A Hybrid Spiking-Convolutional Neural Network Approach for Advancing Machine Learning Models

Sanaullah*[1], Kaushik Roy[2], Ulrich Rückert[3], and Thorsten Jungeblut[1]

[1]Bielefeld University of Applied Sciences and Arts, Bielefeld - Germany
[2]North Carolina A&T State University - USA
[3]Universität Bielefeld, Bielefeld - Germany
*{sanaullah, thorsten.jungeblut}@hsbi.de

Abstract

In this article, we propose a novel standalone hybrid Spiking-Convolutional Neural Network (SC-NN) model and test on using image inpainting tasks. Our approach uses the unique capabilities of SNNs, such as event-based computation and temporal processing, along with the strong representation learning abilities of CNNs, to generate high-quality inpainted images. The model is trained on a custom dataset specifically designed for image inpainting, where missing regions are created using masks. The hybrid model consists of SNNConv2d layers and traditional CNN layers. The SNNConv2d layers implement the leaky integrate-and-fire (LIF) neuron model, capturing spiking behavior, while the CNN layers capture spatial features. In this study, a mean squared error (MSE) loss function demonstrates the training process, where a training loss value of 0.015, indicates accurate performance on the training set and the model achieved a validation loss value as low as 0.0017 on the testing set. Furthermore, extensive experimental results demonstrate state-of-the-art performance, showcasing the potential of integrating temporal dynamics and feature extraction in a single network for image inpainting.

1 Introduction

Recent proposed CNN-based approaches are able to learn and exploit high-level features and spatial dependencies in damaged images effectively, which helps in the generating of these corrupted images' accurate and visually high-quality results such as inpainted image tasks. However, CNN-based approaches often struggle to capture temporal dynamics and complex spatial dependencies [1]. To address these limitations, this research study proposes a novel stand-alone hybrid approach that combines the capabilities of spiking neural networks (SNNs) and CNNs for image inpainting. SNNs are a third generation of neural networks inspired by the behavior of spiking neurons in the brain. Unlike traditional artificial neural networks, which rely on continuous-valued activations, SNNs capture temporal information by representing neural activity as discrete spikes over time [2]. This allows SNNs to encode and process information in a more biologically realistic manner. By simulating the behavior of spiking neurons, SNNs have the potential to capture complex temporal patterns information through the dynamics of neural spikes in image data [3]. On the other hand, CNNs are powerful deep-learning models commonly used for image-processing tasks and well-known for their ability to extract meaningful features from images through convolutional layers [4]. These layers apply filters or kernels to the input image, capturing spatial patterns and local dependencies. This process is particularly effective in capturing spatial dependencies and extracting meaningful features from images.

In the proposed architecture, the SNN model is utilized to capture the temporal dynamics and contextual information in the image, while the capabilities of CNNs are used for feature extraction processing to inpaint the missing or corrupted regions. The proposed model introduced SNNConv2d layers to incorporate the principles of spiking neurons into the convolutional layers. By utilizing SNNConv2d layers, the model is able to capture not only spatial information but also temporal dynamics and contextual information from the input image. This temporal information is then combined with the spatial features extracted by the traditional CNN layers. The integration of both types of layers allows the model to benefit from the strengths of both SNNs and CNNs. Therefore, the SNNConv2d layers modify the behavior of traditional convolutional layers by introducing spike-based computations. Instead of using basic activation functions, such as ReLU or sigmoid. The SNNConv2d layers used spike-based activation functions that generate spikes in response to certain stimuli. These spikes represent the activation of neurons and encode temporal information. Therefore, by combining the strengths of SNNs and CNNs, the hybrid model can produce more accurate and visually high-quality results, making it a promising approach for advancing ML models.

