Spikformer For Audio Classification

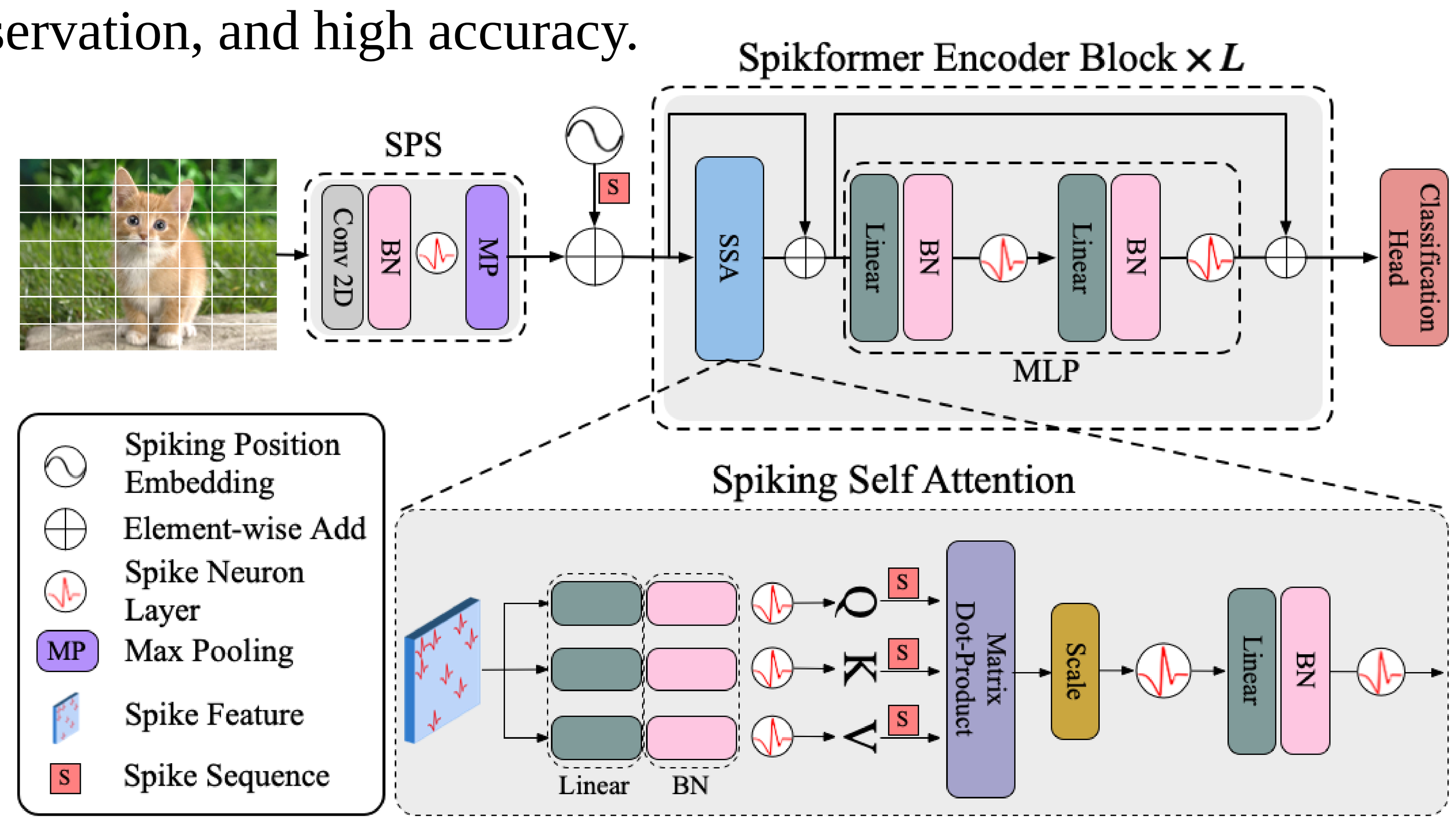
Deep Learning 2024 Fall – Final Project Presentation

