

Neuromorphic Data Augmentation for Training Spiking Neural Networks

Yuhang Li[✉], Youngeun Kim[✉], Hyoungeob Park[✉], Tamar Geller[✉], and
Priyadarshini Panda[✉]

Yale University, New Haven, CT 06511, USA

{yuhang.li, youngeun.kim, hyoungeob.park, tamar.geller, priya.panda}@yale.edu

Abstract. Developing neuromorphic intelligence on event-based datasets with Spiking Neural Networks (SNNs) has recently attracted much research attention. However, the limited size of event-based datasets makes SNNs prone to overfitting and unstable convergence. This issue remains unexplored by previous academic works. In an effort to minimize this generalization gap, we propose Neuromorphic Data Augmentation (NDA), a family of geometric augmentations specifically designed for event-based datasets with the goal of significantly stabilizing the SNN training and reducing the generalization gap between training and test performance. The proposed method is simple and compatible with existing SNN training pipelines. Using the proposed augmentation, for the first time, we demonstrate the feasibility of unsupervised contrastive learning for SNNs. We conduct comprehensive experiments on prevailing neuromorphic vision benchmarks and show that NDA yields substantial improvements over previous state-of-the-art results. For example, the NDA-based SNN achieves accuracy gain on CIFAR10-DVS and N-Caltech 101 by 10.1% and 13.7%, respectively. Code is available on GitHub (URL).

Keywords: Data Augmentation, Event-based Vision, Spiking Neural Networks

1 Introduction

Spiking Neural Networks (SNNs), a representative category of models in neuromorphic computing, have received attention as a prospective candidate for low-power machine intelligence [52]. Unlike the standard Artificial Neural Networks (ANNs), SNNs deal with binarized spatial-temporal data. A popular example of this data is the DVS event-based dataset.¹ Each pixel inside a DVS camera is operated independently and asynchronously, reporting new brightness when it changes, and staying silent otherwise [61]. The spatial-temporal information encoded in DVS data can be suitably leveraged by the temporally evolving spiking neurons in SNNs. In Fig. 1, we describe the process of data recording with a

¹ In this paper, most event-based datasets we use are collected with Dynamic Vision Sensor (DVS) cameras, therefore we also term them as DVS data for simplicity.

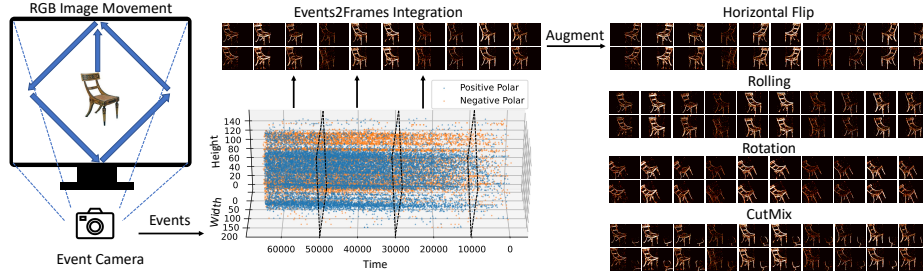


Fig. 1: *Left:* Event data collection of a chair image from N-Caltech101 [47], these events are too sparse to process, thus we integrate them into frames of sparse tensors. *Right:* Our neuromorphic data augmentation on events data.

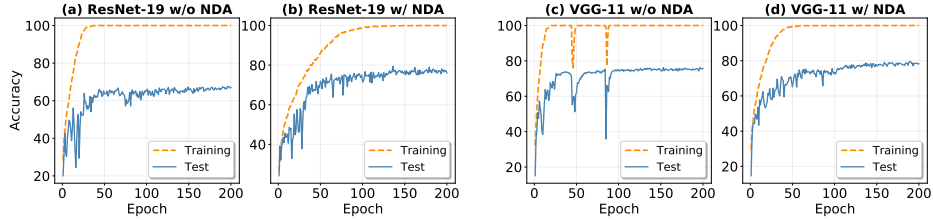


Fig. 2: Training/test accuracy curves of ResNet-19 and VGG-11 on CIFAR10-DVS dataset. Networks trained with NDA tends to have better convergence.

DVS camera. First, an RGB image is programmed to move in a certain trajectory (or the camera moves in a reverse way), and then the event camera records the brightness change and outputs an event stream. Note that the raw event-based DVS data is too sparse to extract features. Thus, we integrate the events into multiple frames and we study this sparse frame-based data [63] (see the events2frames integration in Fig. 1).

Due to the high cost of collecting DVS data [43], existing DVS datasets usually contain limited data instances [37]. Consequently, models trained on raw DVS datasets exhibit large training-test performance gaps due to overfitting. For instance, CIFAR10-DVS [37] only contains 10k data points while CIFAR-10 RGB dataset contains 60k data points. In Fig. 2, we plot the training-test accuracies on CIFAR10-DVS using VGG-11 [56] and ResNet-19 [28]. It is obvious that training accuracy increases swiftly and smoothly. On the contrary, the test accuracy oscillates and remains at half of the training accuracy. Although advanced training algorithms [69,17] have been proposed to improve the generalization of SNNs, data scarcity remains a major challenge and needs to be addressed. A naive way to increase DVS data samples is using the DVS camera to record additional augmented RGB images. However, during training, this carries prohibitively high latency overhead due to the cost of DVS recording. Because of this, we focus on a data augmentation technique that can be directly applied to DVS data in order to balance efficiency and effectiveness.

We propose Neuromorphic Data Augmentation (NDA) to transform the off-the-shelf recorded DVS data in a way that prevents overfitting. To ensure the consistency between augmentation and event-stream generation, we extensively investigate various possible augmentations and identify the most beneficial family of augmentations. In addition, we show that the proposed data augmentation techniques lead to a substantial gain of training stability and generalization performance (cf. Fig. 2). Even more remarkably, NDA enables SNNs to be trained through unsupervised contrastive learning without the usage of labels. Fig. 1 (right) demonstrates four examples where NDA is applied.

