Comparison of Artificial Neural Networks (ANN) and Spiking Neural Networks (SNN)

Rishona Daniels and Ido Becher

February 11, 2025

Abstract

Spiking neural networks (SNN) have gained attention due to their potential as third-generation neural networks. They have the potential to be used to build low power deep neural network architectures. In this project, the performance of SNNs has been compared with standard Artificial Neural Network (ANN) architectures on image recognition tasks with metrics like accuracy, training time, inference time, and model size. A hybrid method that combines a pre-trained vision transformer (ViT) with SNN has also been implemented. The study highlights the potential of SNNs in low-power, event-driven applications and discusses the future of hybrid AI models. The codes used in this work can be found at `https://github.com/rishonadaniels/Comparasion_of_ANN_and_SNN`

1 Introduction

Neural networks have been instrumental in the advancement of AI applications such as computer vision [1] [2], speech processing [3] [4], etc. ANNs are loosely inspired by biological neurons and are widely used in deep learning. However, ANNs require tremendous computational power and resources for training. OpenAI's GPT3 is estimated to have consumed approximately 190,000 kWh for training [5]. In comparison, the brain is considered to be the most efficient computer and can perform all its operations in less than 20 Watts of power [6]. SNNs attempt to bridge the gap between biology and ANNs by more closely mimicking the functioning of biological neurons through spike-based processing [8].

In ANNs, activation functions such as Rectified Linear Unit (ReLU) and Gaussian Error Linear Unit (GELU) are used to add non-linearity to the networks. However, these functions are not generally observed in biological brains. SNNs use activation functions derived from biological neuron models. There are several models of biological neurons with varying complexity like the Hodgkin Huxley model, Izhekevich model, Leaky Integrate and Fire (LIF) model, etc. Of these, the LIF model is considered suitable for deep learning applications due to its relative simplicity as compared to other bio-inspired neuron models. The LIF model can be described by the following equations[7];

$$U[t+1] = \underbrace{\beta U[t]}_{\text{decay}} + \underbrace{WX[t+1]}_{\text{input}} - \underbrace{R[t]}_{\text{reset}} \tag{1}$$

where $U[t]$ is the membrane potential, W is a learnable weight, β is the decay rate and $R[t]$ is the reset potential.

The LIF is said to produce an output spike $S[t]$ if the membrane potential exceeds a particular threshold value;

$$S[t] = \begin{cases} 1, & \text{if } U[t] > U_{\text{thr}} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

1.1 Motivation

Although ANNs have achieved remarkable success, they suffer from high computational and energy costs. SNNs, being biologically inspired, offer a promising alternative for energy-efficient computing. In

SNNs, data are represented as discrete ON-OFF events called spikes. These spikes encode information about the input data, unlike ANNs, in which data is represented as continuous floating-point values. Since SNNs use binary spike trains, the memory required for the hardware implementations of these networks is lower. However, the activation functions used in SNNs are non-differentiable and hence techniques like surrogate gradients and backpropagation through time are required for training [9, 10]. A potential solution is to use a hybrid model in which the model is pre-trained using ANN methods and then converted to an SNN for inference.

1.2 Objective of the Study

This project aims to compare ANNs and SNNs in terms of accuracy, training and inference time and model size on the benchmark datasets MNIST, Fashion-MNIST and CIFAR-10. The objective is to try and answer the question, "Can current frameworks of SNNs deliver it's promise of low-power computing with sufficient accuracy?". We explore whether SNNs can overcome some of the challenges faced by ANNs and whether they represent the next breakthrough in AI.

1.3 Previous Work

There are existing studies that explore the comparison between ANN and SNN models [11]. However, our work aims to evaluate their performance on multiple datasets including metrics like model size and training and inference time that have not been compared in previous studies. Additionally, we investigate a hybrid model that combines aspects of both paradigms to leverage their respective advantages.