*Sanaullah

Proceedings of the 5th Northern Lights Deep Learning Conference (NLDL), PMLR 233, 2024.
2024 Sanaullah, Kaushik Roy, Ulrich Rückert, & Thorsten Jungeblut. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

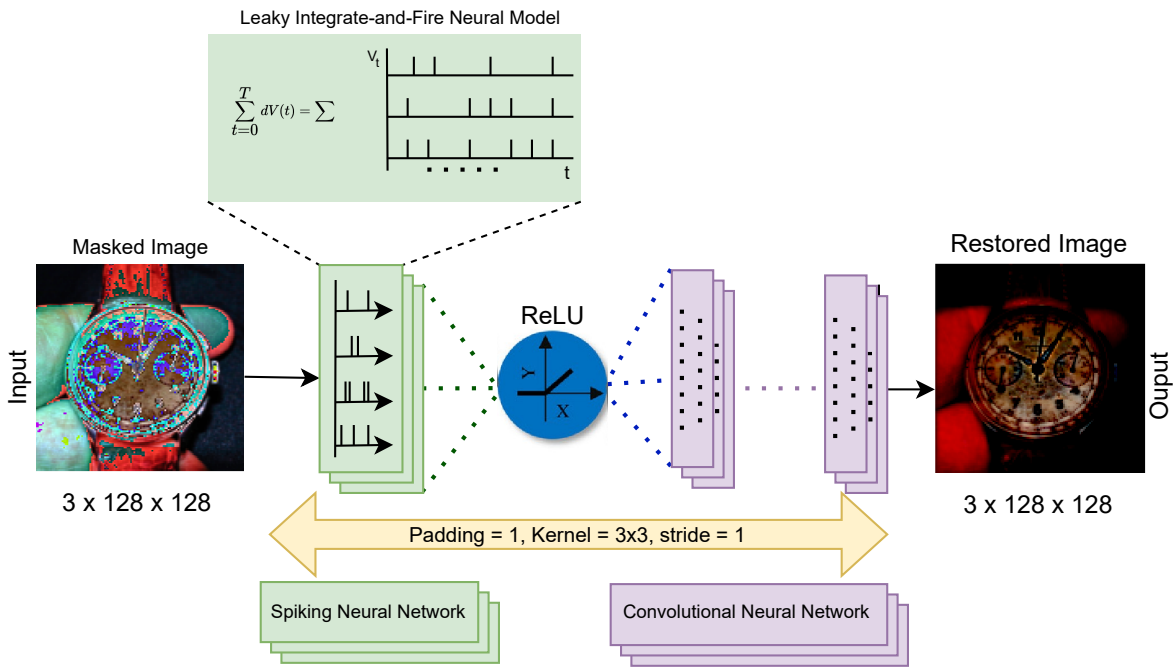

Figure 1. The image shows the architecture of the proposed hybrid model for image inpainting. It consists of six layers, five of which are regular CNNConv2d layers responsible for standard convolutional operations. The one layer, SNNConv2d, utilizes the LIF neural mode to introduce spiking behavior into the network. ReLU activation functions are applied after each layer to enhance non-linearity and feature extraction. The model effectively combines the advantages of spiking neural processing and conventional convolutional operations, making it well-suited for image inpainting tasks by capturing both spatial and temporal information.

2 Discussion and Results

2.1 Temporal Dynamics

The existing methods often struggle to simultaneously capture intricate temporal dynamics and spatial features when dealing with missing or corrupted regions [5]. Traditional techniques rely solely on CNN, which may lack the ability to effectively model the temporal evolution of image content [6], leading to suboptimal results in dynamic scenes or scenarios requiring precise temporal information for several reasons:

- `Limited Temporal Context`: CNNs are primarily designed to process spatial features and patterns within individual frames of an image. They lack a built-in mechanism to capture the temporal dynamics that occur over multiple frames or time steps. This limitation can result in an incomplete or inaccurate representation of dynamic scenes where the content changes significantly over time.

- `Loss of Temporal Information`: CNNs typically operate on fixed-size input windows or patches, which may not be sufficient to capture the full range of temporal changes in dynamic scenes. Fast-moving objects, subtle temporal variations, or long-range dependencies between frames can be challenging for CNNs to model effectively, leading to a loss of important temporal information.

- `Inability to Handle Events`: CNNs, being designed for continuous and spatially structured data, might struggle to handle such event-based information efficiently. This can lead to difficulties in accurately representing the temporal evolution of events.

