SUPPLEMENTARY INFORMATION FOR - DIET-SNN: A LOW-LATENCY SPIKING NEURAL NETWORK WITH DIRECT INPUT ENCODING & LEAKAGE AND THRESHOLD OPTIMIZATION

Anonymous authors
Paper under double-blind review

1 DATASETS

CIFAR: The CIFAR10 and CIFAR100 datasets Krizhevsky et al. (2009) consists of 32×32 pixel images with RGB channels in 10 and 100 classes, respectively. Each dataset is divided into $50,000$ training and $10,000$ test images. The pre-processing involves normalizing the pixlel values with mean and standard deviation, zero-padding of 4 pixels on all four sides, 32×32 random crop, and horizontally flipping the image with a probability of 0.5.

ImageNet: The dataset Deng et al. (2009) consists of approximately 1.2 million high-resolution training images and $50,000$ test images in 1000 classes. The pre-processing involves normalizing the pixel values with mean and standard deviation, 224×224 random crop, and horizontally flipping the image with a probability of 0.5.

2 NETWORK ARCHITECTURE

The VGG and ResNet architectures are modified slightly to minimize the loss during ANN-SNN conversion. The ResNet architecture is appended with 3 plain convolutional layers of 64 filters after the input layer Sengupta et al. (2019). Both the architectures employ average pooling (2×2) and the basic block in ResNet uses a stride of 2 when the number of filters increase as described below. The shortcut connection in ResNet uses 1×1 convolutions in basic blocks where the number of filters in input and output differ (see Fig. 1). For ANN, ReLU is used in place of LIF and the probability of dropout is higher compared to SNN. The activations are already sparse in SNN due to lower number of timesteps, therefore smaller dropout probability is required for stable training. The number of operations in ANN in each layer for both VGG and ResNet architectures is shown in Fig. 2. The number of operations in corresponding SNN depends on the spike rate in each layer and is described by equation 18 in the main text. The layerwise spike rate for ResNet20 on CIFAR datasets is shown in Fig. 3.

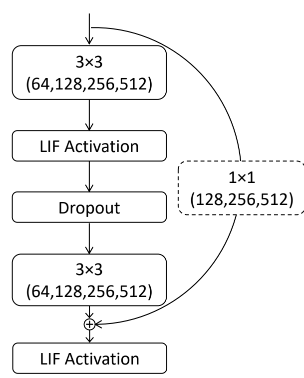

Figure 1: ResNet basic block

VGG6: {64, A, 128, 128, A}, Linear
VGG16: {64, D, 64, A, 128, D, 128, A, 256, D, 256, D, 256, A, 512, D, 512, D, 512, A, 512, D, 512, D, 512}, Linear
ResNet20: {64, D, 64, D, 64, A, 64BB, 64BB, 128BB (/2), 128BB, 256BB (/2), 256BB, 512BB (/2), 512BB}

BB: basic block, Linear: {4096, D, 4096, D, num_classes}, D: Dropout (ANN: 0.2-0.5, SNN: 0.1), A: Average Pooling (kernel size = 2×2)

REFERENCES

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*,

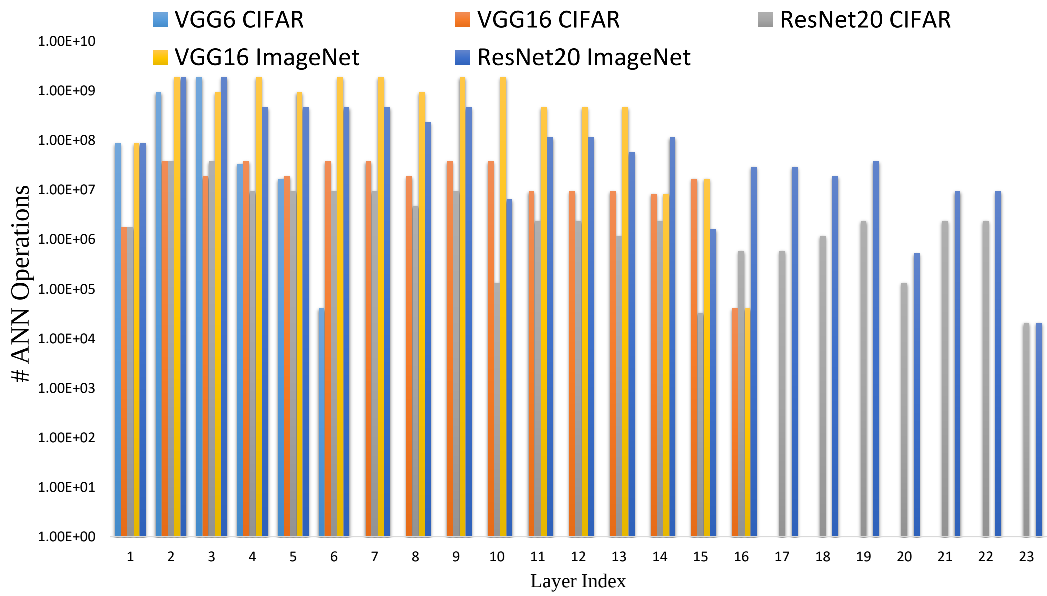

Figure 2: Layerwise number of operations in ANN for different architectures

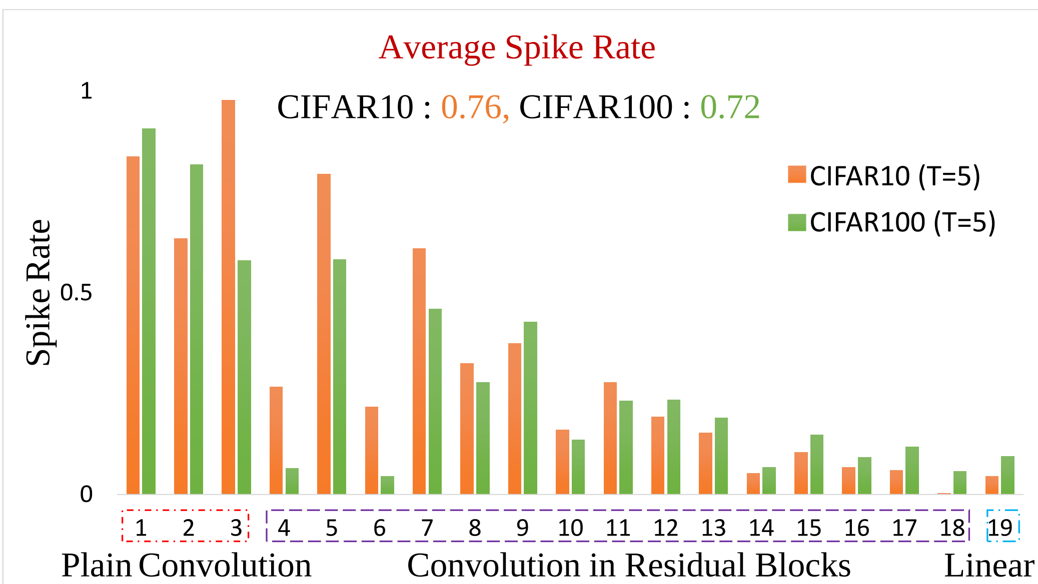

Figure 3: Layerwise spike rate for ResNet20 during inference over entire test-set. Average spike rate is calculated as total ${}^{\#spikes}/_{\#neurons}$ for the entire test set.

pp. 248–255. Ieee, 2009.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, and Kaushik Roy. Going deeper in spiking neural networks: Vgg and residual architectures. *Frontiers in neuroscience*, 13, 2019.