2 Methodology

In order to compare ANNs and SNNs, standard deep neural network architectures namely feedforward multilayer perception (MLP), convolutional neural network (CNN) and the recent Vision Transformer (ViT) have been implemented with the PyTorch library. For the ANN case, the ReLU activation function has been used and in the ViT case, the GELU activation function has been used in the above architectures. For the SNN case, the same architectures have been used and the ReLU and GELU activations have been replaced by the leaky activation function from the snnTorch library [12]. This activation function has a tunable hyperparameter called beta (β) represents the decay rate.

2.1 Network Architectures

We implemented and compared the following architectures:

- **Feedforward ANN:** Fully connected layers trained using backpropagation.
- **CNN:** Convolutional layers for feature extraction and classification.
- **SNN:** Feedforward network with spiking neurons using the Leaky Integrate-and-Fire (LIF) model trained using surrogate gradient descent.(Fig. 1)
- **Vision Transformer (ViT):** A pre-trained transformer-based model fine-tuned for image classification tasks. The ViT leverages self-attention mechanisms to process image patches, providing state-of-the-art accuracy on complex datasets like CIFAR-10. We specifically used the ViT-B/16 architecture with pre-trained weights from ImageNet.
- **Hybrid ViT-SNN Model (ViT-SNN):** A novel architecture combining Vision Transformer (ViT) with Spiking Neural Networks (SNNs). The model is based on a pre-trained ViT-Tiny model, where standard activation functions such as ReLU and GELU were replaced with spiking neurons using the Leaky Integrate-and-Fire (LIF) model. This hybrid approach aims to leverage the powerful feature extraction capabilities of transformers while benefiting from the energy efficiency and biologically inspired processing of SNNs.

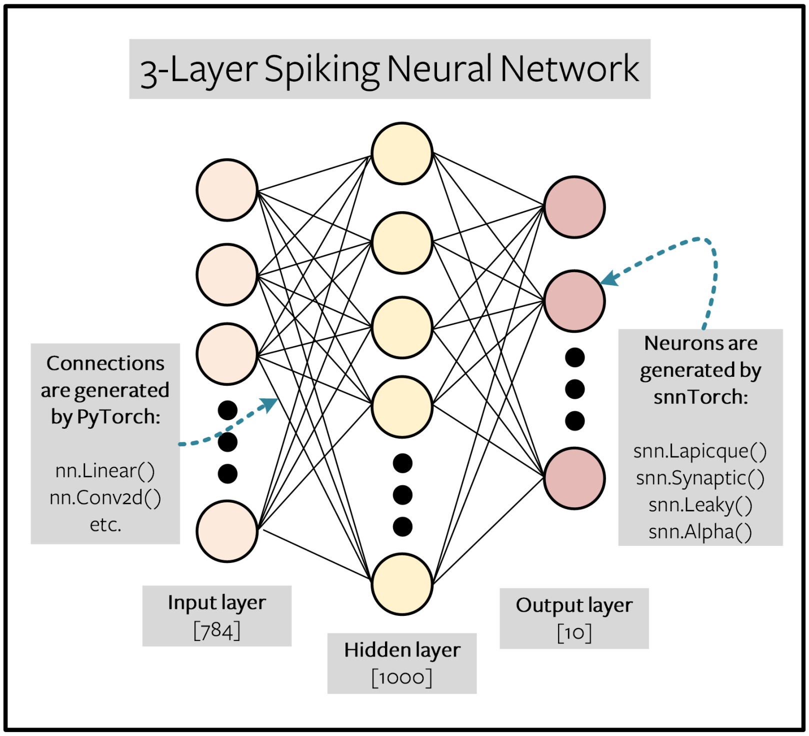

Figure 1: Feedforward network model. In the ANN case, ReLU activations are used and in the SNN case, LIF activation is used. [12]