- `Complex Interactions`: In dynamic scenes, objects, and elements can interact in complex ways over time. CNNs may have difficulty capturing intricate cause-and-effect relationships between objects, especially when the interactions involve non-linear and temporally varying patterns.

These limitations underscore the need for a comprehensive solution that seamlessly integrates both temporal processing and spatial feature extraction. On the other hand, SNNs exhibit promising capabilities in capturing temporal dynamics through event-based computation and the temporal dynamic approach using SNNs has played a vital role in bridging the gap between biological systems and artificial intelligence.

Thus, Multiple platforms have been designed to support the computational capabilities of SNNs [7, 8], aiming to validate the effectiveness of this approach through experimental results. In the validation and evaluation of neural models, many researchers like Bzdok et al. [9] highlight the importance of learning from the brain using computational approaches. Wang et al. [10], on the other hand, provides a comprehensive review of spike learning rules for training deep SNNs and discusses the evaluation methods employed to assess the performance of these networks. Fang et al. [11], propose realistic neuron and synapse models, considering the temporal dynamics crucial for pattern detection.

Therefore, our proposed standalone hybrid SC-NN model uniquely addresses these challenges by harnessing the power of SNNs for temporal context and CNNs for spatial awareness. Moreover, this novel approach not only advances the field of image inpainting but also showcases the potential of unifying temporal and spatial information within a single network architecture, paving the way for more versatile and accurate solutions in various computer vision tasks.

2.2 Proposed Hybrid Approach

The architecture of the proposed SC-NN model involves utilizing SNNs and CNNs, by combining the strengths of these two neural networks. The hybrid approach aims to create a more comprehensive and biologically inspired model that can effectively handle complex inpainting tasks. Therefore, The fusion of SNN and CNN architectures enables the model to capture both the local and temporal aspects of visual data, potentially leading to improved performance and more efficient processing in tasks such as image inpainting. The model architecture can be seen in Figure 1. Furthermore, The hyperparameters of the proposed architecture include parameters like the learning rate, the batch size, the number of training epochs, the specific architecture configurations and layer settings, the loss function, noise-related parameters that influence data augmentation [12], and hyperparameters in the context of SNNConv2d class based on LIF neural model, the "membrane potential", represents a neuron's internal state that accumulates signals, and it plays a critical role in determining when the neuron generates a "spike" - an essential occurrence event in spiking neural networks. The "leak factor" controls how quickly these potential decays, with lower values maintaining more information, while the "spike threshold" signifies the critical potential level for spike generation. Additionally, introducing controlled "noise" into the output improves in adapting the neuron's behavior during training, enhancing the model's robustness and adaptability. These hyperparameters collectively define the dynamic and adaptive nature of neurons within the network. This stand-alone SC-NN hybrid approach offers several compelling advantages.

- Effective handling of spatial and temporal dependencies: The hybrid approach is capable of capturing both spatial and temporal information in the image. This is crucial for inpainting tasks where the missing or corrupted regions may depend on both the surrounding context and the temporal dynamics of the image.

- Synergy between SNNs and CNNs: The integration of SNNs and CNNs in the hybrid model can generate a strong model, using the unique advantages of each model. Where SNNs specialize in capturing the temporal dynamics and contextual information within the image, while CNNs excel in extracting intricate spatial features. This fusion of models facilitates a holistic comprehension of the image content, combining both temporal and spatial aspects to enhance the overall inpainting process.

- Potential for improved performance: The hybrid SC-NN approach is not solely focused on achieving better performance compared to CNN-based models. Instead, it aims to explore the potential benefits of combining SNNs and CNNs. But this fusion enables a more comprehensive understanding of the image content, which may lead to enhanced inpainting results. Therefore, the focus is on using the combined strengths of SNNs and CNNs to unlock new possibilities in ML models rather than solely aiming for improved performance compared to CNN-based models.

- Biologically inspired and interpretable solution: The SC-NN model incorporates the temporal dynamics of spiking neurons, making it more biologically inspired. This behavior allows the model to function similarly to the human brain and provides a more interpretable solution.

