Spiking-EEG2TEXT：結合多視角時空特徵之脈衝 Transformer 腦波轉文字系統

Spiking-EEG2TEXT:
A Brain-to-Text System Combining Multi-View Spatio-Temporal Features and Spiking Transformers

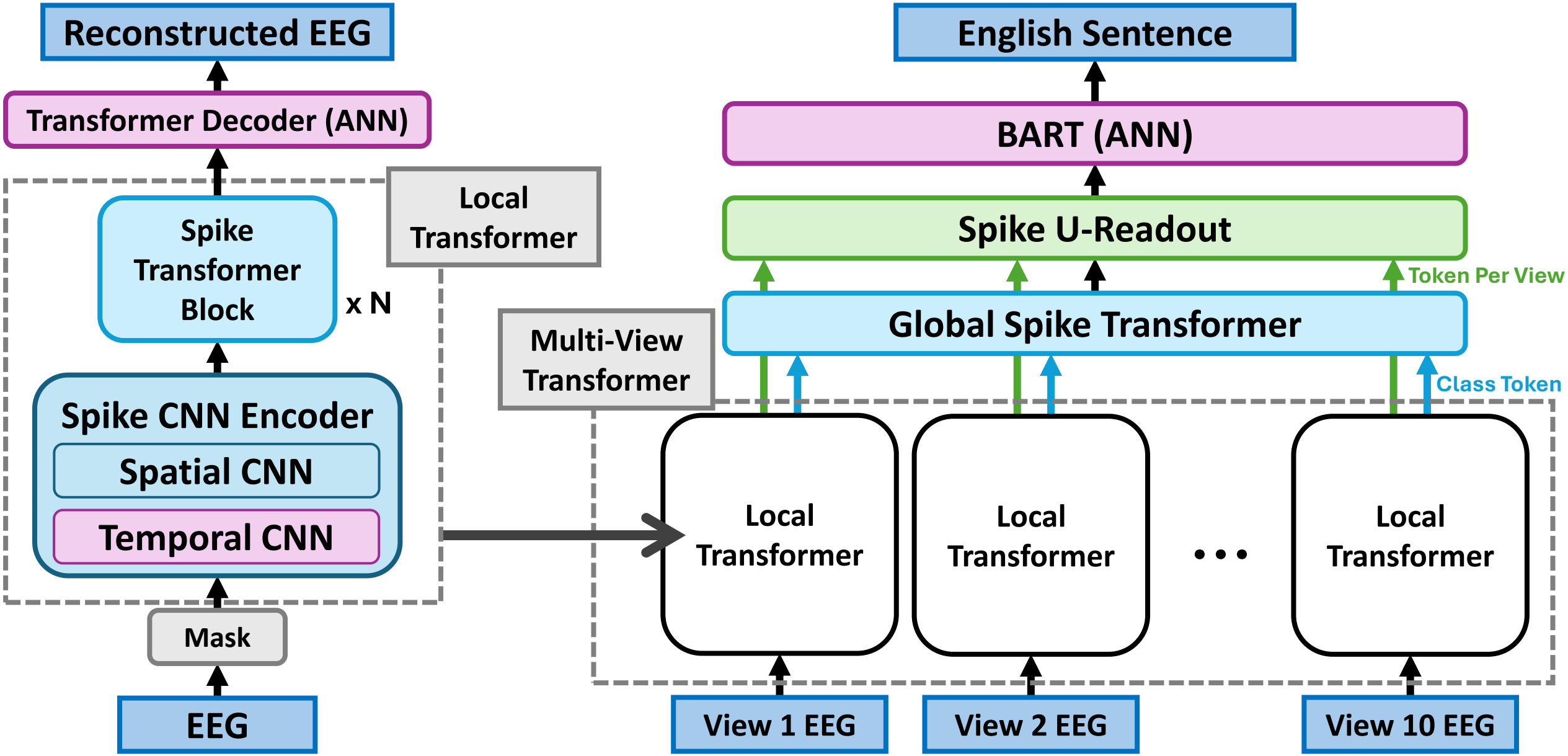
詹慧伶 教授、王俊哲、張哲元

Introduction

- **現況與挑戰**：主流 EEG-to-Text（Electroencephalography-to-Text）依賴 Artificial Neural Networks（ANNs）架構，面臨高昂的推理成本
- **技術新契機**：Spiking Neural Networks（SNNs）具備抗噪與低功耗優勢，適合 EEG 邊緣運算
- **研究切入點**：當前缺乏結合 SNN 的 EEG 語意解碼研究
- **本研究目標**：以 SOTA 架構（*Multi-view EEG2TEXT (*Liu, Hanwen, 2024*)* [1] ）為 Baseline，評估 SNN 應用於 EEG-to-Text 的可行性與效能邊界

Spiking-EEG2Text Architecture

Multi-view EEG2TEXT SNN version


Reconstructed EEG
Transformer Decoder (ANN)
Spike Transformer Block
x N
Local Transformer
Spike CNN Encoder
Spatial CNN
Temporal CNN
Mask
EEG
English Sentence
BART (ANN)
Spike U-Readout
Token Per View
Global Spike Transformer
Class Token
Multi-View Transformer
Local Transformer
Local Transformer
Local Transformer
View 1 EEG
View 2 EEG
View 10 EEG

Result

BLEU Score

	ANN w/TF		SNN (TS = 1) w/TF		SNN (TS = 4) w/TF	
	w/o noise	w/noise	w/o noise	w/noise	w/o noise	w/noise
BLEU - 1	35.03	35.04	29.93	27.30	31.90	31.04
BLEU - 2	20.22	20.13	15.39	13.36	17.08	16.50
BLEU - 3	11.89	11.81	8.10	6.43	9.38	8.99
BLEU - 4	6.89	6.92	4.36	2.77	5.32	4.95

Note: TF is Teacher Forcing. TS is Time Step.