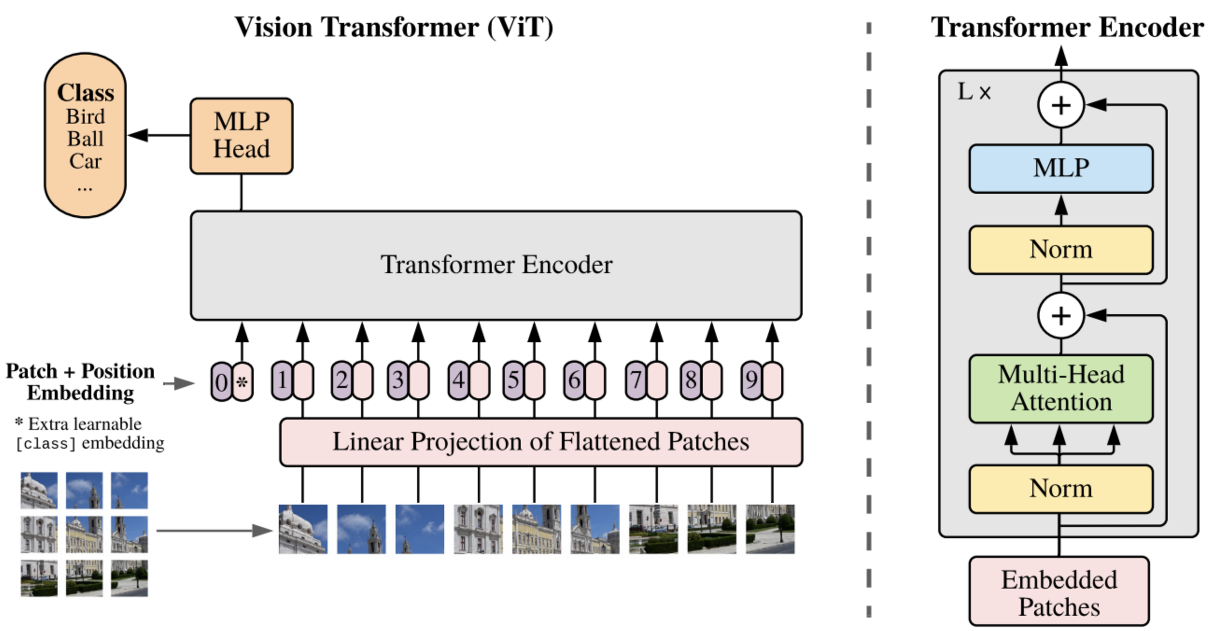

Figure 2: Overview of Vision Trasnformer. The image is split into patches of fixed sizes and linearly embedded with position embeddings. The resulting sequence of vectors is fed to a standard transformer. [13]

2.2 Loss Functions and Training

For ANNs and CNNs, cross-entropy loss with stochastic gradient descent and Adam optimization were used. For SNNs, training was conducted using surrogate gradient descent, which approximates gradients to allow backpropagation through discrete spike events. Each image was given for a set number of time steps as SNNs need their data as temporal sequences over a period of time. The number of time steps is also a hyperparameter in SNNs.

3 Experiments and Results

3.1 Datasets

We conducted experiments with the following datasets:

- **MNIST**: Handwritten digit recognition with 10 classes where the training set consists of $60,000$ images and test set consists of $10,000$ images, where each image is of size 28×28 pixels and each pixel is an 8-bit greyscale value.

- **Fashion-MNIST**: Clothing classification with 10 classes where the training set consists of $60,000$ images and test set consists of $10,000$ images, where each image is of size 28×28 pixels and each pixel is an 8-bit greyscale value.

- **CIFAR-10**: Object classification with 10 classes where the training set consists of 60,000 images and test set consists of 10,000 images, where each image is of size 32×32 pixels and each pixel has 3 channels for RGB value. Each pixel is 3 bytes (8 bits $\times 3 = 24$ bits $= 3$ bytes)

3.2 Experimental Setup

For each model, we used functions from the PyTorch library. For the SNNs, we used functions from the snnTorch library. Each model was trained for multiple epochs, and hyperparameters such as model size, batch size, learning rate, and number of epochs were tuned. For SNNs, the decay rate β and the number of time steps were additional hyperparameters. Accuracy, training time, inference time, and model size were used as performance metrics.