The main contributions of this paper are:

1. We propose Neuromorphic Data Augmentation for training Spiking Neural Networks on event datasets. Our proposed augmentation policy significantly improves the test accuracy of the model with negligible cost. Furthermore, NDA is compatible with existing training algorithm.
2. We conduct extensive experiments to verify the effectiveness of our proposed NDA on several benchmark DVS datasets like CIFAR10-DVS, N-Caltech 101, N-Cars, and N-MNIST. NDA brings a significant accuracy boost when compared to the previous state-of-the-art results.
3. In a first of its kind analysis, we show the suitability of NDA for unsupervised contrastive learning on the event-based datasets, enabling SNN feature extractors to be trained without labels.

2 Related Work

2.1 Data Augmentation

Today, deep learning is a prevalent technique in various commercial, scientific, and academic applications. Data augmentation [54] plays an indispensable role in the deep learning model, as it forces the model to learn invariant features, and thus helps generalization. The data augmentation is applied in many areas of vision tasks including object recognition [8,42,9], objection detection [71], and semantic segmentation [51]. Apart from learning invariant features, data augmentation also has other specific applications in deep learning. For example, adversarial training [19,59] leverages data augmentation to create adversarial samples and thereby improves the adversarial robustness of the model. Data augmentation is also used in generative adversarial networks (GAN) [1,31,23,3], neural style transfer [30,20], and data inversion [68,41]. For event-based data, there are few data augmentation techniques. EventDrop [22], for example, randomly removes several events due to noise produced in the DVS camera. Our work, on the other hand, tackles the consistency problem of directly applying data augmentations to event-based data. A related augmentation technique is video augmentation [4], which also augmentation spatial-temporal data. The main difference is that video augmentation can utilize photometric and color augmentations while our NDA can only adopt geometric augmentations.

2.2 Spiking Neural Networks

Spiking Neural Networks (SNNs) are often recognized as the third generation of generalization methods for artificial intelligence [2]. Unlike traditional Artificial Neural Networks (ANNs), SNNs apply the spiking mechanism inside the network. Therefore, the activation in SNNs is binary and adds a temporal dimension. Current practices consist of two approaches for obtaining an SNN: *Direct training* and *ANN-to-SNN conversion*. Conversion from a pre-trained ANN can guarantee good performance [11,15,53,38,39,25], however, in this paper, we do not study conversion since our topic is a neuromorphic dataset that requires direct training. Training SNNs [64,63,26,40,12,24] requires spatial-temporal backpropagation. Recently, more methods and variants for such spatio-temporal training have been proposed: [50] present hybrid training, [69,32] propose a variant of batch normalization [29] for SNN; [49,17] propose training threshold and potential for better accuracy. However, most of these works focus on algorithm or architecture optimization to achieve improved performance. Because of their sparse nature, DVS datasets present an orthogonal problem of overfitting that is not addressed by the previously mentioned works.

3 Conventional Augmentations for RGB Data

We generally divide the current data augmentation techniques for ANN training on natural RGB image datasets into two types, namely photometric & color augmentations and geometric augmentations.

Photometric and Color Augmentations. This type indicates any augmentations that can transform the image illumination and color space. Fig. 3 (2nd-5th examples) demonstrates some examples of increasing contrast, saturation, and imposing gray scale as well as Gaussian blur. Typically, photometric augmentation is applied to an image by changing each pixel value. For instance, the contrast enhancement will use $f(x) = \text{clip}(ax - \frac{1}{2}a + \frac{1}{2}, 0, 1)$ where $a > 1$, to push pixel values close to black (zero value) and white (one value) and is applied in a pixel-wise manner. The color augmentation includes some color space transformation by casting more red, blue, or green colors on the image (*e.g.* saturation and grayscale). Generally, both Photometric and Color augmentations can be categorized as value-based augmentation where a transformation $f(x)$ applies to all pixels. Therefore, we also categorize augmentations like a Gaussian filter where a convolution with a Gaussian kernel is applied to this class.

Geometric Augmentations. Unlike *value-based* augmentations, geometric augmentations do not seek to alter every pixel value with a certain criterion. Rather, they change the images based on the coordinate of each pixel, *i.e.* the index of each element. In Fig. 3 (6th-9th examples), we visualize several geometric augmentation examples of horizontal flipping, resizing, rotation and cutout. For example, horizontal flipping reverses the order of each pixel and turns the image by 180 degrees. Rotation can be viewed as moving the location of pixels to another place. Cutout [13] applies a 0/1 mask to the original image. All of these augmentations are *index-based*.



Fig. 3: Types of RGB data augmentation. The 2-5th examples include the photometric & color augmentations and the 6-9th examples contain the geometric augmentations.

4 Neuromorphic Augmentations for Event-Based Data

The event-based datasets contain sparse, discrete, and time-series samples, which have fundamentally different formats when compared to RGB images. As a result, the above conventional augmentations cannot all be directly applied. To explore which augmentation is useful for event data, we study the case of photometric and geometric augmentations separately. We also discuss the potential application of neuromorphic data augmentation.

4.1 DVS Data and Augmentation

DVS cameras are data-driven sensors: their output depends on the amount of motion or brightness change in the scene [18]. Mathematically, the output of event camera $\mathbf{x}_E \in \mathbb{B}^{t \times p \times w \times h}$ is a 4-D tensor ($\mathbb{B}$ is the binary domain $\{0, 1\}$ and here we treat it as a sparse binary tensor, *i.e.* we also record 0 if there are no events), where t is the time step, p is the polarity ($p = 2$ corresponding to positive and negative polars) and w, h are the width and the height of the event tensor. The event generation (see details in [18]) can be modeled as

$$\mathbf{x}_E(t, 1, x, y) = \begin{cases} 1 & \text{if } \log V(t, x, y) - \log V(t - \Delta t, x, y) > \alpha \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

$$\mathbf{x}_E(t, 0, x, y) = \begin{cases} 1 & \text{if } \log V(t, x, y) - \log V(t - \Delta t, x, y) < -\alpha \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where, $\mathbf{x}_E(t, 1, x, y)$ is the generated event stream at time step t , spatial coordinate (x, y) and positive polarity. V is the photocurrent (or “brightness”) of the original RGB image. Δt is the time elapsed since the last event generated at the same coordinate. During DVS data recording, the RGB images are programmed to move in a certain trajectory (or the camera moves in a reverse way). As time evolves, if the pixel value changes fast and exceeds a certain threshold α , an event will be generated, otherwise it will stay silent, meaning that the output is 0 in $\mathbf{x}_E$. The event stream will be split into two channels, *i.e.* two polarities. Positive events are integrated into positive polarity and vice versa.