國立陽明交通大學 NATIONAL YANG MING CHIAO TUNG UNIVERSITY

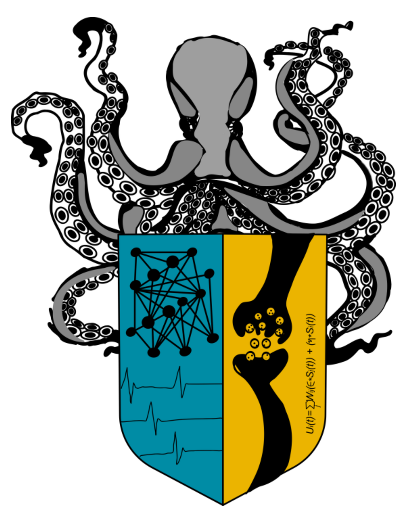
Advisor: 張添烜 Students: 李品辰、陳孟頡

1 Spikformer

Spikformer is a sophisticated deep learning framework that seamlessly integrates biologically inspired **Spiking Neural Networks** with the **self-attention mechanism**, achieving exceptional efficiency, energy conservation, and high accuracy.

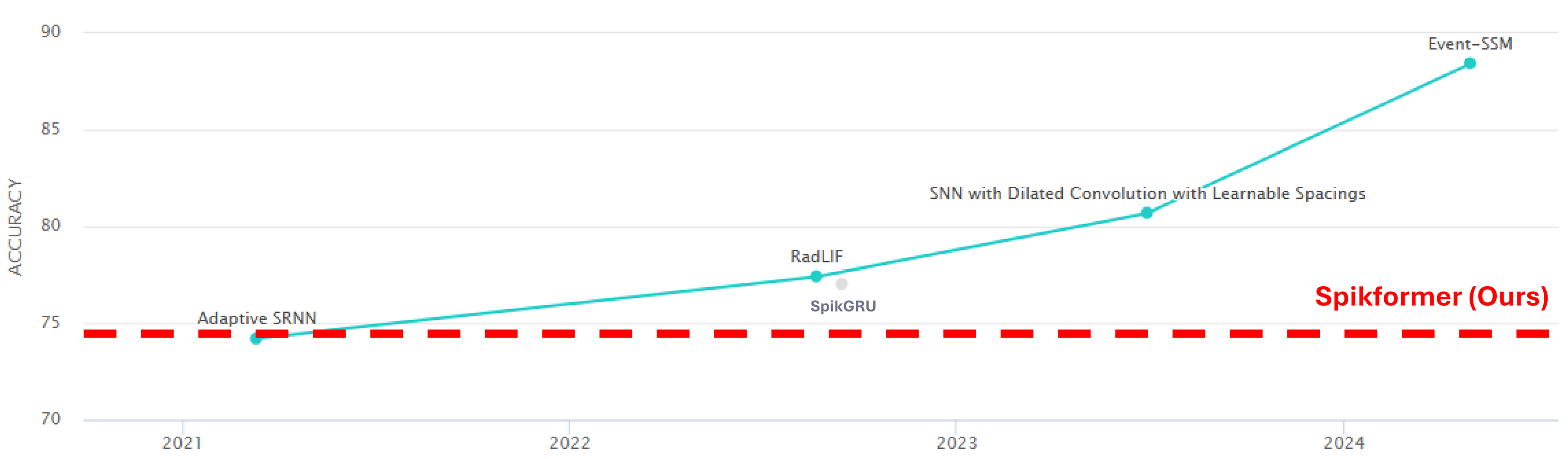


2 Spiking Speech Commands v0.2



SSC is a neuromorphic dataset designed for **benchmarking SNNs in speech recognition tasks**. It is derived from Google Speech Commands dataset, converted into spike-based data using event-driven encoding methods. Training on SSC is challenging due to its temporal nature, sparse and noisy signals, limited labeled data, and the need for energy-efficient models.