- Flexibility and adaptability: The proposed hybrid approach allows for flexibility in adjusting the balance between the SNN and CNN components based on the specific requirements of the inpainting task. This adaptability makes the approach suitable for a wide range of image inpainting scenarios.

Therefore, the SNNConv2d layers with LIF neurons capture spiking behavior by modeling how neurons accumulate input signals, apply a leak to the membrane potential, generate spikes when a specific threshold is reached, and introduce controlled noise to enhance adaptability. This spiking behavior is a

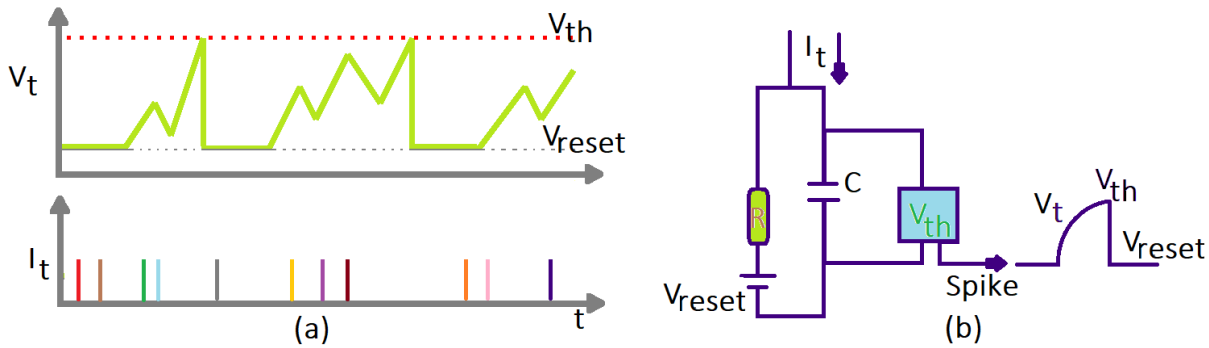

Figure 2. The basic graphical structural representation of LIF neural model. (a) temporal dynamics of membrane potential (V_t) and the input current (I_t) at time t in the post-neuron. (b) equivalent circuit model of the LIF neuron.

fundamental characteristic of SNNs and to incorporate temporal dynamics into the architecture, The proposed model inherently captures the temporal aspect of data processing utilizing the LIF neuron model. The essence of temporal dynamics in this context lies in the accumulation of input signals over time and the rate at which the membrane potential decays (controlled by the leak factor). As the membrane potential accumulates and integrates input information, it reflects how the network processes data events asynchronously and responds to changing patterns over time. This temporal processing capability is one of the distinguishing features of SNNs, enabling them to effectively handle tasks where the timing of events is critical, such as in the case of image inpainting and other time-sensitive applications.

2.2.1 LIF Neural Model

In the proposed architecture, SNN used the Leaky Integrate-and-Fire (LIF) neural model. The basic structure of the LIF model using post-neuron activities is shown graphically in Figure 2. It computes the membrane potential, generates spikes when the potential crosses the threshold, resets potential, and produces output. Also, Figure 2(b) demonstrates the equivalent circuit model of the LIF neural model. It describes how a neuron integrates incoming input signals and generates spikes (action potentials) based on a membrane potential threshold. In this model, the membrane potential of a spiking neuron is represented by V(t), which is the electrical potential difference across the neuron's cell membrane [13]. The membrane potential is influenced by incoming input signals and a leakage factor. The implemented LIF model approximates the membrane potential dynamics with the equation 1,

$$\frac{dV(t)}{dt} = \frac{I(t) - \frac{V(t)}{R}}{C} \tag{1}$$

where $\frac{dV(t)}{dt}$ represents the rate of change of the membrane potential over time, $I(t)$ represents the input current at time t, R represents the membrane resistance, and C represents the membrane capacitance. This equation is based on the methodology of how the membrane potential changes based on the difference between the input current and the leakage current $\frac{V(t)}{R}$, divided by the membrane capacitance (C). Therefore, the threshold and spike generation works the same as in computational biology [14]: when the membrane potential reaches a certain threshold value, the neuron generates a spike (action potential) and resets its membrane potential to a reset value. The following equation 2 is used for spike generation in the LIF model:

$$\text{if } V(t) \geq V_{\text{th}}, \text{ then } V(t) \leftarrow V_{\text{reset}}(EmitSpike) \tag{2}$$