Consider the RGB data as $\mathbf{x} \in \mathbb{C}^{3 \times w \times h}$ ($\mathbb{C}$ is the continuous domain in $[0, 1]$). We use the function $f(\mathbf{x}) : \mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{C}^{3 \times w \times h}$ to augment the RGB data. Note

Table 1: Comparison among all potential augmentation for DVS data. f_P and f_G mean photometric and geometric augmentations. $\mathbb{C}$ stands for the continuous domain in $[0, 1]$ and $\mathbb{B}$ stands for the binary domain $\{0, 1\}$.

Aug.	Combination	Input-output	Pros	Cons
f_P	$g \circ f_P(\mathbf{x})$	$\mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{B}^{t \times p \times w \times h}$	i. Effective augmentation	i. Impractical to record huge amount of DVS data
f_P	$f_P \circ g(\mathbf{x})$	$\mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{C}^{t \times p \times w \times h}$	i. Practical	i. Not effective, ii. Creates continuous data
f_G	$g \circ f_G(\mathbf{x})$	$\mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{B}^{t \times p \times w \times h}$	i. Effective augmentation	i. Impractical to record huge amount of DVS data
f_G	$f_G \circ g(\mathbf{x})$	$\mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{B}^{t \times p \times w \times h}$	i. Practical and effective, ii. Approximates $g \circ f_G(\mathbf{x})$	None

that $f(\mathbf{x})$ can be both photometric or geometric augmentation, and is randomly sampled from a set of augmentations. The optimization objective of training an ANN with RGB augmented data can be given by $\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \ell(\mathbf{w}, f(\mathbf{x}_i), \mathbf{y}_i)$.

Now consider event-based data $\mathbf{x}_E$. We define RGB to DVS data function $\mathbf{x}_E = g(\mathbf{x}) : \mathbb{C}^{3 \times w \times h} \rightarrow \mathbb{B}^{t \times p \times w \times h}$.² A naive approach to augment DVS data when training SNN is to first augment RGB data and then translate them to event-stream form, *i.e.* $g \circ f(\mathbf{x})$. This method can ensure the augmentation is correctly implemented as well as yield the event-stream data. In fact, training with Poisson encoding [14, 52] uses such form $g \circ f(\mathbf{x})$ where g is the encoding function that translates the RGB images to spike trains. However, unlike Poisson encoding which can be implemented with one line of code, it would be very time-consuming and expensive to generate a large amount of DVS augmented data, *i.e.* $g \circ f(\mathbf{x})$. We propose a more efficient method, the neuromorphic data augmentation f_{NDA} which is directly applied to the DVS data $f_{\text{NDA}} \circ g(\mathbf{x})$. As a result, we avoid the expensive $g \circ f(\mathbf{x})$ in the training phase.

Ideally, NDA is supposed to satisfy $f_{\text{NDA}} \circ g(\mathbf{x}) \approx g \circ f(\mathbf{x})$. To fulfill this commutative law, the NDA data augmentation function must have the mapping of $f_{\text{NDA}} : \mathbb{B}^{t \times p \times w \times h} \rightarrow \mathbb{B}^{t \times p \times w \times h}$. Without loss of generality, a core component is to evaluate whether an arbitrary augmentation $f(\cdot)$ can achieve

$$f(H_\alpha [\log V(t) - \log V(t - \Delta t)]) \approx H_\alpha [\log f(V(t)) - \log f(V(t - \Delta t))], \quad (3)$$

where $H_\alpha[\cdot]$ is the Heaviside step function (*i.e.* returns 1 when the input is greater than α otherwise 0). Note that we omit the spatial coordinate here. Recall that photometric & color augmentation are *value-based* augmentation schemes. This brings two problems: first, on the left hand side of Eq. (3), the augmented DVS data is not event-stream since the *value-based* transformation outputs continuous values. Second, on the right hand side of Eq. (3), predicting the event is

² Note that using event camera $g(\mathbf{x})$ to generate DVS data is expensive and impractical during run-time. It is easier to pre-collect the DVS data with a DVS camera and, then work with the DVS data during runtime.

intractable. Due to brightness difference change, it is unclear whether the event is activated or remains silent after augmentation ($f(\log V(t)) - f(\log V(t-1))$). Therefore, we cannot use this type of augmentation in NDA.

It turns out that *index-based* geometric augmentation is suitable for NDA. First, the geometric transformation only changes the position, and therefore the augmented data is still maintained in event-stream. Second, assume an original event which has $\log V(t) - \log V(t-1) > \alpha$ is still generated in the case of geometric augmentation $f(\log V(t)) - f(\log V(t-1))$. The only difference is position change which can be effectively predicted by the left hand side of Eq. (3). For example, rotating a generated set of DVS data and generating a rotated set of DVS data have the same output. Therefore, our NDA consists of several geometric augmentations:

Horizontal Flipping. Flipping is widely used in computer vision tasks. It turns the order of horizontal dimension w . Note that for DVS data, the polarity and the time dimension are kept intact. We set the probability of flipping to 0.5.

Rolling. Rolling means randomly shifting the geometric position of the DVS image. Similar to a bit shift, rolling can move the pixels left or right in the horizontal dimension. In the vertical dimension, we can also shift the pixels up or down. Note that both circular and non-circular shifts are acceptable since DVS image borders are likely to be 0. Each time we apply rolling, the horizontal shift value a and the vertical shift value b are sampled from an integer uniform distribution $\mathcal{U}(-c, c)$, where c is a hyper-parameter.

Rotation. The DVS data will be randomly rotated in the clockwise or counter-clockwise manner. Similar to rolling, each time we apply rotation, the degree of rotation is sampled from a uniform distribution $\mathcal{U}(-d, d)$, where positive degree means clockwise rotation and vice versa. d is a hyperparameter.