3 Benchmark Model Comparisons



Adaptive SRNN address the limitations of LIF by introducing adaptive mechanisms, such as trainable time constants and thresholds in Adaptive LIF neurons.

SpikGRU is a gated version of the Current-based LIF (Cuba LIF) model, spikes are integrated into a current variable prior to the membrane potential.

Rad-LIF (Rate-Adaptive Leaky Integrate-and-Fire) model incorporate rate adaptation to dynamically adjust firing rates based on input characteristics.

SNN with DCLS use 1D DCLS layer with Gaussian interpolation across time to simulate learnable delays between consecutive layers, without using recurrent connections.

Event-SSM is an event-based state-space model (SSM) that leverages continuous-time methods to capture interactions between events that are widely separated in the spatial domain.

Model	SSC Accuracy (%)	Parameters (M)	Features	Recurrent Topology
Adaptive SRNN	74.2	N/a	Adaptive-LIF	✓
SpikGRU	77	0.28	Gated Cuba-LIF	✓
RadLIF	77.4	3.9	Rad-LIF	✓
SNN with DCLS	80.69	2.5	Learnable Delays	×
Event-SSM	88.4	0.6	Async. events	✓
Simple SNN (Ours)	**55.74**	**0.2**	**N/a**	×
Spikformer (Ours)	**74.63**	**17.3**	**Self-Attention**	×

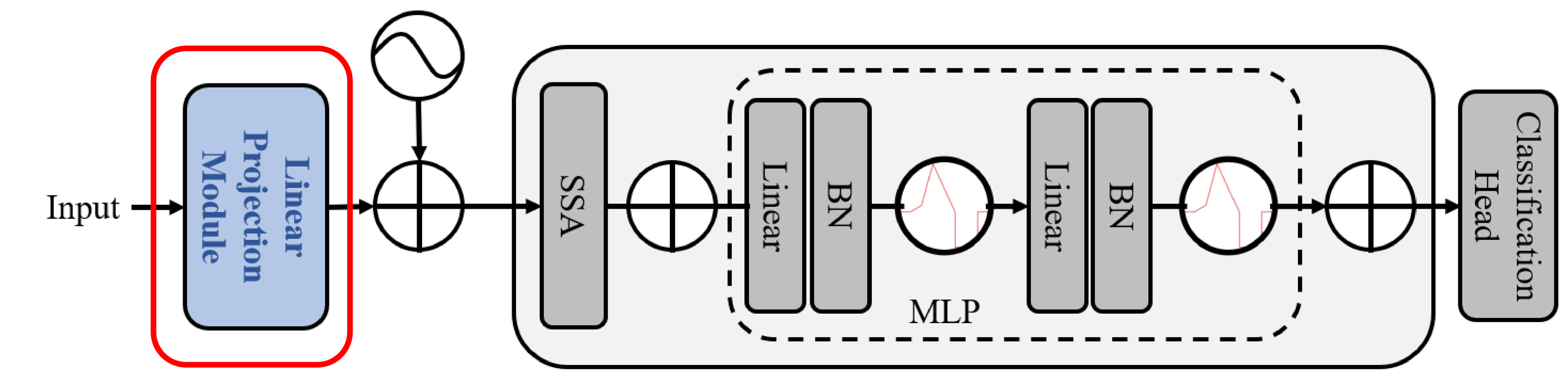
4 Spikformer on SSC

I. Why Spikformer

Spikformer is better suited for the SSC dataset compared to traditional ANNs because of the **event-driven and energy-efficient** characteristics while incorporating the **powerful feature extraction capabilities** of self-attention mechanisms. This makes it particularly effective for temporal spike-based data in the SSC dataset, offering superior efficiency and performance tailored to neuromorphic computing tasks.

