Exploring Temporal Information Dynamics in Spiking Neural Networks

Youngeun Kim[1], Yuhang Li[1], Hyoungseob Park[1], Yeshwanth Venkatesha[1], Anna Hambitzer[2], Priyadarshini Panda[1]

[1] Department of Electrical Engineering, Yale University, New Haven, CT, USA
[2] Technology Innovation Institute, Abu Dhabi, UAE
{youngeun.kim, yuhang.li, hyoungseob.park, yeshwanth.venkatesha, priya.panda}@yale.edu, Anna.Hambitzer@tii.ae

Abstract

Most existing Spiking Neural Network (SNN) works state that SNNs may utilize temporal information dynamics of spikes. However, an explicit analysis of temporal information dynamics is still missing. In this paper, we ask several important questions for providing a fundamental understanding of SNNs: *What are temporal information dynamics inside SNNs? How can we measure the temporal information dynamics? How do the temporal information dynamics affect the overall learning performance?* To answer these questions, we estimate the Fisher Information of the weights to measure the distribution of temporal information during training in an empirical manner. Surprisingly, as training goes on, Fisher information starts to concentrate in the early timesteps. After training, we observe that information becomes highly concentrated in earlier few timesteps, a phenomenon we refer to as *temporal information concentration*. We observe that the temporal information concentration phenomenon is a common learning feature of SNNs by conducting extensive experiments on various configurations such as architecture, dataset, optimization strategy, time constant, and timesteps. Furthermore, to reveal how temporal information concentration affects the performance of SNNs, we design a loss function to change the trend of temporal information. We find that temporal information concentration is crucial to building a robust SNN but has little effect on classification accuracy. Finally, we propose an efficient iterative pruning method based on our observation on temporal information concentration. Code is available at https://github.com/Intelligent-Computing-Lab-Yale/Exploring-Temporal-Information-Dynamics-in-Spiking-Neural-Networks.

Introduction

Within the last decade, Spiking Neural Networks (SNNs) have received huge attention as a low-power alternative to Artificial Neural Networks (ANNs) (Roy, Jaiswal, and Panda 2019; Christensen et al. 2022; Lobo et al. 2020; Wang, Lin, and Dang 2020). SNNs process visual information in an event-driven manner using sparse binary spikes over multiple timesteps that make them an attractive option for low-power neuromorphic hardware implementation (Akopyan et al. 2015; Davies et al. 2018; Furber et al. 2014). Recent years have witnessed a surge in SNN algorithmic works that aim to improve accuracy of SNNs on standard image recognition tasks while maintaining higher efficiency than ANNs (Wu et al. 2019; Comsa et al. 2020; Mostafa 2017; Li et al. 2021a; Deng et al. 2022).

Although most of the existing works on SNN assert that they might improve (or leverage) the temporal dynamics of spikes (Wu et al. 2018; Fang et al. 2021b; Kim and Panda 2020; Masquelier, Albantakis, and Deco 2011; Neftci, Mostafa, and Zenke 2019), yet, an explicit analysis of temporal information dynamics is still missing. In this paper, we ask several important questions for understanding such fundamental characteristics of SNNs: *What are temporal information dynamics inside SNNs? How can we measure the temporal information dynamics? How do the temporal information dynamics affect the overall learning performance?* Understanding the dynamics will enable us to apprehend the learning representations inside SNNs which may help to develop better temporal training algorithms, find new use-cases for SNNs for conventional AI (and possibly computational neuroscience) applications, and also explore new theoretical directions for SNN research.

To this end, we present a first-of-its-kind study to understand temporal information dynamics in SNNs through the lens of Fisher information. In this study, we select a ResNet model with Leak-Integrated-and-Fire spiking neuron as a baseline, which is widely used in classification task (Zheng et al. 2020; Li et al. 2021b; Fang et al. 2021a). We measure the Fisher information of an SNN at each timestep, where we find the temporal information distribution varies as training goes on. Specifically, we find that information in an SNN shifts from latter timesteps to earlier timesteps as training progresses. We call this phenomenon as *temporal information concentration (TIC)*. This is a novel observation, and therefore, we investigate whether the TIC phenomenon is a function of specific training variables such as architecture, dataset, and optimization strategy. Through extensive experiments, we found that TIC is a common characteristic of SNNs during training.

Furthermore, we explore the impact of TIC on the performance of networks (*i.e.*, accuracy and robustness). To this end, we design a loss function that can manipulate the Fisher information at each timestep. The results show that TIC significantly affects the robustness of SNN against both adversarial perturbations and standard input noise (such as

Gaussian and Blur). Unlike robustness, changes in temporal information due to TIC manipulation do not show any meaningful effect on classification accuracy. Finally, we propose an efficient iterative pruning strategy for SNNs using TIC, where we found the pruning performance is almost preserved with less number of timesteps.

In summary, our key contributions are as follows: (1) For the first time, we quantitatively analyze the temporal dynamics of SNNs. (2) Using Fisher information, we find that temporal information concentration (TIC) is a general trend in SNNs during training. (3) By designing a loss forcing SNNs to have a different trend of TIC, we find that TIC plays a crucial role in imparting robustness to SNNs. (4) Finally, we apply the TIC observation to propose an efficient iterative pruning method for SNNs.

Modeling Spiking Neural Networks

In this section, we briefly introduce the neuron type and input encoding technique used in our analysis. We use Spiking Neural Networks (SNNs) with discretized Leaky Integrate-and-Fire (LIF) neurons (Roy, Jaiswal, and Panda 2019; Wu et al. 2019; Fang et al. 2021b; Kim et al. 2022b) using simulation step $dt = 1$, which can be formulated by

$$U_l^t = (1 - \frac{1}{\tau})U_l^{t-1} + \frac{1}{\tau}W_l O_{l-1}^t, \tag{1}$$

where U_l^t denotes membrane potential at time-step t for layer l, and O_{l-1}^t stands for the spike output from the previous layer. Also, W_l represents weight connections at layer l, and τ is a time constant for decaying the membrane potential. Note, capital letters (*e.g.*, U_l^t and O_{l-1}^t) represent matrices. The neuron generates a spike output if its membrane potential exceeds a firing threshold. Then, the membrane potential is reset to zero after firing. For training weight parameters, we use spatio-temporal back-propagation (STBP), which accumulates the gradients over all timesteps (Wu et al. 2018; Neftci, Mostafa, and Zenke 2019). We can formulate the gradients at the layer l by chain rule, given by

$$\Delta W_l = \frac{\partial L}{\partial W_l} = \sum_t (\frac{\partial L}{\partial O_l^t}\frac{\partial O_l^t}{\partial U_l^t} + \frac{\partial L}{\partial U_l^{t+1}}\frac{\partial U_l^{t+1}}{\partial U_l^t})\frac{\partial U_l^t}{\partial W_l}. \tag{2}$$