3.3 Results Comparison

Tables 1, 2 and 3 summarizes the accuracy and efficiency of different models for the MNIST, Fashion MNIST and CIFAR-10 datasets respectively.

For the MNIST dataset, all models performed exceptionally well in terms of accuracy as its a very simple task. However, the CNN and SNN trained quite quickly as compared to the other models. The ViT took a very long time to fine-tune. The SNN train faster than the ANN. The SNN model was $3\times$ larger than the ANN.

For Fashion MNIST, the accuracy of the SNN was quite low as compared to the other models. The CNN by far outperformed all other models including the fine-tuned ViT. As in the MNIST case, the model size and training time were significantly larger than the other models.

For CIFAR-10, the feedforward ANN was not able to give satisfactory accuracy. However, a relatively small CNN model was able to give very high accuracy. The pre-trained and fine-tuned ViT gave the best result but it took a very long time for fine tuning. In the hybrid SNN ViT model, the pre-trained ViT was fine-tuned on CIFAR-10 by replacing its GELU activations with the LIF activation. It took significantly longer to train and only gave accuracy of 73%. We were only able to train it for 5 epochs. We believe, it can get much higher accuracy with better hyperparameter selection and training time.

Model	Accuracy	Training time	model size
ANN	97.5%	137.82 sec	0.39 MB
CNN	99.1%	62.81 sec	1.61 MB
SNN	94.22%	67.88 sec	3.04 MB
ViT	99.28%	4237.52 sec	327.38 MB

Table 1: Comparison of Model Performance for MNIST dataset

Model	Accuracy	Training time	model size
ANN	89.25%	294.02 sec	1.04 MB
CNN	93.08%	285.56 sec	3.69 MB
SNN	86.2%	835.68 sec	3.04 MB
ViT	90.01%	4394.52 sec	327.38 MB

Table 2: Comparison of Model Performance for Fashion MNIST dataset

Model	Accuracy	Training time	model size
ANN	57.68%	921.66 sec	90.83 MB
CNN	91.08%	2172.99 sec	12.42 MB
ViT	93.45%	8424.16 sec	327.38 MB
ViTSNN	73.38%	12267.26 sec	NA

Table 3: Comparison of Model Performance for CIFAR-10 dataset

4 Conclusion and Future Work

The experimental results suggest that SNNs do not currently offer a significant reduction in model size or training time. However, with proper optimization, they can achieve accuracy levels comparable to traditional neural networks. This study utilized static datasets, which were transformed into time sequences by repeating the images. It is believed that the use of spiking datasets could enhance the performance of SNNs in comparison to ANNs. Further research is required to investigate effective approaches for converting ANNs to SNNs. Moreover, hybrid ANN-SNN models are considered a promising direction for the future of AI applications.

5 Ethics Statement

1. Introduction
Student names: Rishona Daniels, Ido Becher
Comparison of Artificial Neural Networks (ANN) and Spiking Neural Networks (SNN) This project compares Artificial Neural Networks (ANNs), Spiking Neural Networks (SNNs), Convolutional Neural Networks (CNNs), and Vision Transformers (ViT) models to evaluate their performance across different datasets. The goal is to analyze their architectural differences, training methods, accuracy, model size and training time. By conducting experiments on datasets such as MNIST, Fashion-MNIST, and CIFAR-10, the project aims to identify the strengths and weaknesses of each model and explore potential hybrid approaches for improved AI efficiency.
2. Large Language Model Answer