Cutout. Cutout was originally proposed in [13] to address overfitting. This method randomly erases a small area of the image, with a similar effect of dropout [58]. Mathematically, first, a square is generated with random size and random position. The side length is sampled from an integer uniform distribution $\mathcal{U}(1, e)$, and then a center inside the image is randomly chosen to place the square. All pixels inside the square are masked off.

Shear. Shear mapping originated from plane geometry. It is a linear map that displaces each point in a fixed direction [62]. Mathematically, a horizontal shear (or ShearX) is a function that takes point (x, y) to point $(x + my, y)$. All pixels above x -axis are displaced to the right if $m > 0$, and to the left if $m < 0$. Note that we do not use vertical shear (ShearY) following prior arts [8,9]. The shear factor m is also sampled from some uniform distribution $\mathcal{U}(n, n)$.

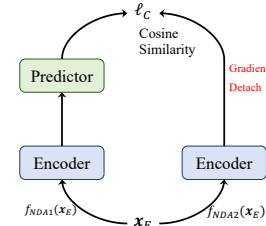
CutMix. Proposed in [67], CutMix is an effective linear interpolation of two input data and labels. Consider two input data samples $\mathbf{x}_{N1}$ and $\mathbf{x}_{N2}$ and their corresponding label $\mathbf{y}_1, \mathbf{y}_2$. CutMix is formulated as

$$\tilde{\mathbf{x}}_n = \mathbf{M}\mathbf{x}_{N1} + (\mathbf{1} - \mathbf{M})\mathbf{x}_{N2}, \quad \tilde{\mathbf{y}} = \beta\mathbf{y}_1 + (1 - \beta)\mathbf{y}_2, \quad (4)$$

where $\mathbf{M}$ is a binary mask (similar to the one used in Cutout), and β is the ratio between the area of one and the area of zero in that mask.

Table 2: Look-up table for exchanging N with augmentation hyper-parameters.

N	Rolling c	Rotation d	Cutout e	Shear n
1	3	15	8	0.15
2	5	30	16	0.30
3	7	45	24	0.45

**Fig. 4:** NDA for SSL-SNN.

4.2 Augmentation Policy

With each above augmentation, one can generate more training samples. In this section, we explore the combination and intensity of the considered augmentations. We first set flipping and CutMix as the default augmentation, which means they are always enabled. The flipping probability is set to 0.5 and the CutMix interpolation factor is sampled from $\beta(1, 1)$. Second, for all time step of input data, we randomly sample several augmentations from {Rolling, Rotation, Cutout, ShearX}, inspired by prior work [9,46,70]. We define two hyper-parameters: M for the number of augmentations and N for the intensity of augmentations. That is, before each forward pass, we randomly sample M augmentations with N -level intensity (the higher the N , the greater the difference between augmented and original data). M can be set from 1 to 4. In Table 2, we describe the corresponding relationship between N and the augmentation hyper-parameters. For instance, using M2N3 policy during a forward pass yields 2 randomly sampled augmentations with $N = 3$ intensity as shown in Table 2. Our algorithm can be simply implemented and can be an add-on extension to existing SNN training methods. In the following experiments, we will show how the number of augmentations and the intensity of augmentations impact test accuracy.

4.3 Application to Unsupervised Contrastive Learning

Besides improving the supervised learning of SNNs, we show another application of NDA, *i.e.* unsupervised contrastive learning for SNNs. Unsupervised contrastive learning [27,5,6,21] is a widely used and well-performing learning algorithm without the requirement of labels. The idea of contrastive learning is to learn the similarity between paired input images. Each paired output can be either similar or different. It would be easy to identify different images, as they are naturally distinct. However, for a similar image pair, it is required to augment the same image to different tensors and optimize the network to output the same feature. This task is a simpler task that doesn't require any labels as compared to image classification or segmentation, which makes it perfect for learning some low-level features and transferring the model to some high-level vision tasks.

In this paper, we implement unsupervised contrastive learning for SNNs based on a recent work Simple Siamese Network [7] with our proposed NDA, as

illustrated in Fig. 4. First, a DVS data input $\mathbf{x}_E$ is augmented into two samples: $f_{\text{NDA1}}(\mathbf{x}_E)$ and $f_{\text{NDA2}}(\mathbf{x}_E)$. Our goal is to maximize the similarity of two outputs and one output is further processed by a predictor head to make the final prediction. The contrastive loss is defined as the cosine similarity between the predictor output and the encoder output. In contrastive learning, there is a problem called *latent space collapsing*, meaning that the output of a network is the same, irrespective of different inputs which can render the network useless. This collapsing can always yield minimum loss. To address this problem, gradient detach is applied to the branch without predictor. It is noteworthy to mention that all the contrastive learning schemes require data augmentation to make model learn invariant features. As a broader impact, this could be helpful when the event camera collects new data that is not easy to be labeled because of raw events.

After the unsupervised pre-training stage, we drop the predictor and only save the encoder part. This encoder is used for transfer learning, *i.e.* construct a new head (the last fully-connected layer *a.k.a* the classifier) and finetune the model. We will provide transfer results in the experiments section below.

5 Experiments

In this section, we will verify the effectiveness and efficiency of our proposed NDA. In section 5.2, we compare our model with existing literature. In section 5.3, we analyze our method in terms of sharpness and efficiency. In section 5.1, we give an ablation study of our methods. In section 5.4, we provide the results of unsupervised contrastive learning.

Implementation details. We implement our experiments with the Pytorch package. All our experiments are run on 4 GPUs. We use ResNet-19 [28,69] and VGG-11 [56,17] as baseline models. Note that all ReLU layers are changed to the Leaky integrate-and-fire module and all max-pooling layers are changed to average pooling. We use tdBN [69] as our baseline SNN training method. For our own implementation, we only change the data augmentation part and keep other training hyper-parameters aligned. We use M1N2 NDA configuration for all our experiments. The total batch size is set to 256 and we use Adam optimizer. For all our experiments, we train the model for 200 epochs and the learning rate is set to 0.001 followed by a cosine annealing decay [45]. The weight decay is $1e-4$. We verify our method on the following DVS benchmarks:

CIFAR10-DVS [37]. CIFAR10-DVS contains 10K DVS images recorded from the original CIFAR10 dataset. We apply a 9 : 1 train-valid split (*i.e.* 9k training images and 1k validation images). The resolution is 128×128 , we resize all of them to 48×48 in our training and we integrate the event data into 10 frames per sample.