While computing backward gradients, we use an approximate piece-wise linear function gradient $\frac{1}{\pi}arctan(\pi x) + \frac{1}{2}$ in order to address non-differentiability of LIF neurons (Fang et al. 2021b). Overall, according to gradient descent, the network parameters are updated as $W_l = W_l - \eta\Delta W_l$, where η represents learning rate.

In this work, we focus on static image recognition where the majority of prior SNN works have focused so far (Roy, Jaiswal, and Panda 2019). We choose to generate spikes in an end-to-end manner, *i.e.*, directly encode the images in the first layer, due to its simplicity, flexibility, and high performance on large-scale datasets (Wu et al. 2019; Zheng et al. 2020; Zhang and Li 2020; Fang et al. 2021b; Lee et al. 2020; Li et al. 2021a; Kim et al. 2022a). A given image is shown for T timesteps $\{1, 2, \ldots, T\}$ to an SNN, and the final prediction is computed by accumulating the spikes at the output layer across T steps.

Fisher Information Analysis in Time Dimension

The Fisher Information Matrix (FIM) quantifies the amount of information inside a model obtained from a given data, when the model parameters are perturbed (Fisher 1925). If weight perturbation brings small (or large) change in the model's prediction, we can say that the model contains small (or large) information with respect to the corresponding data. FIM can be interpreted as the second-order gradient of KL divergence between the original model and the weight-perturbed model (Achille, Rovere, and Soatto 2018). Mathematically, given a network's approximate posterior distribution $f_\theta(y|x)$ with weight parameters θ, input image x sampled from data distribution D, output variable y, FIM can be formulated as follows:

$$M = \mathbb{E}_{x\sim D, y\sim f_\theta(y|x)}[\nabla_\theta log f_\theta(y|x) \nabla_\theta log f_\theta(y|x)^T]. \tag{3}$$

Existing works have utilized FIM to explore the characteristic of the loss landscape (Keskar et al. 2016; Liang et al. 2019; Soen and Sun 2021), model capacity (Ly et al. 2017), and model training dynamics (Achille, Rovere, and Soatto 2018).

Different from the conventional ANN model, SNN predicts class probabilities by accumulating a given input through multiple timesteps. Our objective is to analyze the information dynamics of the model across time. Therefore, we introduce a metric to measure the amount of accumulated FIM in SNN from timestep 1 to timestep t:

$$M_t = \mathbb{E}[\nabla_\theta log f_\theta(y|x_{i\leq t}) \nabla_\theta log f_\theta(y|x_{i\leq t})^T], \tag{4}$$

where, $i \in \{1, 2, ..., T\}$ is a positive integer that represents the index of timestep. Note that it is difficult to measure the amount of information in one timestep independently since the posterior distribution of SNNs is based on the information from all previous timesteps with LIF neurons. One major problem of FIM on deep neural networks is the size of the matrix, which is too large to compute completely. To address this, following previous works (Achille, Rovere, and Soatto 2018; Kirkpatrick et al. 2017), we use the trace of FIM to measure the accumulated information I_t stored in weight parameters from timestep 1 to t:

$$I_t = Tr(M_t) = \mathbb{E}_{x\sim D, y\sim f_\theta(y|x_{i\leq t})}[\|\nabla_\theta log f_\theta(y|x_{i\leq t})\|^2]. \tag{5}$$

Given N training samples, the expectation in Eq. 5 can be replaced by the empirical mean across observed data (Amari, Park, and Fukumizu 2000; Martens and Grosse 2015; Karakida, Akaho, and Amari 2019):

$$I_t = \frac{1}{N}\sum_{n=1}^{N} \|\nabla_\theta log f_\theta(y|x_{i\leq t}^n)\|^2. \tag{6}$$

Across all experiments, we use Eq. 6 for quantifying the amount of Fisher information (or sometimes we use *information* interchangeably) across different timesteps.

In ANN literature, the empirical FIM trace (Eq. 6) can be interpreted as a measure for the importance of weight connections (Achille, Rovere, and Soatto 2018; Kirkpatrick et al. 2017) with respect to training data. For example, Kirkpatrick *et al*. (Kirkpatrick et al. 2017) address catastrophic forgetting in continual learning (*i.e.*, a network learns different tasks

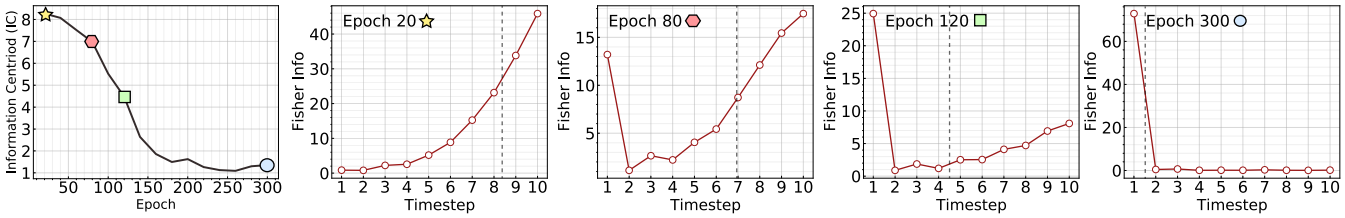


Figure 1: Illustration of temporal dynamics across training epochs. Starting from the left panel, we present information centroid (Eq. 7) across epochs. We also report Fisher information in temporal dimension (Eq. 6) at four different epochs (marked with different colors). As training goes on, the early timesteps obtain more information, while the information decreases in the latter timesteps. Accordingly, the information centroid (vertical dotted line) moves towards early timesteps as training progresses.