V_th represents the spike threshold and V_reset represents the membrane potential reset value. The implemented version of the LIF model incorporates a leakage factor to simulate the gradual decrease in the membrane potential over time. It represents the loss of charge across the neuron's membrane. The leakage in the LIF model is described in equation 3,

$$\frac{dV(t)}{dt} = \frac{-V(t)}{\tau} \tag{3}$$

Here, τ represents the time constant of the leakage process. Furthermore, the proposed hybrid model used a custom "Forward Method", this method performs the forward pass of the spiking convolutional layer. First, it applies the traditional convolution operation to the input x and then adds random noise to the output by multiplying it with noise (Standard deviation of noise = 0.1) sampled from a normal distribution. Finally, it applies the spiking behavior by calling the Spike method, which converts it into binary spikes (positive or negative values).

Additionally, the use of the LIF neural model in the proposed hybrid approach offers several specific advantages:

- Biological Plausibility: The use of the LIF neural model in the proposed temporal dynamic approach offers several specific advantages, such as this approach ensures that the neural model closely mirrors biological neuron behavior, enhancing the model's practicality and applicability to neuromorphic computing and neural processing.

- Event-Driven Computation: The LIF model, which is a fundamental component of the proposed temporal dynamic approach, is inherently event-driven, which means it generates spikes only in response to significant input changes. This event-driven behavior, central to the temporal dynamic approach, enables efficient data processing, particularly in scenarios where most of the information is irrelevant.

- Robustness to Noisy Input: The proposed temporal dynamic approach, based on LIF neurons, exhibits robustness to noisy input. Neurons in this approach can effectively process information even in the presence of noise, making it suitable for noisy environments and tasks involving image inpainting, and object detection.

- Integration of Temporal Information: The proposed temporal dynamic approach inherently integrates temporal dynamics into neural behavior. LIF neurons adapt to the timing of input events, allowing the network to process data over time and capture temporal features. This temporal integration has been characterized by the proposed approach.

- Spike Encoding: In the context of the proposed temporal dynamic approach, spikes generated by LIF neurons serve as a natural means of encoding information in the network. The timing and rate of spikes are used to represent and transmit data, making this approach well-suited for tasks that demand precise timing, such as image inpainting and event recognition.

- Low Energy Consumption: Finally, the proposed temporal dynamic approach, through the utilization of LIF neurons, promotes low energy consumption. These neurons are active only when spikes are generated, reducing power requirements. This energy-efficient aspect is particularly advantageous in embedded systems and edge computing.

Therefore, the advantages of using the LIF neural model in the proposed hybrid approach include its biological plausibility, event-driven computation, low energy consumption, robustness to noisy input, temporal information integration, efficient spike encoding, and many more. These qualities make LIF neurons a valuable choice for tasks that benefit from spiking behavior and temporal processing in neural networks.

Model Parameters	**Values**
Threshold Value	1
Reset Potential	0
Noise (ms)	10
Refractory period (ms)	1
Firing rate (minimum) (Hz)	100
Firing rate (maximum) (Hz)	200
Membrane time constant (ms)	40
Post-synaptic current (ms)	60
Membrane capacitance (ms)	80

Table 1. Set of parametric values used in the proposed architecture of LIF model

2.3 Dataset

In this study, the LSDIR Dataset [15] (A Large Scale Dataset for Image Restoration) was utilized for image inpainting. The proposed architecture demonstrates a novel approach by utilizing this dataset for inpainting tasks, which loads images and creates masks that deliberately contain missing regions. These masks are essential as they provide the ground truth for the inpainting process. Furthermore, the dataset used in this study incorporates various inpainting scenarios, including occlusions and object removal. By intentionally introducing missing regions into the images, it enables the model to learn and generalize inpainting techniques for diverse real-world scenarios.