N-Caltech 101 [47]. N-Caltech 101 contains 8831 DVS images recorded from the original Caltech 101 dataset. Similarly, we apply 9 : 1 train-valid split and resize all images to 48×48 . We use the spikingjelly package [16] to process the data and integrate them into 10 frames per sample.

Table 3: Ablation study: comparison between photometric/color augmentation and geometric augmentation, and comparison with the intensity and the number of the augmentation per data.

Dataset	Photo/Color	Geo-M1N1	Geo-M1N2	Geo-M2N2	Geo-M3N3
CIFAR10-DVS	62.8	73.4	78.0	75.1	71.4
N-Caltech101	64.0	74.4	78.6	72.7	65.1

N-MNIST [47]. The neuromorphic MNIST dataset is a converted dataset from MNIST. It contains 50K training images and 10K validation images. We preprocess it in the same way as in N-Caltech 101.

N-Cars [57]. Neuromorphic cars dataset is a binary classification dataset with labels either from *cars* or from *background*. It contains 7940 car and 7482 background training samples, 4396 car and 4211 background test samples. We preprocess it in the same way as in N-Caltech 101.

5.1 Ablation Study

Augmentation Choice. In theory, photometric & color augmentations are not supposed to be used for DVS data. To verify this, we **compulsorily cast Color-Jitter and GaussianBlur augmentation to the DVS data** (note that the data is not event-stream after these augmentations) and compare the results with geometric augmentation. The results are shown in Table 3 (all entries are results trained with ResNet-19). We find that photometric and color augmentation (Jitter + GaussianBlur) performs much worse than geometric augmentation, regardless of the dataset. This confirms our analysis that *value-based* augmentations are not suitable for NDA.

Augmentation Policy. We also test the augmentation intensity as well as the number of the augmentation. In Table 3, we show that the intensity of the augmentation satisfies some bias-variance trade-off. The augmentation can become neither too simple so that the data is not diverse enough, nor too complex so that the data does not contain useful event information.

5.2 Comparison with Existing Literature

We first compare our NDA method on the CIFAR10-DVS dataset. The results are shown in Table 4. We compare our method with Gabor-SNN, Streaming rollout, tdbN, and PLIF [57,44,64,34,69,17]. Among these baselines, tdbN and PLIF achieve better accuracy. We reproduce tdbN with our own implementations. When training with NDA, we use the best practice M1N2 for sampling augmentations. With NDA, our ResNet-19 reaches 78.0% top-1 accuracy, outperforming the baseline without augmentation by a large margin. After scaling the data to a resolution of 128, our method gets 81.7% accuracy on VGG-11 model. We also compare our method on N-Caltech 101 dataset. There aren't many SNN

Table 4: Accuracy comparison with different methods on CIFAR10-DVS, N-Caltech 101, N-Cars, we use tdbN in our model. Acc. is referred as the top-1 accuracy.

Method	Model	CIFAR10-DVS		N Caltech-101		N-Cars	
		<i>T</i> Step	Acc.	<i>T</i> Step	Acc.	<i>T</i> Step	Acc.
HOTS [35]	N/A	N/A	27.1	N/A	21.0	10	54.0
Gabor-SNN [57]	2-layer CNN	N/A	28.4	N/A	28.4	-	-
HATS [57]	N/A	N/A	52.4	N/A	64.2	10	81.0
DART [48]	N/A	N/A	65.8	N/A	66.8	-	-
CarSNN [60]	4-layer CNN	-	-	-	-	10	77.0
CarSNN [60]	4-layer CNN ²	-	-	-	-	10	86.0
BNTT [32]	6-layer CNN	20	63.2	-	-	-	-
Rollout [34]	VGG-16	48	66.5	-	-	-	-
SALT [33]	VGG11	20	67.1	20	55.0	-	-
LIAF-Net [65]	VGG-like	10	70.4	-	-	-	-
tdBN [69]	ResNet-19 ¹	10	67.8	-	-	-	-
PLIF [17]	VGG-11 ²	20	74.8	-	-	-	-
tdBN (w/o NDA) ³	ResNet-19 ¹	10	67.9	10	62.8	10	82.4
tdBN (w/. NDA)	ResNet-19 ¹	10	78.0	10	78.6	10	87.2
tdBN (w/o NDA) ³	VGG-11	10	76.2	10	67.2	10	84.4
tdBN (w/. NDA)	VGG-11	10	79.6	10	78.2	10	90.1
tdBN (w/o NDA) ³	VGG-11 ²	10	76.3	10	72.9	10	87.4
tdBN (w/. NDA)	VGG-11 ²	10	81.7	10	83.7	10	91.9

¹ Quadrupled channel number, ² 128 × 128 resolution, ³ Our implementation.

works on this dataset. Most of them are non-neural network based methods using hand-crafted processing [35,57,48]. NDA can obtain a high accuracy network with 15.8% absolute accuracy improvement over the baseline without NDA. Our VGG-11 even reaches 83.7% accuracy with full 128×128 resolution. This indicates that only improving the network training strategy is less effective. Next, we test our algorithm on the N-Cars dataset. The primary comparison work is CarSNN [60] which optimizes a spiking network and deploys it on the Loihi chip [10]. tdbN trained ResNet-19 achieve 82.4% using 48×48 resolution. Simply applying NDA improves the accuracy by 4.8% without any additional modifications. When training with full resolution, we have 4.5% accuracy improvement with VGG-11.

Finally, we validate our method on the N-MNIST dataset (as shown in Table 5). N-MNIST is harder than the original MNIST dataset, but most baselines get over 99% accuracy. We use the same architecture as PLIF. Our NDA uplifts the model by 0.12% in terms of accuracy. Although this is a marginal improvement, the final test accuracy is quite close to the 100% mark.

Evaluating NDA on ANN. We also evaluate if our method is applicable to ANNs. We mainly compare our method with EventDrop on frame data [22], another work that augments the event data by random dropping some events (like Cutout used in NDA). In Table 6, we find NDA can greatly improve the performance of ANN. For VGG-19, our NDA can boost the accuracy to 82.8%. For N-Cars

Table 5: N-MNIST accuracy comparison with different methods, we use tdBN in our model.