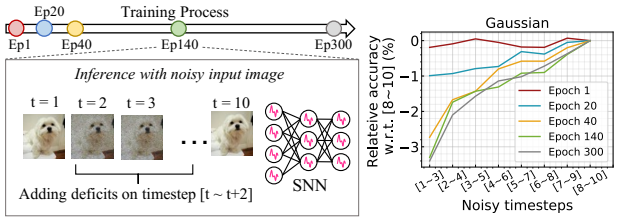


Figure 2: Additional experiments support TIC observation at inference. We use a ResNet19 architecture on the CIFAR10 dataset. We select five different SNN models during training (with clean data): The early training phase (epoch 1, 20), the intermediate phase (epoch 40, 140), and the final trained model (epoch 300). Then, we measure the test accuracy of the model at each point by adding deficits (*i.e.*, Gaussian noise) to the input image for time window $[t \sim t+2]$. We change t from 1 to 8. In the right panels, we report the relative test accuracy w.r.t. noise in $[8 \sim 10]$, *i.e.*, $Acc_{[t\sim t+2]} - Acc_{[8\sim 10]}$. We provide experimental details in Appendix.

sequentially) by maintaining connections having high FIM trace from the prior tasks. Achille *et al*. (Achille, Rovere, and Soatto 2018) use empirical FIM trace in order to discover the important training epochs for standard deep neural networks, where they show that epochs having a higher FIM trace impact more on the accuracy. Similarly, in our work, the SNN model having a high FIM trace at timestep t represents that the weight connections inside the SNN contain important information for the given input until timestep t.

Temporal information concentration in SNN. To reveal the temporal information dynamics in SNNs, we first present temporal Fisher information of SNNs across training epochs. In Fig. 1, we visualize temporal Fisher information of a ResNet19 architecture on CIFAR10 dataset. Interestingly, as training goes on, the amount of Fisher information in the latter timesteps steadily decreases while the early-time information increases. In the final trained SNN model (epoch 300), Fisher information concentrates on the first few timesteps, and maintains a near-zero value till the end of timesteps. We call this phenomenon as *temporal information concentration (TIC)* - an information shift from latter timesteps to the early timesteps as training progresses.

Further, to provide better visualization of the overall trend of information dynamics across epochs, we introduce a metric called *Information Centroid* (IC). Given SNNs with T timesteps, IC can be formulated by

$$IC = \frac{\sum_{t\in\{1,\dots,T\}} tI_t}{\sum_{t\in\{1,\dots,T\}} I_t}. \tag{7}$$

Thus, high IC means Fisher information (I_t) increases with timestep t (*e.g.*, epoch 20 in Fig. 1). In Fig. 1, SNN at epoch 20 yields highest IC value which can be understood from its increasing I_t trend across different timesteps. Small IC denotes that information concentrates in the early timesteps (*e.g.*, epoch 300 in Fig. 1 where I_t is highest at $t = 1$ and becomes nearly 0 for $t = 2, .., 10$).

Moreover, we found that TIC is closely related to performance degradation with deficits at inference. As shown in Fig. 2, we select models from five different epochs, and add the deficits to the input image for a certain time window. If the accuracy degrades significantly in a specific time window, those timesteps are likely to convey critical information for prediction (Achille, Rovere, and Soatto 2018). At the beginning of the training phase (or early epochs), all timesteps show similar noise sensitivity. However, interestingly, as training goes on, early timesteps show higher noise sensitivity compared to the later timesteps. The results support our observation on TIC, where the early timesteps contain important information as training evolves.

Layer-wise analysis with Fisher Information. Previous works (Kirkpatrick et al. 2017; Achille, Rovere, and Soatto 2018) use Fisher information to measure the importance of layers (or weight connections). If a layer contains high Fisher information, the layer has a high contribution to the prediction for the given data. Here, we conduct layer-wise Fisher information analysis to observe the importance of each layer across timesteps. To this end, we measure Fisher information contained in each residual block of a ResNet19 SNN model (Fig. 3). The overall trend of temporal information dynamics in each layer follows the TIC trend. Intriguingly, we find that, after the SNN model is trained for several epochs (*e.g.*, epoch 120 and 300), shallow and deep layers contain high Fisher information at the latter and early timesteps, respectively. Such observation can be interpreted through the role of shallow and deep layers in feature representation. The prior ANN and SNN works (Selvaraju et al. 2017; Kim and Panda 2021) have revealed that shallow and deep layers of a model contain low-level (*e.g.*, texture) and high-level (*e.g.*, semantic) features, respectively. Thus, high Fisher information in shallow/deep layers means that weight connections related to

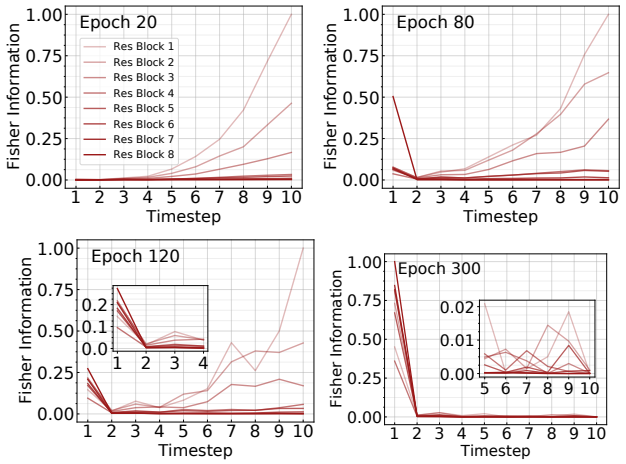


Figure 3: Normalized Fisher information contained in each residual block (in ResNet19 trained on CIFAR10) as a function of timestep. We visualize Fisher information at Epoch 20, 120 and 300. Here, we provide the normalized Fisher information for better visualization on relative Fisher information across all layers.

low/high-level features are important for prediction. Accordingly, we conclude that the SNN model focuses on high-level features at the early timesteps, while low-level features are important at the latter timesteps for prediction.

Is temporal information concentration trend always shown in SNNs? Having observed that TIC emerges in SNNs, we next study whether such phenomenon can be varied by various configurations such as different timesteps, time constant, learning rate, dataset and different architectures. We visualize the change of Information Centroid (IC) across epochs in Fig. 4. The default setting for all experiments is as follows: timestep 10, time constant 2, SGD optimizer with learning rate 3e-1, weight decay 5e-4, CIFAR10 dataset, and ResNet19 architecture.