Best	Ground Truth	Kennedy's other decorations of the Second **World War include the Purple Heart**, **Asiatic-Pacific** Campaign **Medal**, **and the** World **War II Victory Medal**.
	ANN (45.56)	edy was portrayal notable include the year **World War include the Purple Heart**, **theiatic-Pacific** Medal **Medal**, **and the** Presidential **War II Victory Medal**.
	SNN (10.84)	edy's performance half are the year **World War** are the Medal **Heart**, thepenatic-American **,,,and the** Medal **War II** Medal **Medal**.
Near BLEU-4	Ground Truth	After World **War II,** Kennedy entered politics (partly to fill the void of his popular brother, Joseph P. **Kennedy,** Jr., on whom his family had pinned many of their **hopes** but who was killed in the war).
	ANN (6.92)	a **War II**, he became the asasly as run the vacancy left the father father, President **Kennedy**. Kennedy, who.) who the he father had been their **hopes** their hopes) who had now in World Battle)
	SNN (7.73)	a **War II**, he was the andasly as serve a gap left a father father, who). Kennedy), who.), who the he father had been their **hopes** their hopes). who had not in the war.
Worst	Ground Truth	In 1946 McNamara joined Ford Motor **Company**, which he later said had been the result of a Life magazine **article** which reported how few college-educated managers there were at the then unprofitable **company**.
	ANN (1.92)	the,amara was the's **Company** as where was served became was a a largest of his " Time **article** that had that the people studentseducated men were were in Ford time-incitable Ford.
	SNN (3.31)	a,amara was the's **Company** as where was left became was been a biggest of his desireboat **article**. had on the people studentseducated men were were. the time-incitable **company**.

Energy Consumption

	Architectural Layers	ANN		SNN	
		Operations	Energy (mJ)	Operations	Energy (mJ)
Spike Encoding	Temporal Convolution	MAC	38.640	MAC	38.640
Spiking Core	Spatial Convolution Transformer U-Readout	MAC	10.158	**AC**	**0.177**
Language Model	BART	MAC	307.428	MAC	307.428
Total Energy			356.227		346.246

Note: E_{MAC} and E_{AC} denote the energy consumption per 32-bit operation, set to 4.6 pJ for Multiply-Accumulate (MAC) and 0.9 pJ for Accumulate (AC), respectively. Both values are grounded on a standard $45\mathrm{-nm}$ CMOS process at a $0.9\mathrm{V}$ supply voltage.

Conclusion

SNN vs. ANN

- Performance Gaps
  - 在對抗雜訊的表現上，SNN 現階段仍不如 ANN
  - 受限於脈衝與 surrogate function，SNN 的 BLEU 分數仍低於 ANN

- Energy Trade-offs
  - 單看 Spiking Core，其高度稀疏性確實帶來了極低的局部能耗
  - 由於 Spike Encoding 與 Language Model 占比較大，導致整體能耗並未顯著降低

Future Work

Future Work

- Performance Optimization
  - 引入 $\{-1, 0, 1\}$ 的三元編碼機制取代目前的二元脈衝，大幅提升時序編碼的資訊承載量
  - 棄用目前 CNN 編碼，改採將腦波的序列維度與 SNN 的時間維度直接對齊的編碼方式

- Energy Reduction
  - 訓練 Spike 版本的 BART

References

[1] H. Liu, D. Hajialigol, B. Antony, A. Han, and X. Wang, ‘Eeg2text: Open vocabulary eeg-to-text translation with multi-view transformer’, in *2024 IEEE International Conference on Big Data (BigData)*, 2024, pp. 1824–1833.

[2] 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, 2023.

[3] N. Hollenstein, J. Rotsztejn, M. Troendle, A. Pedroni, C. Zhang, and N. Langer, ‘ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading’, *Scientific data*, vol. 5, no. 1, p. 180291, 2018.

[4] N. Hollenstein, M. Troendle, C. Zhang, and N. Langer, ‘ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation’, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 138–146.

References

[5] A. Gulati et al., ‘Conformer: Convolution-augmented transformer for speech recognition’, *arXiv preprint* arXiv:2005. 08100, 2020.

[6] Z. Wang and H. Ji, ‘Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification’, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, pp. 5350–5358.

[7] Z. Zhou et al., ‘Spikformer: When Spiking Neural Network Meets Transformer’, in *The Eleventh International Conference on Learning Representations*, 2023.

[8] C. Lv et al., ‘Spikebert: A language spikformer learned from bert with knowledge distillation’, *Neural Networks*, p. 108482, 2025.

[9] J. Wang et al., ‘$\mathrm{S}^2\mathrm{M}$-Former: Spiking Symmetric Mixing Branchformer for Brain Auditory Attention Detection’, *Advances in Neural Information Processing Systems*, vol. 38, pp. 126593–126620, 2026.

References

[10] Z. Zhou et al., ‘Spikformer v2: Join the high accuracy club on imagenet with an snn ticket’, *arXiv preprint* arXiv:2401. 02020, 2024.

[11] P. Gao, T. Ma, H. Li, J. Dai, and Y. Qiao, ‘ConvMAE: Masked Convolution Meets Masked Autoencoders’, *arXiv preprint* arXiv:2205. 03892, 2022.

[12] N. Zheng, H. Xia, Z. Sun, W. Li, and Y. Wang, ‘S3T-Former: A Purely Spike-Driven State-Space Topology Transformer for Skeleton Action Recognition’, *arXiv preprint* arXiv:2603. 18062, 2026.