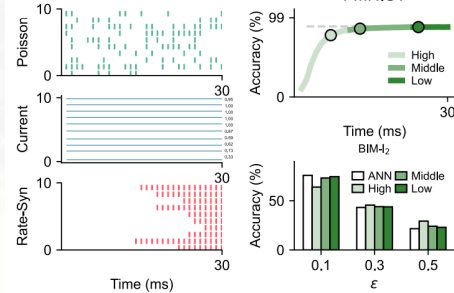
Improving the robustness of SNNs is crucial for their real-life deployment

Some work focuses on SNN robustness analysis



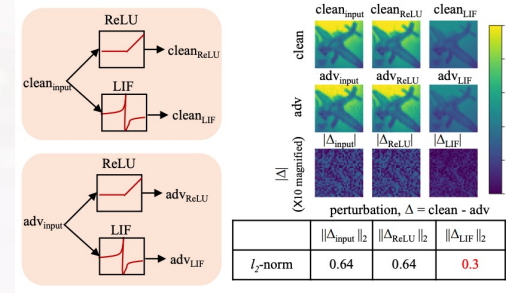
Voltage threshold

[Rida El-Allami et al, DATE 21]



Spiking timing

[Jianhao Ding et al, arxiv23]

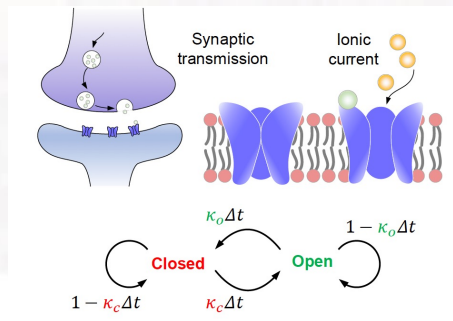


Non-Linear Activations

[Saima Sharmin et al, ECCV20]

- Solely on experimental analysis
- Lack theoretical support

Other work aims to improve the robustness of SNN from biological aspects.



Stochasticity

[Jianhao Ding et al, AAAI24]

- Result in the loss of the original information

■ Design a highly robust SNN model that more closely mimics the biological nervous system

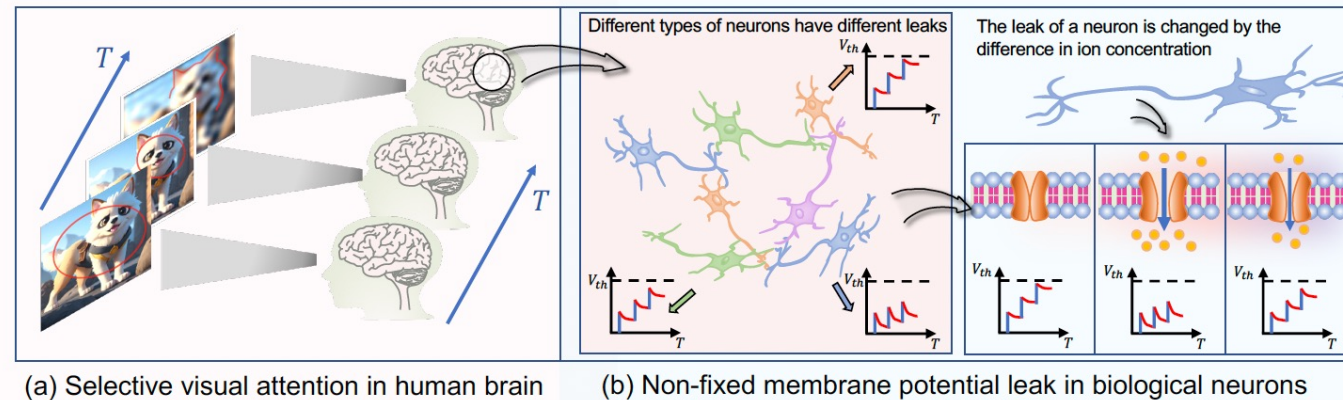


Figure 1: Illustration of the (a) selective visual attention and (b) non-fixed membrane potential leak in biological nervous system.

➤ Selective visual attention mechanism

Selectively focuses on stimuli of different frequencies over time and can filter out unwanted information


➤ Non-fixed membrane potential leak

The changes in membrane potential in biological neurons are determined by ion concentration inside and outside the cell membrane. Different environments and types of nerve axon fibers can affect the degree of leak of the membrane potential.