Timestep: We first train SNN on CIFAR10 with timesteps 4, 6, 8, and 10. We observe that all timestep configurations show a similar trend, *i.e.*, decreasing IC values as training goes on. Note that longer timesteps have a higher initial IC value at the beginning of training. One interesting observation is that a longer timestep starts late IC transition from high to low. For example, IC value of t=4 (green curve) starts to drop fast at the very early epochs, however, t=10 (red curve) shows IC transition near epoch 80. This implies that a longer timestep contains more information, thus requires more training epochs to concentrate them in early timesteps.

Time constant: As shown in Eq. 1, time constant τ controls the decaying intensity of the membrane potential of an LIF neuron. A higher time constant means that the LIF neuron relies more on the previous information rather than the current input. To evaluate the impact of time constant τ on TIC behavior, we train an SNN model with different time constants 0.75, 2, and 4. The results show that the smaller time constants ($\tau = 0.75$ and $\tau = 2$) achieve low IC in the middle of training (see the yellow curves in the right panels of Fig. 4(b)). On the other hand, the higher time constants ($\tau = 4$) shows

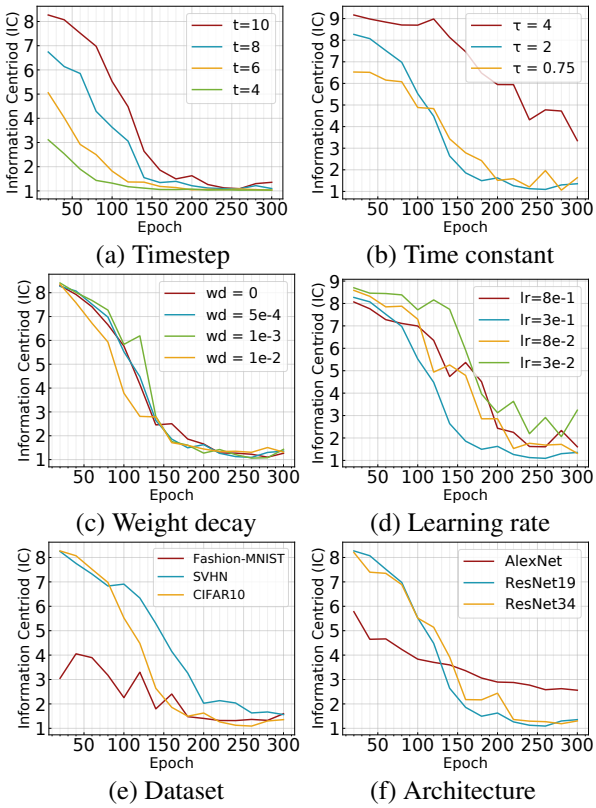


Figure 4: Information Centriod (IC) change with different factors. We observe that most cases show TIC at the later epochs, where the early timesteps show higher Fisher information.

relatively higher IC compared to the others. This is because, with high time constant, information is slowly propagated through layers, thus the information concentration begins in late timesteps (see the red curves in the right panels).

Weight decay and Learning rate: We also analyze how learning rate affects TIC. To this end, we use four different learning rate (lr) configurations: optimal lr (1e-3), lower lr (3e-2 and 8e-2), and higher lr (8e-1). Compared to an optimal lr setting (blue curve), a lower lr (green and yellow curves) shows late IC transition from late timesteps to early timesteps. Similarly, a larger lr (red curve) also does not show quick IC transition. We also illustrate temporal Fisher information change for lr=3e-2 and lr=3e-1 in the right panels of Fig. 4(d). This results suggest that the early IC transition can be an indicator for choosing a proper learning rate in training process. On the other hand, in Fig. 4(c), we further conduct weight decay analysis, which shows weight decay does not make significant change on the temporal concentration behavior in SNN.

Dataset and Architecture: In Fig. 4(e) and Fig. 4(f), we show IC transition with different datasets and architectures. For dataset ablation studies, we use CIFAR10 (Krizhevsky and Hinton 2009) and SVHN (Netzer et al. 2011) that contain natural RGB images, as well as gray-scale Fashion-MNIST dataset (Xiao, Rasul, and Vollgraf 2017). All datasets show IC transition from late timesteps to early timesteps across training epochs. Furthermore, we analyze on CNN architectures

without skip connections (AlexNet (Krizhevsky, Sutskever, and Hinton 2012)), with skip connections (ResNet19 (He et al. 2016)), and deeper architecture (ResNet34 (He et al. 2016)) with skip connections. AlexNet shows quick IC transition in the early epochs, however, it does not show very low IC value at the end of epochs. This phenomenon also can be shown with temporal Fisher information visualization (right panels, Fig. 4(f)) where AlexNet shows slow Fisher information decrease across time at epoch 300. Different from AlexNet, ResNet34 shows quick information concentration in timesteps 1$\sim$3. This suggests that AlexNet capacity is limited compared to ResNet34, therefore they require more timesteps to concentrate information.

Overall, we conclude that TIC is a common characteristic of SNNs, but IC transition speed can be changed according to various factors such as optimization settings, architecture, and datasets.

Analysis on TIC: Robustness Perspective

While the previous section shows temporal information concentration (TIC) is widely shared in SNNs, a key open question remains: *what is the role of temporal concentration in SNN? does it bring higher accuracy or better robustness?* To find the answer, in this section, we force the SNN to have specific Fisher information trend across timesteps. In this way, we can investigate the characteristics and role of Fisher information in timesteps.

To this end, we design a loss function to control the Fisher information trend in SNN. Before designing the loss, we first define the approximated relation between a loss function and Fisher information, as shown in the following definition.