Method	Model	Time Step	Top-1 Acc.
Lee <i>et al.</i> [36]	2-layer CNN	N/A	98.61
SLAYER [55]	3-layer CNN	N/A	99.20
Gabor-SNN [57]	2-layer CNN	N/A	83.70
HATS [57]	N/A	N/A	99.10
STBP [64]	6-layer CNN	10	99.53
PLIF [17]	4-layer CNN	10	99.61
tdBN (w/o NDA)	4-layer CNN	10	99.58
tdBN (w/. NDA)	4-layer CNN	10	99.70

Table 6: Comparison with EventDrop [22] using ANNs for DVS datasets. EventDrop models are ImageNet pre-trained.

Model	Method	N-Caltech 101	N-Cars
ResNet-34	Baseline [22] ¹	77.4	91.8
	EventDrop [22] ¹	78.2	94.0
	Ours (w/o NDA)	67.7	90.5
	Ours (w/. NDA)	81.2	95.5
VGG-19	Baseline [22] ¹	72.3	91.6
	EventDrop [22] ¹	75.0	92.7
	Ours (w/o. NDA)	65.4	90.3
	Ours (w/. NDA)	82.8	94.5

dataset, we adopt the same pre-processing as EventDrop, *i.e.* using 8 : 2 train-validation split to optimize the model, and our method attains 1.8% higher accuracy even without any ImageNet pre-training.

5.3 Analysis

Model Sharpness. Apart from the test accuracy, we can measure the *sharpness* of a model to estimate its generalizability. A sharp loss surface implies that a model has a higher probability of mispredicting unseen samples. In this section, we will use two simple and popular metrics: (1) Hessian spectra and (2) Noise injection. Hessian matrix is the second-order gradient matrix. The second-order gradient contains the curvature information of the model. Here, we measure the topk Hessian eigenvalues and the trace to estimate the sharpness. Low eigenvalue and low trace lead to low curvature and better performance. We compute the 1st, 5th eigenvalues as well as the trace of the Hessian in ResNet-19 using PyHessian [66]. The model is trained on CIFAR10-DVS with and without NDA. We summarize the results in Table 7. We show that the Hessian spectra of the model trained with NDA is significantly lower than that without NDA. Moreover, the trace of the Hessian also satisfies this outcome.

Another way to measure the sharpness is noise injection. We randomly inject Gaussian noise sampled from $\mathcal{N}(0, \gamma)$ into the model weights, where γ is a hyper-parameter controlling the range of the noise. We run 5 times for each γ and record the mean & standard deviation, which are summarized in Fig. 5. It can be seen that the model with NDA is much more robust to the noise we imposed on the weights. For example, when we set the noise to $\mathcal{N}(0, 0.01)$, the model with NDA only suffers a slight 1.5% accuracy drop, while the model without NDA has a 19.4% accuracy drop. More evidently, when casting $\mathcal{N}(0, 0.02)$ noise, our NDA model suffers 11.3% accuracy decline, while the counterpart suffers a drastic 43.4% accuracy decline. This indicates that NDA serves a similar effect with regularization technique.

Algorithm Efficiency. Our NDA is efficient and easy to implement. To demonstrate its efficiency, we hereby report the time cost of NDA. We use Intel XEON(R) E5-2620 v4 CPU to test the data loader. When loading CIFAR10-DVS, the CPU

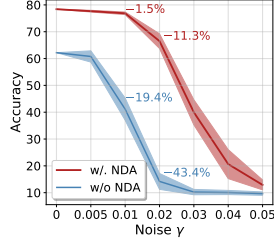


Fig. 5: Comparison of robustness under Gaussian noise injected to weights.

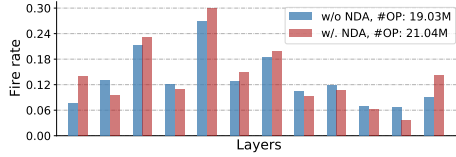


Fig. 6: Fire Rate of several LIF layers of ResNet-19 trained on CIFAR10-DVS.

Table 7: Hessian spectra comparison. λ_1, λ_5, Tr refer to the 1st, 5th highest eigenvalue and the trace of the Hessian matrix. We record the model at 100, 200, 300 epochs. (The lower the Hessian spectrum, the flatter the converged minimum is).

Epoch	w/o NDA			w/. NDA		
	λ_1	λ_5	Tr	λ_1	λ_5	Tr
100	910.7	433.4	6277	424.3	73.87	1335
200	3375	1416	21342	516.4	155.3	1868
300	3404	1686	20501	639.7	187.5	2323

Table 8: Ablation Study: the effect of validation dataset.

Method	Acc. No Valid.	Acc. 15% Valid.
PLIF [17]	74.8	69.0
w/o NDA	63.4	58.5
w/. NDA	78.0	74.4

expends additional 15.2873 seconds in applying NDA to 9000 DVS training images. The average cost basically amounts to 1.7ms per image.

We also estimate the energy consumption of the model using NDA. Specifically, we use the number of operations and roughly assume energy consumption is linearly proportional to the number of operations [49,69]. We do not count the operation if the spike does not fire. Thus, the number of operations can be computed as $FireRate \times MAC$. In Fig. 6, we visualize the *FireRate* of several layers in the trained model with and without NDA. We can find that the *FireRate* is similar (note, NDA is slightly higher) and the final number of operations are very close in both cases. Model with NDA only expends a 10% higher number of operations than the model without NDA, demonstrating the efficiency of NDA.

Training with validation dataset. Following [17], we test our algorithm on a challenging task. We take 15% of the data from the training dataset to build a *validation dataset*. Then, we report the *test accuracy* when the model reaches best *validation accuracy*. We run experiments on CIFAR10-DVS with ResNet-19. In this case (Table 8), our NDA model yields 16.3% accuracy improvement over a model without NDA.

5.4 Unsupervised Contrastive Learning

In this section, we test our NDA algorithm with unsupervised contrastive learning. Usually, the augmentation in unsupervised contrastive learning requires stronger augmentation than that in supervised learning [5]. Thus we use NDA-M3N2 for augmentation. We pre-train a VGG-11 on the N-Caltech 101 dataset

Table 9: Unsupervised transfer learning results on CIFAR10-DVS. All the model is pre-trained on N-Caltech 101.