(Continued) 2. Large Language Model Answer
2a. List 3 types of stakeholders that will be affected by the project. • **AI Researchers:** AI and machine learning researchers who study and develop neural network models. • **Companies Deploying AI Models:** Organizations that may implement these models in AI-driven applications, particularly in low-power computing and event-driven processing. • **General Users of AI Applications:** Individuals who use AI-powered systems based on these models, including consumers of edge AI applications and automated decision-making systems.
2b. What will an explanation that is given to each stakeholder look like? • **AI Researchers:** "This project provides a comparative analysis of neural network architectures, including SNNs, ANN-based models, and hybrid approaches. It offers insights into training efficiency, model accuracy, and energy consumption, helping researchers optimize AI designs for different applications." • **Companies:** "The study evaluates the practical advantages and limitations of various AI models, helping businesses make informed decisions when adopting energy-efficient AI solutions in edge computing and real-time applications." • **General Users:** "By exploring AI model efficiency, this research can lead to better, faster, and more sustainable AI-powered technologies, reducing energy costs and improving user experience in AI applications such as smartphones and IoT devices."
2c. Who is responsible for giving the explanation to each stakeholder? • **AI Researchers:** Responsibility lies with the project authors and the research community to publish findings transparently through papers, conferences, and open-source platforms. • **Companies:** Research teams and technical consultants should communicate findings to businesses through technical reports, whitepapers, or direct industry collaborations. • **General Users:** Public outreach should be conducted through simplified explanations in blogs, presentations, or media discussions to ensure broader awareness of AI advancements.
3. Reflect on the AI Output
What do you think needs to be added/changed in the Generative AI responses to make the explanation more ethical? The AI's response does not include potential biases that could be included in the model evaluation. The documentation must be clear so that each of the stakeholders can properly understand the positive and negative impacts of the evaluation. The limitations of the study, such as the exclusion of some kinds of architecture, are also missing. Here, we have studied the vision task, which is of particular interest to stakeholders like law enforcement agencies and governments. Privacy concerns are also of ethical importance.

Table 4: Ethics statement.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097–1105, 2012.

[2] R. Girshick, J. Donahue, T. Darrell and J. Malik, ”Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 580-587.

[3] A. Graves and N. Jaitly, ”Towards End-To-End Speech Recognition with Recurrent Neural Networks,” *Proceedings of the 31st International Conference on Machine Learning*, Vol. 32, No. 2, pp. 1764-1772, June 2014.

[4] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, ”Listen, attend and spell: A neural network for large vocabulary conversational speech recognition,” *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4960–4964.

[5] L. F. W. Anthony, B. Kanding, and R. Selvan, “Carbontracker: Tracking and predicting the carbon footprint of training deep learning models,” arXiv:2007.03051, 2020.

[6] W. B. Levy and V. G. Calvert, “Computation in the human cerebral cortex uses less than 0.2 Watts yet this great expense is optimal when considering communication costs,” *bioRxiv*, April 2020.

[7] J. K. Eshraghian et al., ”Training Spiking Neural Networks Using Lessons From Deep Learning,” *Proceedings of the IEEE*, Vol. 111, No. 9, pp. 1016-1054, September 2023.

[8] G. Gallego et al., ”Event-Based Vision: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, no. 1, pp. 154-180, 1 Jan. 2022,

[9] E. Neftci et al., ”Surrogate Gradient Learning in Spiking Neural Networks,” *IEEE Transactions on Neural Networks*, 2019.

[10] A. Tavanaei et al., ”Deep Learning in Spiking Neural Networks,” *Neural Networks*, 2019.

[11] L. Deng et. al, ”Rethinking the performance comparison between SNNS and ANNS,” *Neural Networks*, Vol. 121, pp. 294-307, 2020.

[12] SNNTorch Documentation, 2024. [Online]. Available: https://snntorch.readthedocs.io/en/latest/index.html. [Accessed: Feb. 11, 2025].

[13] A. Dosovitskiy et. al, ”An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *9th International Conference on Learning Representations, ICLR 2021.*