Definition 1. *The log posterior $log f_\theta(y|x_{i\leq t})$ can be represented as a loss function $L_t(\theta)$, e.g., cross-entropy loss, where the final layer's outputs are accumulated t timesteps before they are converted to probabilities (e.g., with a softmax layer). Thus, we can rewrite Eq. 6 as:*

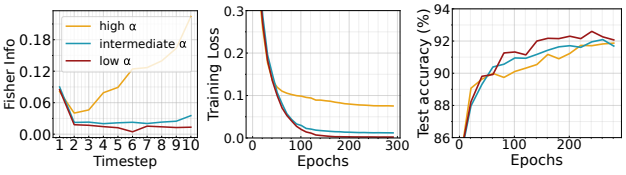
$$I_t(\theta) = \mathbb{E}[\|\nabla L_t(\theta)\|^2]. \tag{8}$$

Here, decreasing loss $L_t(\theta)$ will bring Fisher information $I_t(\theta)$ degradation because gradient $\nabla L_t(\theta)$ goes smaller as the loss converges to local minima. On the other hand, if we disturb the model to converge a loss value, Fisher information cannot become smaller. According to the aforementioned hypothesis, we control the Fisher information value at each timestep, by manipulating the loss function during training. Specifically, we force the loss function to have a value around α:

$$L_t(\theta, \alpha) = |L_t(\theta) - \alpha|. \tag{9}$$

The above equation represents that, if the loss function goes below α, weights are updated with gradient ascent, otherwise a standard gradient descent is applied. Here, adding or subtracting α does not affect gradients for weight parameters. We apply Eq. 9 across T timesteps in order to make sure that Fisher information shows a similar trend for all timesteps, which can be formulated as:

$$L(\theta, \alpha) = \frac{1}{T}\sum_{t=1}^{T} L_t(\theta, \alpha). \tag{10}$$



(a) (b) (c)

Figure 5: (a) Fisher information across all layers after training (b) Training loss (c) Test accuracy of three α configurations. We train ResNet19 architecture on CIFAR10 dataset. We observe that Fisher information successfully changes according to the magnitude of α. Also, (b) and (c) show our loss function provides stable convergence. For loss visualization, we show the value of original loss $L_t(\theta)$ in Eq. 9. We present the SVHN and CIFAR100 results in the Appendix.

By changing α value, we can approximately control *relative* magnitude of Fisher information. We find that our loss $L(\theta, \alpha)$ can successfully control Fisher information, *i.e.*, smaller α shows less Fisher information across timesteps, as shown in Fig. 5. In our experiments, three types of SNNs trained with different α values are investigated: α_{low}, $\alpha_{intermediate}$, and α_{high} ($\alpha_{low} < \alpha_{intermediate} < \alpha_{high}$). We select $\alpha_{cifar10}$=[1e-3, 1e-2, 7e-2], α_{svhn}=[1e-4, 1e-2, 7e-2], $\alpha_{cifar100}$=[1e-4, 1e-3, 1e-2], for [α_{low}, $\alpha_{intermediate}$, α_{high}]. The α hyperparameter selection is based on dataset complexity where they have different sensitivity w.r.t to α.

Compared to α_{low}, using $\alpha_{intermediate}$ forces SNN to slightly increase the amount of Fisher information. α_{high} further forces the model to have high Fisher Information, therefore the Fisher information increases as time goes on. By comparing these configurations, we can reveal what is the advantage if Fisher information becomes smaller through time *i.e.*, TIC. Note, the loss function (Eq. 10) applied for Fig. 5 and Table 1 is different from the other experiments where we only apply CrossEntropy loss on the accumulated spike activation in the last layer. Our objective here is to manipulate Fisher information to explore their impact on the robustness of an SNN model.

Classification Accuracy. We first measure the accuracy of SNNs trained with three different α (with the same number of epochs) on CIFAR10, CIFAR100 (Krizhevsky and Hinton 2009), and SVHN (Netzer et al. 2011), and report the results in Table 1. We use ResNet19 as a baseline architecture. All α configurations achieve similar accuracy across all datasets, regardless of dataset complexity. This implies that TIC is not an essential factor for SNNs to obtain high accuracy. In fact, in the TIC ablation experiments in Fig. 4, we found that IC transition speed changes based on different configurations, but, there was no conspicuous effect on accuracy. Table 1 results on the relation between TIC and accuracy further corroborate Fig. 4 (except for architecture and data experiments) results, where we achieve almost similar accuracy for all configurations.

Robustness against Noise / Adversarial sample. We analyze the relation between robustness and TIC. Specifically, we first corrupt the input image using two types of noise: Gaussian noise and Blur. For Gaussian noise experiments,

Method	Dataset	Clean Acc. (%)	Gaussian Noise	Blur	FGSM	PGD
low α	SVHN	96.03	93.48 (-2.55)	95.56 (-0.47)	91.69 (-4.34)	90.87 (-5.16)
inter. α	SVHN	96.01	93.24 (-2.77)	95.56 (-0.56)	77.93 (-18.08)	49.84 (-46.17)
high α	SVHN	95.91	92.85 (-3.06)	95.12 (-0.79)	55.39 (-40.52)	4.46 (-91.45)
low α	CIFAR10	92.04	69.01 (-23.03)	58.11 (-33.93)	77.22 (-14.82)	74.63 (-17.41)
inter. α	CIFAR10	91.89	68.09 (-23.80)	56.88 (-35.01)	69.29 (-22.60)	58.65 (-33.24)
high α	CIFAR10	91.87	61.01 (-30.86)	54.55 (-37.32)	53.50 (-38.37)	32.58 (-59.29)
low α	CIFAR100	68.17	37.60 (-30.57)	51.18 (-16.99)	44.62 (-23.55)	38.64 (-29.52)
inter. α	CIFAR100	68.47	36.98 (-31.49)	51.02 (-17.45)	43.39 (-25.08)	35.03 (-33.43)
high α	CIFAR100	67.95	35.56 (-32.39)	49.09 (-18.86)	38.36 (-29.59)	30.33 (-37.62)

Table 1: Classification accuracy and robustness of SNNs trained with three different α. We train ResNet19 architecture on three public datasets including SVHN, CIFAR10, CIFAR100. For robustness experiments, we report both accuracy and relative accuracy drop w.r.t. clean accuracy.

we add Gaussian noise whose l_2 norm is set to 50% of the norm of the given input. For generating Blur effect to the image, we follow the method used in (Achille, Rovere, and Soatto 2018) - images are downsampled to smaller resolution (*e.g.*, 16×16) and then upsampled back to its original resolution (*e.g.*, 32×32) with bilinear interpolation, which is an efficient implementation for destroying small-scale details in images. To further evaluate the robustness, we also evaluate the robustness of SNN models against two representative adversarial attacks. FGSM attack (Goodfellow et al. 2014) is a single-step attack based on backward gradients where the noise is the sign of gradients scaled by ϵ. Projected Gradient Descent (PGD) attack (Madry et al. 2017) is an iterative adversarial attack characterized by three parameters- maximum perturbation ϵ, perturbation step size α and the number of iterations n. In our experiments, we use $\epsilon = \frac{8}{255}$ for FGSM attack, and [$\epsilon = \frac{8}{255}$, $\alpha = \frac{4}{255}$, $n = 10$] for PGD attack.