II. Implementation

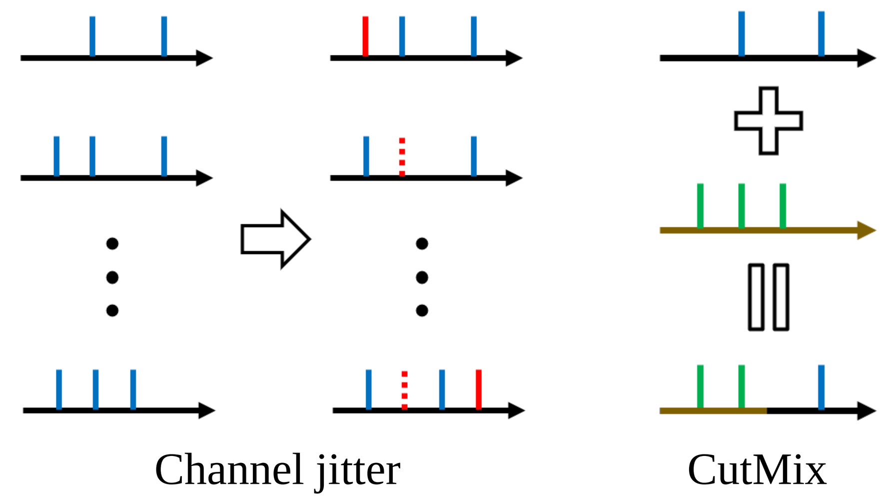
II.1 Architecture Modification



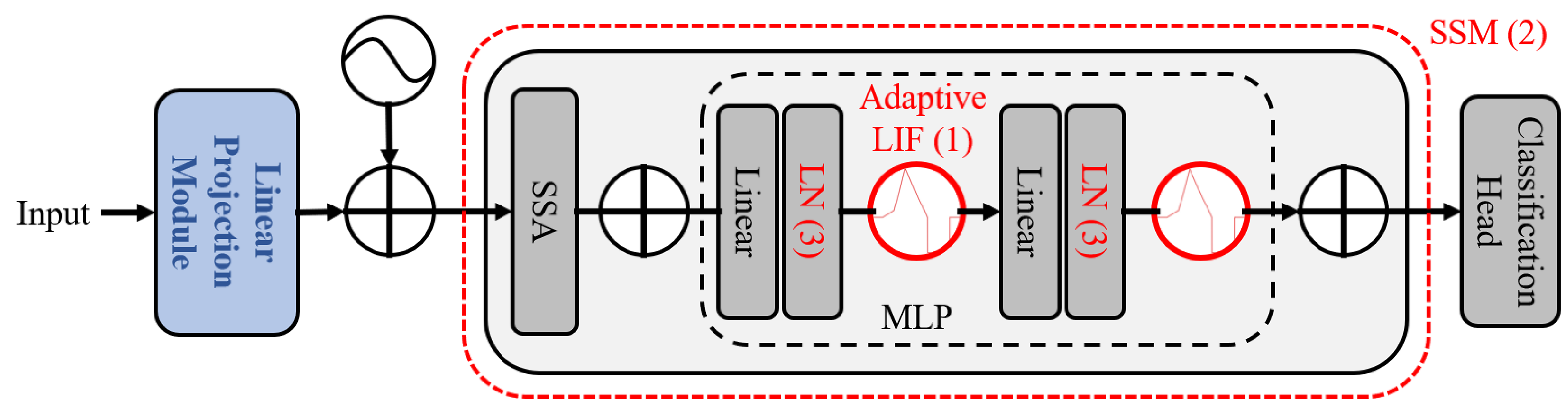
The SPS head is modified into **a linear projection module**, which maps input spikes to a higher embedding dimension.

II.2 Regularization

Various **regularization techniques** were employed to address overfitting issue.