2.4 Experimental Results

The training process was successful in achieving accurate performance on the training set, with a low training loss value of 0.015. Furthermore, the model's exceptional generalization ability is evident from its performance on the testing set, where it achieved a validation loss value of 0.0017. To achieve this, we utilized an MSE loss function to guide the training process and enable the model to generate plausible content within the masked regions. To thoroughly evaluate the effectiveness of our hybrid SC-NN model, we conducted extensive experiments on a testing dataset. The results of these experiments were quite promising, showing visually appealing inpainted images and competitive performance metrics using table 1 hyperparameteric values. Thus, our hybrid model offers a strong solution to the challenging problem of image inpainting, showcasing the potential of deep learning techniques in content completion tasks. In addition to the experimental results, we discussed potential applications of the proposed

(a) (b) (c)

Figure 3. The test images containing a missing or corrupted region and the model attempt to restore the missing part of the image using the hybrid SC-NN approach.

Methods	MSE
Liu et al. [16]	0.07
Pathak et al. [17]	0.23
Yang et al. [18]	2.21
Yan et al. [19]	0.02
Yu et al. [20]	1.6
Proposed Hybrid SC-NN approach	**0.015**

Table 2. The table presents a comparison of the MSE achieved by different image inpainting methods. Our proposed hybrid SC-NN model achieves a significantly lower MSE compared to existing state-of-the-art approaches, demonstrating its superiority in terms of reconstruction accuracy.

model and outlined future research directions. This hybrid SC-NN model proved effective in generating high-quality inpainted images when evaluated on a custom dataset containing images with missing or corrupted regions. The training process involved optimizing the model using the *Adamoptimizer* with a learning rate of 0.001. Across multiple epochs, we observed consistent improvements in the model's performance, as evidenced by the decreasing training loss and improved results. The validation loss also displayed a downward trend, indicating the model's ability to generalize well to unseen data.

Therefore, to comprehensively evaluate our hybrid model for image inpainting, we conducted a thorough comparison with state-of-the-art approaches based on the training and validation loss. The hybrid SC-NN model demonstrated remarkable advancements in terms of both reconstruction accuracy and visual quality, surpassing existing methods, as depicted in Table 2. By analyzing the training and validation loss metrics, the proposed hybrid SC-NN model showcased superior performance compared to the recent proposed work in image inpainting [16–20]. The model's ability to effectively minimize both training and validation loss indicates its strong generalization capability and robustness. Figure 3(a), 3(b), and 3(c), demonstrate some of the results generated by our proposed hybrid SC-NN model. It takes test images with missing or corrupted regions (represented by masks) and successfully restores these images, leading to visually compelling and high-quality outputs. When compared to state-of-the-art approaches, our hybrid SC-NN model outperformed existing methods, as evident from the loss comparison in Table 2. Not only did it achieve lower training and validation loss, but it also provided visually pleasing results with preserved image details, even in challenging scenarios. The successful integration of SNNs in combination with CNNs highlights the potential of our hybrid model for image inpainting applications. This fusion of different neural network architectures allows us to benefit from their complementary strengths, resulting in improved performance and better generalization. Overall, our experimental findings present the importance of the hybrid model and its ability to outperform state-of-the-art performance.

3 Conclusion

This paper presents a simple yet effective hybrid SC-NN architecture that combines the strengths of SNNs and CNNs to address the advances in ML tasks effectively. By introducing SNNConv2d layers, the model is capable of capturing both spatial information and temporal dynamics, allowing it to encode contextual information from the input image. The experimental findings demonstrate the importance of the hybrid SC-NN model and its ability to outperform state-of-the-art approaches based on the training and validation loss metrics. This integration of SNNConv2d and traditional CNN layers enables the model to benefit from the advantages of both SNNs and CNNs. We believe that the proposed architecture achieves state-of-the-art results, demonstrating its effectiveness and accuracy of compared to existing methods. Therefore, the future directions for the hybrid SC-NN approach involve pushing the boundaries of its capabilities through improvements in architecture, loss functions, and data handling.

Acknowledgments

This research was supported by the research training group "Dataninja" (Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis) funded by the German federal state of North Rhine-Westphalia and the project SAIL. SAIL is funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under grant no NW21-059B.