In Table 1, we report the corrupted accuracy with respect to each noise. Interestingly, for all noise configuration, SNNs trained with low α show less performance drop compared to SNNs trained with high α. This difference is especially huge for adversarial attacks. This results can be explained by connecting the KL divergence to Fisher information matrix (FIM). Given that a small perturbation δ is added to input x, it will make a difference in the probability $f_\theta(y|x_t + \delta)$. Then, we can measure the difference in the output probabilities using KL divergence, where we can apply second-order Taylor approximation as

$$D_{KL}(f_\theta(y|x_t)||f_\theta(y|x_t + \delta)) \approx \frac{1}{2}\delta^T M_t \delta. \qquad (11)$$

The above equation shows that the output perturbation from the given input noise can be represented by the function of FIM. If the eigen values of the FIM have larger values, the output perturbation is likely to be severe. Thus, the trace of FIM (*i.e.*, sum of eigen values) should be small in order to suppress the noise. Therefore, the SNN model that shows temporal concentration behavior (smaller Fisher trace as time goes on) might have better robustness.

Application: Efficient SNN Pruning Using TIC

We further explore the usage of TIC for applications. Here, we apply a standard iterative magnitude-based pruning (Han et al. 2015) to SNNs and propose an efficient pruning process using TIC. Basically, as shown in Fig. 6(a), the original pruning method (Han et al. 2015) starts with training SNNs using timestep T. After training, $p\%$ of low-magnitude weight connections are pruned by thresholding, followed by retraining the remaining parameters. It was found that the iterative pruning strategy (*i.e.*, multiple pruning-retraining cycles) brings better pruning results. Following this, in our experiments, we prune 50% of connections for 5 cycles. As such iterative pruning causes non-trivial training time, we aim to reduce the computational cost using the TIC phenomenon in SNNs.

To bring efficiency, we focus on the observation that the information is concentrated in early timesteps after the first training stage. In Fig. 6(b), we illustrate the amount of Fisher

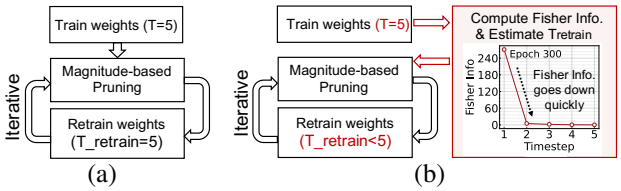


Figure 6: (a) Iterative pruning method (Han et al. 2015) for SNNs. Here, we assume the SNN are trained with timestep 5. (b) We propose the concept of efficient pruning using TIC (colored with red). For a retraining-pruning cycle, the SNN model is trained with a less number of timesteps.

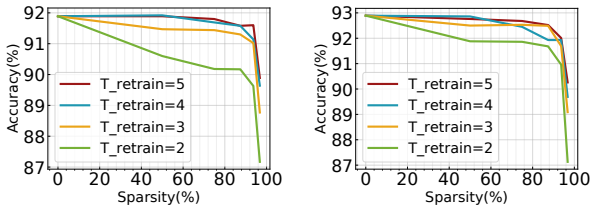


Figure 7: Accuracy vs. sparsity of (a) VGG16 (b) ResNet19 architectures on CIFAR10. In these experiments, we use timestep 5 for the first training stage, then we change $T_{retrain}$ in retraining-pruning cycles.

information w.r.t timesteps, which shows near-zero values after T=3. This allows us to use less number of timesteps $T_{retrain}$ for pruning-retraining cycles since we will not lose information with a smaller number of timesteps. In Fig. 7, we compute the accuracy across various sparsity levels with different $T_{retrain}$ values. For $T_{retrain}$ that shows near-zero Fisher information values (*i.e.*, $T_{retrain}$=[3, 4, 5]), we achieve similar pruning performance. On the other hand, $T_{retrain}$=2 shows huge performance degradation.

Using small $T_{retrain}$ brings energy-efficiency. Suppose we train the SNN model for N epochs in the first training round, and $N_{retrain}$ epochs for R rounds of the pruning-retraining process. With the original approach (Han et al. 2015), the computational cost is approximately proportional to $NT + N_{retrain}RT$, where T is the number of timesteps. If we apply $T_{retrain}$ for the pruning-retraining process based on TIC observation, the computational cost is approximately proportional to $NT + N_{retrain}RT_{retrain}$. Overall, the compute efficiency can be obtained as: $\frac{N_{retrain}R(T-T_{train})}{NT+N_{retrain}RT} \times 100$. In our experiment, we set $N = 300$, $N_{retrain} = 60$, $R = 5$, $T = 5$. In this case, if we apply $T_{retrain} = 3$, we will obtain 20% compute efficiency improvement.

Discussion and Conclusion

In our work, we observe temporal information concentration (TIC) in SNNs, an information shift from latter timesteps to the early timesteps as training progresses. This new observation enables us to understand the learning representations inside SNNs, providing several discussions on the connection to previous studies and future research directions for SNNs.

Connection with bio-plausibility of SNNs. TIC reveals some connections between SNNs and biological features observed in human brain. Firstly, TIC also can be founded in the human visual cortex. Previous neuroscience research (Stigliani, Jeska, and Grill-Spector 2017; Boynton et al. 1996) presented that the V1 activity responses to stimuli increase steeply at the early time and show little change afterward. Also, layer-wise Fisher analysis (Fig. 3) implies that the SNN model focuses on high-level features (*e.g.*, semantic meaning of the given image) at the early timesteps, while low-level features (*e.g.*, texture) are important at the latter timesteps for prediction. Such new observation in SNN is similar with primate vision where they focus on the semantic context of a given image at the early time, and then, look into finer details (Brady et al. 2009; Hollingworth and Henderson 2002).