Pre-training Method	Finetuning Method	Test Acc.
No Pre-training	Train@300	81.7
Supervised	Finetune@100	77.4
Supervised	Finetune@300	80.9
Unsupervised	Finetune@100	80.8
Unsupervised	Finetune@300	82.1

with SimSiam learning (cf. Section 4.3, Fig. 4). N-Caltech 101 contains more visual categories and can be a good pre-training dataset. In this experiment, we use the original 128×128 resolution DVS data, and use the simple-siamese method to train the network for 600 epochs. The batch size is set to 256.

After pre-training, we replace the predictor with a zero-initialized fully-connected classifier and finetune the model for 100/300 epochs on CIFAR10-DVS. We also add supervised pre-training with NDA and no pre-training (*i.e.* directly train a model on downstream task) as our baselines. We show the transferability results in Table 9. Interestingly, we find supervised pre-training has no effect on transfer learning with such DVS datasets, which is different from conventional natural image classification transferability results. This may be attributed to the distance between dataset domains that is larger in DVS than that in RGB dataset. The model pre-trained on N-Caltech 101 with supervised learning only achieves 80.9% accuracy, which is even 0.8% lower than the no pre-training method. The unsupervised learning with NDA achieves significantly better transfer results: finetuning only for 100 epochs takes the model to 80.8% accuracy on CIFAR10-DVS, and 300 epochs finetuning yields 82.1% test accuracy, establishing a new state of the art.

6 Conclusions

We introduce the Neuromorphic Data Augmentation technique (NDA), a simple yet effective method that improves the generalization of SNNs on event-based data. NDA allows users to generate new high-quality event-based data instances that force the model to learn invariant features. Furthermore, we show that NDA acts like regularization that achieves an improved bias-variance trade-off. Extensive experimental results validate that NDA is able to find a flatter minimum with the higher test accuracy and enable unsupervised pre-training for transfer learning. However, the current NDA lacks value-based augmentations for events, which may be realized by logical operations and studied in the future.

Acknowledgment. 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

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International conference on machine learning. pp. 214–223. PMLR (2017)
2. Basegmez, E.: The next generation neural networks: Deep learning and spiking neural networks. In: Advanced Seminar in Technical University of Munich. pp. 1–40. Citeseer (2014)
3. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
4. Budvytis, I., Sauer, P., Roddick, T., Breen, K., Cipolla, R.: Large scale labelled video data augmentation for semantic segmentation in driving scenarios. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 230–237 (2017)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
6. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
7. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15750–15758 (2021)
8. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501 (2018)
9. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020)
10. Davies, M., Srinivasa, N., Lin, T.H., China, G., Cao, Y., Choday, S.H., Dimou, G., Joshi, P., Imam, N., Jain, S., et al.: Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee Micro* **38**(1), 82–99 (2018)
11. Deng, S., Gu, S.: Optimal conversion of conventional artificial neural networks to spiking neural networks. arXiv preprint arXiv:2103.00476 (2021)
12. Deng, S., Li, Y., Zhang, S., Gu, S.: Temporal efficient training of spiking neural network via gradient re-weighting. arXiv preprint arXiv:2202.11946 (2022)
13. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
14. Diehl, P.U., Neil, D., Binas, J., Cook, M., Liu, S.C., Pfeiffer, M.: Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In: 2015 International joint conference on neural networks (IJCNN). pp. 1–8. iee (2015)
15. Diehl, P.U., Zarella, G., Cassidy, A., Pedroni, B.U., Neftci, E.: Conversion of artificial recurrent neural networks to spiking neural networks for low-power neuromorphic hardware. In: 2016 IEEE International Conference on Rebooting Computing (ICRC). pp. 1–8. IEEE (2016)
16. Fang, W., Chen, Y., Ding, J., Chen, D., Yu, Z., Zhou, H., Tian, Y., other contributors: Spikingjelly. <https://github.com/fangwei123456/spikingjelly> (2020), accessed: YYYY-MM-DD
17. Fang, W., Yu, Z., Chen, Y., Masquelier, T., Huang, T., Tian, Y.: Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2661–2671 (2021)

18. Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., et al.: Event-based vision: A survey. arXiv preprint arXiv:1904.08405 (2019)
19. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. *The journal of machine learning research* **17**(1), 2096–2030 (2016)
20. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2414–2423 (2016)
21. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., Azar, M.G., et al.: Bootstrap your own latent: A new approach to self-supervised learning. arXiv preprint arXiv:2006.07733 (2020)
22. Gu, F., Sng, W., Hu, X., Yu, F.: Eventdrop: data augmentation for event-based learning. arXiv preprint arXiv:2106.05836 (2021)
23. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028 (2017)
24. Guo, Y., Tong, X., Chen, Y., Zhang, L., Liu, X., Ma, Z., Huang, X.: Recdis-snn: Rectifying membrane potential distribution for directly training spiking neural networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 326–335 (2022)
25. Han, B., Roy, K.: Deep spiking neural network: Energy efficiency through time based coding. In: *European Conference on Computer Vision* (2020)
26. Hazan, H., Saunders, D.J., Khan, H., Patel, D., Sanghavi, D.T., Siegelmann, H.T., Kozma, R.: Bindsnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in neuroinformatics* **12**, 89 (2018)
27. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9729–9738 (2020)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
29. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*. pp. 448–456. PMLR (2015)
30. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: A review. *IEEE transactions on visualization and computer graphics* **26**(11), 3365–3385 (2019)
31. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
32. Kim, Y., Panda, P.: Revisiting batch normalization for training low-latency deep spiking neural networks from scratch. arXiv preprint arXiv:2010.01729 (2020)
33. Kim, Y., Panda, P.: Optimizing deeper spiking neural networks for dynamic vision sensing. *Neural Networks* (2021)
34. Kugele, A., Pfeil, T., Pfeiffer, M., Chicca, E.: Efficient processing of spatio-temporal data streams with spiking neural networks. *Frontiers in Neuroscience* **14**, 439 (2020)
35. Lagorce, X., Orchard, G., Galluppi, F., Shi, B.E., Benosman, R.B.: Hots: a hierarchy of event-based time-surfaces for pattern recognition. *IEEE transactions on pattern analysis and machine intelligence* **39**(7), 1346–1359 (2016)
36. Lee, J.H., Delbruck, T., Pfeiffer, M.: Training deep spiking neural networks using backpropagation. *Frontiers in neuroscience* **10**, 508 (2016)