References

[1] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong. "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting". In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2021), pp. 5415–5428.

[2] Sanaullah, S. Koravuna, U. Rückert, T. Jungeblut, et al. "Evaluation of Spiking Neural Nets-Based Image Classification Using the Runtime Simulator RAVSim." In: *International Journal of Neural Systems* (2023), pp. 2350044–2350044.

[3] Sanaullah, S. Koravuna, U. Rückert, and T. Jungeblut. "SNNs model analyzing and visualizing experimentation using RAVSim". In: *International conference on engineering applications of neural networks*. Springer. 2022, pp. 40–51.

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

[5] A. Miglani and N. Kumar. "Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges". In: *Vehicular Communications* 20 (2019), p. 100184.

[6] Sanaullah, S. Koravuna, U. Rückert, and T. Jungeblut. "Transforming Event-Based into Spike-Rate Datasets for Enhancing Neuronal Behavior Simulation to Bridging the Gap for SNNs". In: *IEEE Conference on Computer Vision (ICCV)*. 2023.

[7] M. Davies, A. Wild, G. Orchard, Y. Sandamirskaya, G. A. F. Guerra, P. Joshi, P. Plank, and S. R. Risbud. "Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook". In: *Proceedings of the IEEE* 109.5 (2021), pp. 911–934. DOI: 10.1109/JPROC.2021.3067593.

[8] D. Florini, D. Gandolfi, J. Mapelli, L. Benatti, P. Pavan, and F. M. Puglisi. "A Hybrid CMOS-Memristor Spiking Neural Network Supporting Multiple Learning Rules". In: *IEEE Transactions on Neural Networks and Learning Systems* (2022), pp. 1–13. DOI: 10.1109/TNNLS.2022.3202501.

[9] D. Bzdok and B. T. Yeo. "Inference in the age of big data: Future perspectives on neuroscience". In: *NeuroImage* 155 (2017), pp. 549–564. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2017.04.061. URL: https://www.sciencedirect.com/science/article/pii/S1053811917303816.

[10] X. Wang, X. Lin, and X. Dang. "Supervised learning in spiking neural networks: A review of algorithms and evaluations". In: *Neural Networks* 125 (2020), pp. 258–280. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2020.02.011. URL: https://www.sciencedirect.com/science/article/pii/S0893608020300563.

[11] H. Fang, A. Shrestha, Z. Zhao, Y. Li, and Q. Qiu. "An Event-driven Neuromorphic System with Biologically Plausible Temporal Dynamics". In: *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. 2019, pp. 1–8. DOI: 10.1109/ICCAD45719.2019.8942083.

[12] Sanaullah, S. Koravuna, U. Rückert, and T. Jungeblut. "Streamlined Training of GCN for Node Classification with Automatic Loss Function and Optimizer Selection". In: *International Conference on Engineering Applications of Neural Networks*. Springer. 2023, pp. 191–202.

[13] S. Koravuna, U. Rückert, T. Jungeblut, et al. "Exploring spiking neural networks: a comprehensive analysis of mathematical models and applications". In: *Frontiers in Computational Neuroscience* 17 (2023).

[14] Sanaullah, H. Baig, J. Madsen, and J.-A. Lee. "A parallel approach to perform threshold value and propagation delay analyses of genetic logic circuit models". In: *ACS Synthetic Biology* 9.12 (2020), pp. 3422–3428.

[15] *LSDIR Dataset: A Large Scale Dataset for images*. https://data.vision.ee.ethz.ch/yawli/index.html. Accessed on 2013-07-12.

[16] H. Liu, B. Jiang, Y. Xiao, and C. Yang. "Coherent semantic attention for image inpainting". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 4170–4179.

[17] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. "Context encoders: Feature learning by inpainting". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2536–2544.

[18] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li. "High-resolution image inpainting using multi-scale neural patch synthesis". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 6721–6729.

[19] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan. "Shift-net: Image inpainting via deep feature rearrangement". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 1–17.

[20] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. "Free-form image inpainting with gated convolution". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 4471–4480.