Shrinking timesteps in SNN. Several works (Chowdhury, Rathi, and Roy 2021; Chowdhury, Garg, and Roy 2021) progressively reduce the number of timesteps as training goes on. Although those works significantly reduce the latency of SNN while almost maintaining accuracy, there is a lack of explanation why timestep shrinking is possible in SNN. Based on our observation, we can explain that SNNs can process most information in the early few timesteps, therefore achieving a high accuracy without the later timesteps. We hope our observation (*i.e.*, TIC, Fig. 1) leads to interpretation of other temporal-related techniques in SNNs such as temporal batch norm (Zheng et al. 2020; Kim and Panda 2020) and learnable leak factor (Fang et al. 2021b).

Connection between model capacity and timesteps. In ANN literature, it is a well-known trend that a larger model usually achieves better accuracy (Alzubaidi et al. 2021; Dosovitskiy et al. 2020). Similarly, SNN works (Hu, Tang, and Pan 2018; Fang et al. 2021a) also present that larger architectures achieve a better performance, showing that model capacity plays an important role in SNNs. As our work focuses on understanding the relation between temporal dynamics and model learning, we conjecture that model capacity might affect the number of timesteps required for stable training. Such hypothesis can be supported by observations in Fig. 4(f), where high-capacity model (ResNet19) shows quick TIC compared to low-capacity model (AlexNet). Thus, the results imply high-capacity model can concentrate information within short timesteps. To empirically validate this, as pilot experiments, we measured the accuracy of ResNet19 and AlexNet trained with various timesteps. Interestingly, we found that, ResNet19 shows performance saturation at timestep 4, whereas AlexNet saturates at timestep 8, which supports our statement that high-capacity model can be trained with short timesteps. The results suggest that the trend of TIC can affect the minimum number of timesteps for performance saturation. We present experiment settings and results in Appendix.

Robustness of SNNs with respect to noise and adversarial attacks. Recent SNN works (Sharmin et al. 2020; Liang et al. 2021) highlight that SNNs have better robustness against adversarial and natural noise compared to standard ANNs. In our study, we further provide the analysis how temporal information dynamics affects robustness. As we show that early timesteps are important for inference (thus, vulnerable), one can devise a more efficient attack/defense algorithm for SNN by adding noise at the early few timesteps. We hope our study fosters future work understanding the robustness of SNNs.

Impact of data type, loss function, coding scheme, and learning algorithm. In our experiments, we study the most general case of an SNN model (cross-entropy loss, direct coding, backpropagation for training) used in state-of-the-art works on image classification task. We clarify that the trend TIC might be different with different SNN training configurations such as data type, loss function, coding scheme, and learning algorithm, which points out interesting future work to further understand SNNs. Here we ask several research questions: How information dynamics change with respect to sequential input such as DVS dataset (Calabrese et al. 2019; Li et al. 2017)? Will modifying cross-entropy loss in time axis help SNN training? Will different coding scheme (such as temporal coding (Mostafa 2017; Comsa et al. 2020)) show TIC trend? We hope our work will become a starting point for thinking about a new direction in temporal representation for SNNs.

Acknowledgements

This work was supported in part by C-BRIC, a JUMP center sponsored by DARPA and SRC, Google Research Scholar Award, the National Science Foundation (Grant#1947826), TII (Abu Dhabi) and the DARPA AI Exploration (AIE) program.

References

Achille, A.; Rovere, M.; and Soatto, S. 2018. Critical learning periods in deep networks. In *International Conference on Learning Representations*.

Akopyan, F.; Sawada, J.; Cassidy, A.; Alvarez-Icaza, R.; Arthur, J.; Merolla, P.; Imam, N.; Nakamura, Y.; Datta, P.; Nam, G.-J.; et al. 2015. Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip. *IEEE transactions on computer-aided design of integrated circuits and systems*, 34(10): 1537–1557.

Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; and Farhan, L. 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1): 1–74.

Amari, S.-i.; Park, H.; and Fukumizu, K. 2000. Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural computation*, 12(6): 1399–1409.

Boynton, G. M.; Engel, S. A.; Glover, G. H.; and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, 16(13): 4207–4221.

Brady, T. F.; Konkle, T.; Oliva, A.; and Alvarez, G. A. 2009. Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & integrative biology*, 2(1): 1–3.

Calabrese, E.; Taverni, G.; Awai Easthope, C.; Skriabine, S.; Corradi, F.; Longinotti, L.; Eng, K.; and Delbruck, T. 2019. DHP19: Dynamic vision sensor 3D human pose dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.

Chowdhury, S. S.; Garg, I.; and Roy, K. 2021. Spatio-Temporal Pruning and Quantization for Low-latency Spiking Neural Networks. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–9. IEEE.

Chowdhury, S. S.; Rathi, N.; and Roy, K. 2021. One timestep is all you need: Training spiking neural networks with ultra low latency. *arXiv preprint arXiv:2110.05929*.

Christensen, D. V.; Dittmann, R.; Linares-Barranco, B.; Sebastian, A.; Le Gallo, M.; Redaelli, A.; Slesazeck, S.; Mikolajick, T.; Spiga, S.; Menzel, S.; et al. 2022. 2022 roadmap on neuromorphic computing and engineering. *Neuromorphic Computing and Engineering*.

Comsa, I. M.; Fischbacher, T.; Potempa, K.; Gesmundo, A.; Versari, L.; and Alakuijala, J. 2020. Temporal coding in spiking neural networks with alpha synaptic function. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8529–8533. IEEE.

Davies, M.; Srinivasa, N.; Lin, T.-H.; Chinya, G.; Cao, Y.; Choday, S. H.; Dimou, G.; Joshi, P.; Imam, N.; Jain, S.; et al. 2018. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1): 82–99.

Deng, S.; Li, Y.; Zhang, S.; and Gu, S. 2022. Temporal Efficient Training of Spiking Neural Network via Gradient Re-weighting. *arXiv preprint arXiv:2202.11946*.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Fang, W.; Yu, Z.; Chen, Y.; Huang, T.; Masquelier, T.; and Tian, Y. 2021a. Deep Residual Learning in Spiking Neural Networks. *arXiv preprint arXiv:2102.04159*.