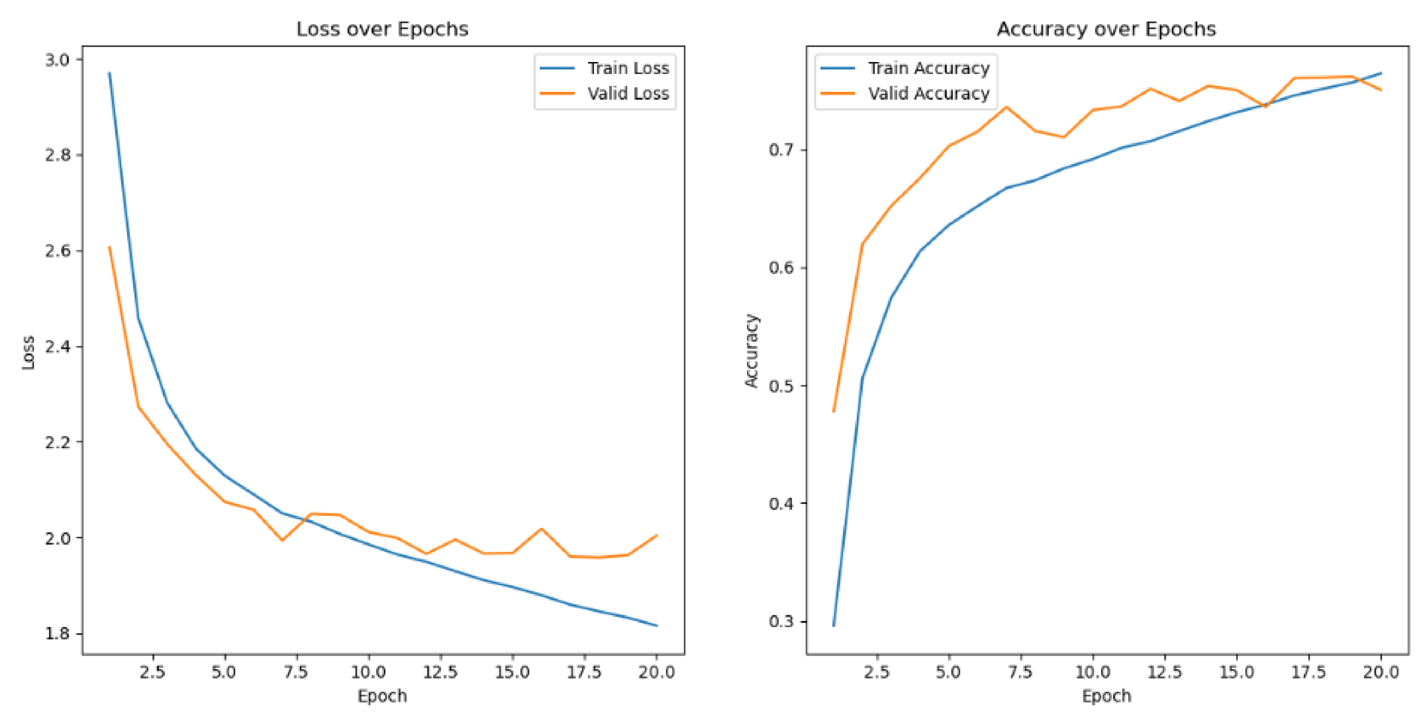
II.3 Method Combination



We **integrate various techniques** from existing research to further enhance the model's performance. However, not all of these approaches prove to be effective.

5 Experimental Result

Spikformer can easily achieve moderate accuracy on the SSC dataset. However, due to severe overfitting and the limited effectiveness of incorporating methods from other studies, our final result reached an **accuracy of 74.63%**, comparable to that of the Adaptive SRNN.



6 Conclusion

The project demonstrates that Spikformer is capable of achieving moderate performance on the SSC dataset. However, challenges such as severe overfitting hindered its potential, resulting in a final accuracy of 74.63%. This highlights the need for **more robust techniques to address overfitting and further optimize model performance**.