■ The robustness analysis of SNNs

- The robustness of the model is quantified as $\mathcal{L}(x + \epsilon) - \mathcal{L}(x)$
- Local linearity technique $\mathcal{L}(x + \epsilon) - \mathcal{L}(x) \leq |\epsilon \odot \nabla_x \mathcal{L}(x)|_1 + g(\epsilon, x)$

Theorem 1 *Given an L -layered SNN intended to inference T time-steps with λ as the leak factor, suppose that there are N_l neurons in layer l for $l = 1, 2, \dots, L$. $\lambda_l \in \mathbb{R}^{N_l \times T}$, $W_{l-1, l} \in \mathbb{R}^{N_l \times N_{l-1}}$, it satisfies:*

$$\min \sum_t \left| \epsilon(t) \odot \frac{\partial \mathcal{L}}{\partial x^t} \right|_1 = \min \sum_t \frac{1}{L} \sum_{l=1}^L \left[\underbrace{\left(\prod_{k=t}^T \epsilon(t) \odot \lambda_l^k \right)}_{\odot} \cdot \prod_{q=2}^l W_{q-1, q} \cdot \prod_{v=1}^l \frac{\partial O_v^t}{\partial u_v^t} \cdot \frac{\partial \mathcal{L}}{\partial O_l^T} \right]_1$$


Frequency encoding

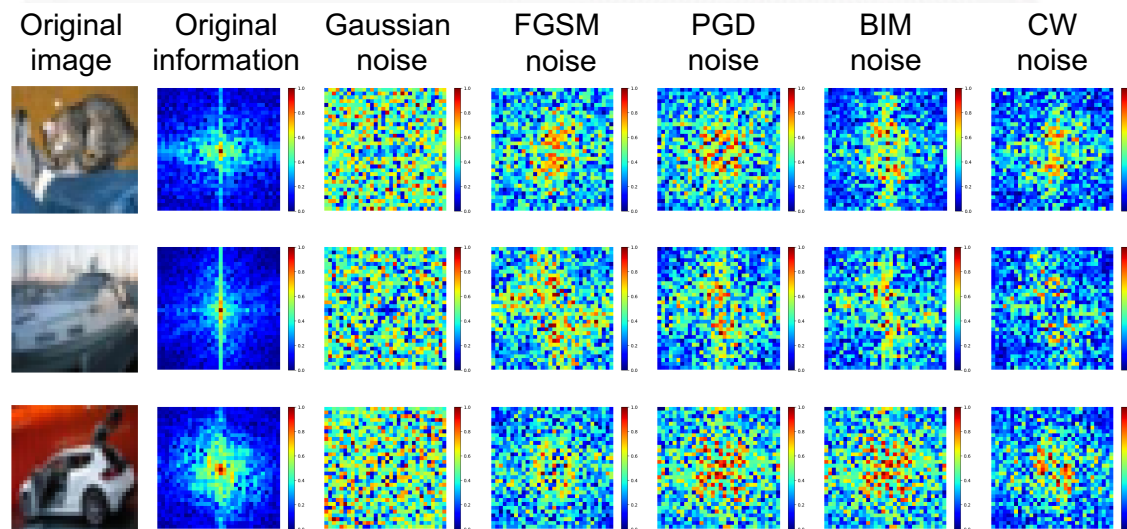


Figure 2. Visualization frequency spectrums for data observation.

Our Frequency Encoding (FE) module at time step T is defined as:

$$\tilde{x}_{r_i}^t \leftarrow \mathcal{F}^{-1} \left(\mathcal{M}_{r_i}^t \odot \mathcal{F}^t(x) \right), \quad i, t \in \{1, 2, \dots, T\}$$

where $\mathcal{F}(\cdot)$ is Discrete Fourier Transform, and \mathcal{M} is the frequency mask.

$$\mathcal{M}_{m,n} = \begin{cases} 1, & 0 \leq |m|, |n| \leq r \\ 0, & \text{else} \end{cases}$$