Fang, W.; Yu, Z.; Chen, Y.; Masquelier, T.; Huang, T.; and Tian, Y. 2021b. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2661–2671.

Fisher, R. A. 1925. Theory of statistical estimation. In *Mathematical proceedings of the Cambridge philosophical society*, volume 22, 700–725. Cambridge University Press.

Furber, S. B.; Galluppi, F.; Temple, S.; and Plana, L. A. 2014. The spinnaker project. *Proceedings of the IEEE*, 102(5): 652–665.

Goodfellow, I. J.; et al. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Han, S.; Pool, J.; Tran, J.; and Dally, W. 2015. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR 2016*, 770–778.

Hollingworth, A.; and Henderson, J. M. 2002. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1): 113.

Hu, Y.; Tang, H.; and Pan, G. 2018. Spiking Deep Residual Networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Karakida, R.; Akaho, S.; and Amari, S.-i. 2019. Universal statistics of fisher information in deep neural networks: Mean field approach. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 1032–1041. PMLR.

Keskar, N. S.; Mudigere, D.; Nocedal, J.; Smelyanskiy, M.; and Tang, P. T. P. 2016. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*.

Kim, Y.; Li, Y.; Park, H.; Venkatesha, Y.; and Panda, P. 2022a. Neural Architecture Search for Spiking Neural Networks. *arXiv preprint arXiv:2201.10355*.

Kim, Y.; Li, Y.; Park, H.; Venkatesha, Y.; Yin, R.; and Panda, P. 2022b. Exploring Lottery Ticket Hypothesis in Spiking Neural Networks. In *European Conference on Computer Vision*, 102–120. Springer.

Kim, Y.; and Panda, P. 2020. Revisiting batch normalization for training low-latency deep spiking neural networks from scratch. *Frontiers in neuroscience*, 1638.

Kim, Y.; and Panda, P. 2021. Visual explanations from spiking neural networks using inter-spike intervals. *Scientific reports*, 11(1): 1–14.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13): 3521–3526.

Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Lee, C.; Sarwar, S. S.; Panda, P.; Srinivasan, G.; and Roy, K. 2020. Enabling spike-based backpropagation for training deep neural network architectures. *Frontiers in Neuroscience*, 14.

Li, H.; Liu, H.; Ji, X.; Li, G.; and Shi, L. 2017. Cifar10-dvs: an event-stream dataset for object classification. *Frontiers in neuroscience*, 11: 309.

Li, Y.; Deng, S.; Dong, X.; Gong, R.; and Gu, S. 2021a. A Free Lunch From ANN: Towards Efficient, Accurate Spiking Neural Networks Calibration. *arXiv preprint arXiv:2106.06984*.

Li, Y.; Guo, Y.; Zhang, S.; Deng, S.; Hai, Y.; and Gu, S. 2021b. Differentiable Spike: Rethinking Gradient-Descent for Training Spiking Neural Networks. *Advances in Neural Information Processing Systems*, 34.

Liang, L.; Hu, X.; Deng, L.; Wu, Y.; Li, G.; Ding, Y.; Li, P.; and Xie, Y. 2021. Exploring adversarial attack in spiking neural networks with spike-compatible gradient. *IEEE transactions on neural networks and learning systems*.

Liang, T.; Poggio, T.; Rakhlin, A.; and Stokes, J. 2019. Fisher-rao metric, geometry, and complexity of neural networks. In *The 22nd international conference on artificial intelligence and statistics*, 888–896. PMLR.

Lobo, J. L.; Del Ser, J.; Bifet, A.; and Kasabov, N. 2020. Spiking neural networks and online learning: An overview and perspectives. *Neural Networks*, 121: 88–100.

Ly, A.; Marsman, M.; Verhagen, J.; Grasman, R. P.; and Wagenmakers, E.-J. 2017. A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80: 40–55.

Madry, A.; et al. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.

Martens, J.; and Grosse, R. 2015. Optimizing neural networks with kronecker-factored approximate curvature. In *International conference on machine learning*, 2408–2417. PMLR.

Masquelier, T.; Albantakis, L.; and Deco, G. 2011. The timing of vision–how neural processing links to different temporal dynamics. *Frontiers in psychology*, 2: 151.

Mostafa, H. 2017. Supervised learning based on temporal coding in spiking neural networks. *IEEE transactions on neural networks and learning systems*, 29(7): 3227–3235.

Neftci, E. O.; Mostafa, H.; and Zenke, F. 2019. Surrogate gradient learning in spiking neural networks. *IEEE Signal Processing Magazine*, 36: 61–63.

Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; and Ng, A. Y. 2011. Reading digits in natural images with unsupervised feature learning.

Roy, K.; Jaiswal, A.; and Panda, P. 2019. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784): 607–617.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.

Sharmin, S.; et al. 2020. Inherent Adversarial Robustness of Deep Spiking Neural Networks: Effects of Discrete Input Encoding and Non-Linear Activations. *arXiv preprint arXiv:2003.10399*.

Soen, A.; and Sun, K. 2021. On the Variance of the Fisher Information for Deep Learning. *Advances in Neural Information Processing Systems*, 34.

Stigliani, A.; Jeska, B.; and Grill-Spector, K. 2017. Encoding model of temporal processing in human visual cortex. *Proceedings of the National Academy of Sciences*, 114(51): E11047–E11056.

Wang, X.; Lin, X.; and Dang, X. 2020. Supervised learning in spiking neural networks: A review of algorithms and evaluations. *Neural Networks*, 125: 258–280.

Wu, Y.; Deng, L.; Li, G.; Zhu, J.; and Shi, L. 2018. Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in neuroscience*, 12: 331.

Wu, Y.; Deng, L.; Li, G.; Zhu, J.; Xie, Y.; and Shi, L. 2019. Direct training for spiking neural networks: Faster, larger, better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 1311–1318.

Xiao, H.; Rasul, K.; and Vollgraf, R. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Zhang, W.; and Li, P. 2020. Temporal spike sequence learning via backpropagation for deep spiking neural networks. *arXiv preprint arXiv:2002.10085*.

Zheng, H.; Wu, Y.; Deng, L.; Hu, Y.; and Li, G. 2020. Going Deeper With Directly-Trained Larger Spiking Neural Networks. *arXiv preprint arXiv:2011.05280*.