37. Li, H., Liu, H., Ji, X., Li, G., Shi, L.: Cifar10-dvs: an event-stream dataset for object classification. *Frontiers in neuroscience* **11**, 309 (2017)
38. Li, Y., Deng, S., Dong, X., Gong, R., Gu, S.: A free lunch from ann: Towards efficient, accurate spiking neural networks calibration. *arXiv preprint arXiv:2106.06984* (2021)
39. Li, Y., Deng, S., Dong, X., Gu, S.: Converting artificial neural networks to spiking neural networks via parameter calibration. *arXiv preprint arXiv:2205.10121* (2022)
40. Li, Y., Guo, Y., Zhang, S., Deng, S., Hai, Y., Gu, S.: Differentiable spike: Rethinking gradient-descent for training spiking neural networks. *Advances in Neural Information Processing Systems* **34**, 23426–23439 (2021)
41. Li, Y., Zhu, F., Gong, R., Shen, M., Dong, X., Yu, F., Lu, S., Gu, S.: Mixmix: All you need for data-free compression are feature and data mixing. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4410–4419 (2021)
42. Lim, S., Kim, I., Kim, T., Kim, C., Kim, S.: Fast autoaugment. *Advances in Neural Information Processing Systems* **32**, 6665–6675 (2019)
43. Lin, Y., Ding, W., Qiang, S., Deng, L., Li, G.: Es-imagenet: A million event-stream classification dataset for spiking neural networks. *Frontiers in Neuroscience* p. 1546 (2021)
44. Liu, Q., Ruan, H., Xing, D., Tang, H., Pan, G.: Effective aer object classification using segmented probability-maximization learning in spiking neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 34, pp. 1308–1315 (2020)
45. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* (2016)
46. Munoz-Bulnes, J., Fernandez, C., Parra, I., Fernández-Llorca, D., Sotelo, M.A.: Deep fully convolutional networks with random data augmentation for enhanced generalization in road detection. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. pp. 366–371. IEEE (2017)
47. Orchard, G., Jayawant, A., Cohen, G.K., Thakor, N.: Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in neuroscience* **9**, 437 (2015)
48. Ramesh, B., Yang, H., Orchard, G., Le Thi, N.A., Zhang, S., Xiang, C.: Dart: distribution aware retinal transform for event-based cameras. *IEEE transactions on pattern analysis and machine intelligence* **42**(11), 2767–2780 (2019)
49. Rath, N., Roy, K.: Diet-snn: Direct input encoding with leakage and threshold optimization in deep spiking neural networks. *arXiv preprint arXiv:2008.03658* (2020)
50. Rath, N., Srinivasan, G., Panda, P., Roy, K.: Enabling deep spiking neural networks with hybrid conversion and spike timing dependent backpropagation. *arXiv preprint arXiv:2005.01807* (2020)
51. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
52. Roy, K., Jaiswal, A., Panda, P.: Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**(7784), 607–617 (2019)
53. Rueckauer, B., Lungu, I.A., Hu, Y., Pfeiffer, M.: Theory and tools for the conversion of analog to spiking convolutional neural networks. *arXiv: Statistics/Machine Learning* (1612.04052), 0–0 (2016)
54. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6**(1), 1–48 (2019)

55. Shrestha, S.B., Orchard, G.: Slayer: Spike layer error reassignment in time. arXiv preprint arXiv:1810.08646 (2018)
56. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
57. Sironi, A., Brambilla, M., Bourdis, N., Lagorce, X., Benosman, R.: Hats: Histograms of averaged time surfaces for robust event-based object classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1731–1740 (2018)
58. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
59. Tramèr, F., Kurakin, A., Papernot, N., Goodfellow, I., Boneh, D., McDaniel, P.: Ensemble adversarial training: Attacks and defenses. arXiv preprint arXiv:1705.07204 (2017)
60. Viale, A., Marchisio, A., Martina, M., Masera, G., Shafique, M.: Carsnn: An efficient spiking neural network for event-based autonomous cars on the loihi neuromorphic research processor. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–10. IEEE (2021)
61. Wikipedia: Event camera — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Event_camera (2021)
62. Wikipedia: Shear mapping — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Shear_mapping (2021)
63. Wu, Y., Deng, L., Li, G., Zhu, J., Shi, L.: Spatio-temporal backpropagation for training high-performance spiking neural networks. Frontiers in neuroscience **12**, 331 (2018)
64. Wu, Y., Deng, L., Li, G., Zhu, J., Xie, Y., Shi, L.: Direct training for spiking neural networks: Faster, larger, better. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 1311–1318 (2019)
65. Wu, Z., Zhang, H., Lin, Y., Li, G., Wang, M., Tang, Y.: Liaf-net: Leaky integrate and analog fire network for lightweight and efficient spatiotemporal information processing. IEEE Transactions on Neural Networks and Learning Systems (2021)
66. Yao, Z., Gholami, A., Keutzer, K., Mahoney, M.W.: Pyhessian: Neural networks through the lens of the hessian. In: 2020 IEEE International Conference on Big Data (Big Data). pp. 581–590. IEEE (2020)
67. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)
68. Zhang, X., Qin, H., Ding, Y., Gong, R., Yan, Q., Tao, R., Li, Y., Yu, F., Liu, X.: Diversifying sample generation for accurate data-free quantization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15658–15667 (2021)
69. Zheng, H., Wu, Y., Deng, L., Hu, Y., Li, G.: Going deeper with directly-trained larger spiking neural networks. arXiv preprint arXiv:2011.05280 (2020)
70. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 13001–13008 (2020)
71. Zoph, B., Cubuk, E.D., Ghiasi, G., Lin, T.Y., Shlens, J., Le, Q.V.: Learning data augmentation strategies for object detection. In: European conference on computer vision. pp. 566–583. Springer (2020)