Evolutionary leak factor

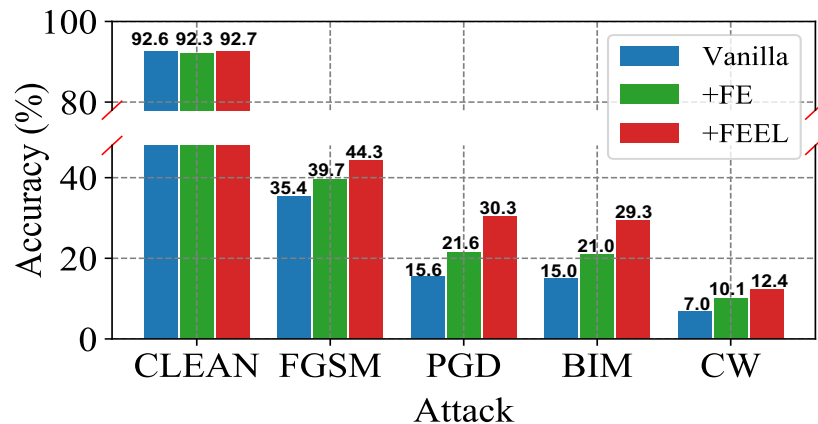
Leveraging the frequency-encoded input, we assign trainable leak factors to different neurons within a layer across time steps to mitigate the propagation of noise information.

$$\lambda_i^t = \lambda_i^t - \eta \Delta \lambda_i^t$$

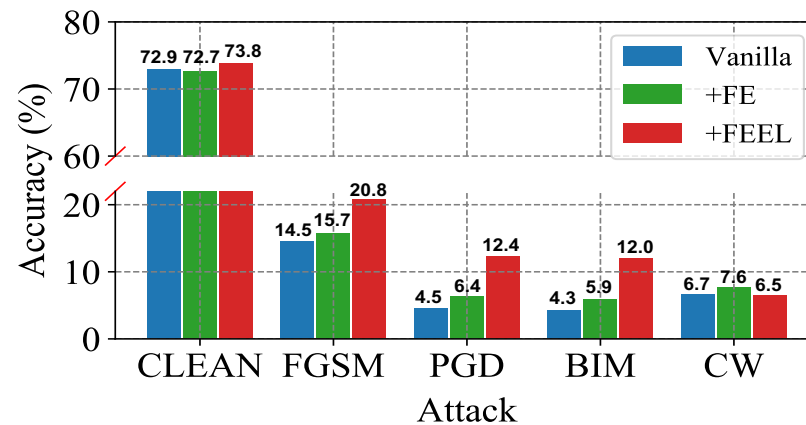
$$\Delta \lambda_i^t = \frac{\partial \mathcal{L}}{\partial \lambda_i^t} = \frac{\partial \mathcal{L}}{\partial O_i^t} \cdot \frac{\partial O_i^t}{\partial u_i^t} \cdot \frac{\partial u_i^t}{\partial \lambda_i^t} = \frac{\partial \mathcal{L}}{\partial O_i^t} \cdot \frac{\partial O_i^t}{\partial u_i^t} \cdot u_i^{t-1}$$

$$\mathcal{L} = \mathcal{L}_{CE}(x, y, W, \lambda)$$

Compare with Vanilla

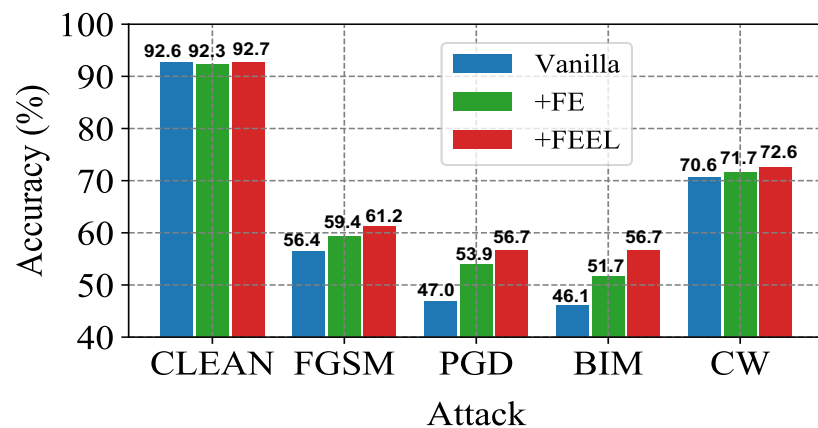


(a) CIFAR10, VGG11, T=4

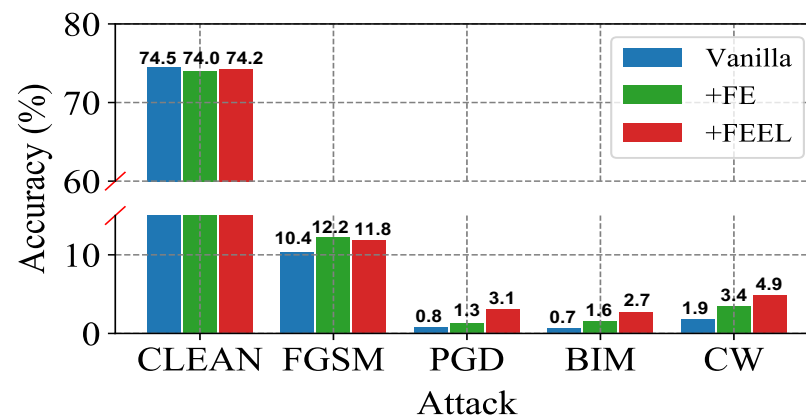


(b) CIFAR100, VGG11, T=4

Figure 4. Performance of the proposed FE and FEEL under different white-box attacks.



(a) CIFAR10, VGG11, T=4



(b) Tiny-ImageNet, resNet19, T=4

Figure 5. Performance of the proposed FE and FEEL under different black-box attacks.

■ Compare with SOTA

Table 1: Performance of the proposed FE and FEEL with different training strategies. The perturbation $\epsilon = 8/255$ for all attacks, and iterative step $k = 7$, step size $\alpha = 0.01$ for PGD, BIM. The dataset is CIFAR100 with $T = 8$, the network is VGG11. The improvement brought by our method is shown in parentheses.

Methods	clean	GN	FGSM	PGD	BIM	CW
Vanilla	72.93	68.93	4.91	0.16	0.14	6.53
Vanilla+FE (Ours)	72.67 (-0.26)	69.40 (+0.47)	5.18 (+0.27)	0.31 (+0.15)	0.24 (+0.10)	7.63 (+1.10)
Vanilla+FEEL (Ours)	73.79 (+0.86)	68.05(-0.88)	9.60 (+4.69)	2.04 (+1.88)	1.81 (+1.57)	6.66 (+0.13)
AT [12]	69.14	68.27	17.21	8.63	8.13	16.54
AT+FE (Ours)	69.34 (+0.20)	68.67 (+0.40)	17.65 (+0.44)	8.92 (+0.29)	8.33 (+0.20)	21.49 (+4.95)
AT+FEEL (Ours)	69.79 (+0.65)	69.02 (+0.75)	18.67 (+1.46)	11.07 (+2.44)	10.56 (+2.43)	21.78 (+5.24)
RAT [8]	70.03	69.26	18.88	8.87	7.93	20.79
RAT+FE (Ours)	69.74 (-0.29)	68.35 (-0.91)	18.74 (-0.14)	9.70 (+0.83)	8.91 (+0.98)	27.16 (+6.37)
RAT+FEEL (Ours)	69.80 (-0.23)	68.46 (-0.80)	19.08 (+0.20)	12.36 (+3.49)	11.96 (+4.03)	25.52 (+4.73)
StoG [10]	72.22	61.63	5.92	0.26	0.20	19.87
StoG+FE (Ours)	73.13 (+0.91)	67.65 (+6.02)	6.95 (+1.03)	0.22 (-0.04)	0.25 (+0.05)	23.02 (+3.15)
StoG+FEEL (Ours)	72.13 (-0.09)	65.96 (+4.33)	9.15 (+3.23)	0.55 (+0.29)	0.31 (+0.11)	24.79 (+4.92)
AT+StoG	69.24	63.35	19.64	9.77	3.23	44.79
AT+StoG+FE (Ours)	69.45 (+0.21)	68.83 (+5.48)	20.06 (+0.42)	10.69 (+0.92)	3.24 (+0.01)	38.56 (-6.23)
AT+StoG+FEEL (Ours)	69.53 (+0.29)	68.47 (+5.12)	18.27 (-1.37)	11.52 (+1.75)	3.90 (+0.67)	45.18 (+0.39)
RAT+StoG	69.12	68.37	29.25	15.43	6.91	32.08
RAT+StoG+FE (Ours)	68.97 (-0.15)	68.52 (+0.15)	31.65 (+2.40)	17.49 (+2.06)	8.57 (+1.66)	47.16 (+15.08)
RAT+StoG+FEEL (Ours)	69.97 (+0.85)	68.15 (-0.22)	31.68 (+2.43)	18.07 (+2.64)	8.89 (+1.98)	50.56 (+18.48)

FEEL-SNN: Robust Spiking Neural Networks with Frequency Encoding and Evolutionary Leak Factor

Mengting Xu^{1,2}, De Ma^{1,2}, Huajin Tang^{1,2},
Qian Zheng^{1,2}*, Gang Pan^{1,2}*

¹ The State Key Lab of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China

² College of Computer Science and Technology, Zhejiang University, Hangzhou, China

THANKS FOR LISTENING!