GAFormer:

Group-aware embeddings for timeseries transformers

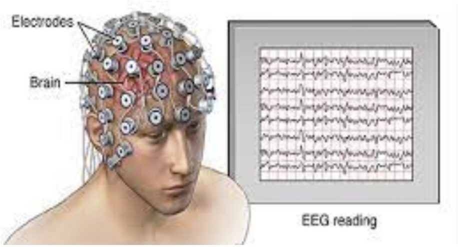
Jingyun Xiao*, Ran Liu, Eva Dyer
Georgia Tech, Machine Learning

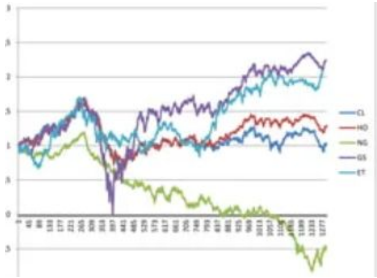
Presented at ICLR 2024

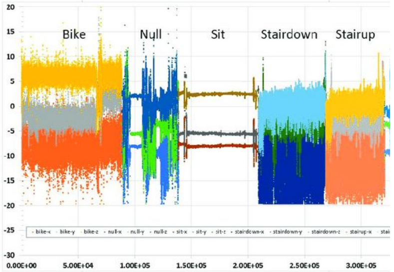
Motivation

Multi-variate timeseries: Multiple variables change over time.

Multi-variate timeseries data is everywhere!







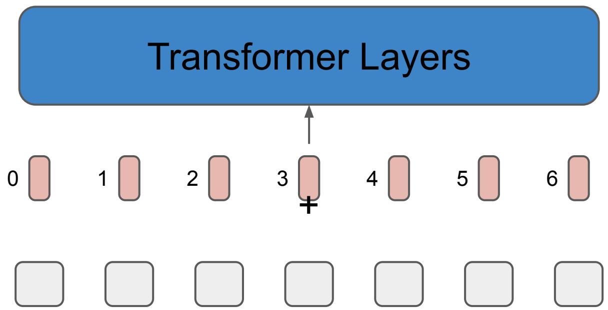
Examples: Electrophysiological signals / Trading trend / Traffic forecasting / Human movement

Challenges

When modeling timeseries with transformers, temporal and spatial information must be encoded through “position embeddings” (PE)

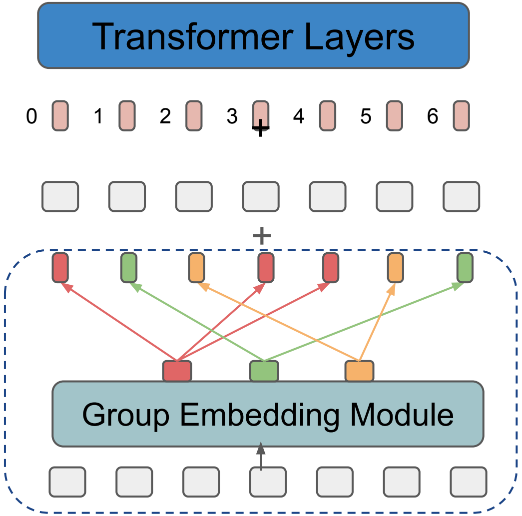
Issues with PEs:

- Not data adaptive - fixed for all samples
- There is no predetermined ordering or spatial “position” for different channels.



Our Approach

Build a data-adaptive token augmentation to group time and space!



Group embedding

$$X \leftarrow X + \mathrm{GE}(X)$$

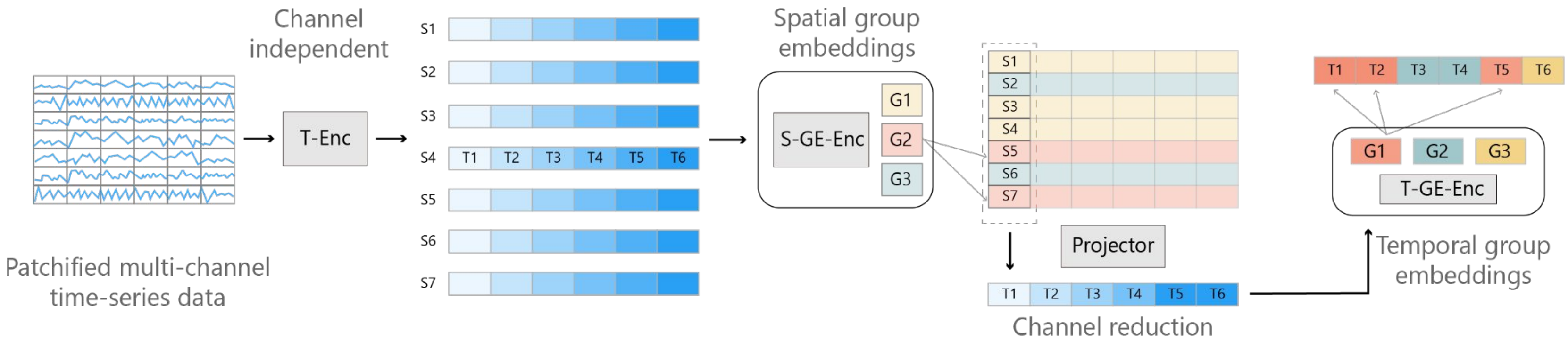
$$\mathrm{GE}(X) = SoftMax(\mathrm{Encoder}(X) \cdot W) \cdot G$$

Data-dependent assignment of tokens

Learnable groups

GAFormer

To group in both time and space, we introduce a new spatiotemporal transformer architecture for time-series data!

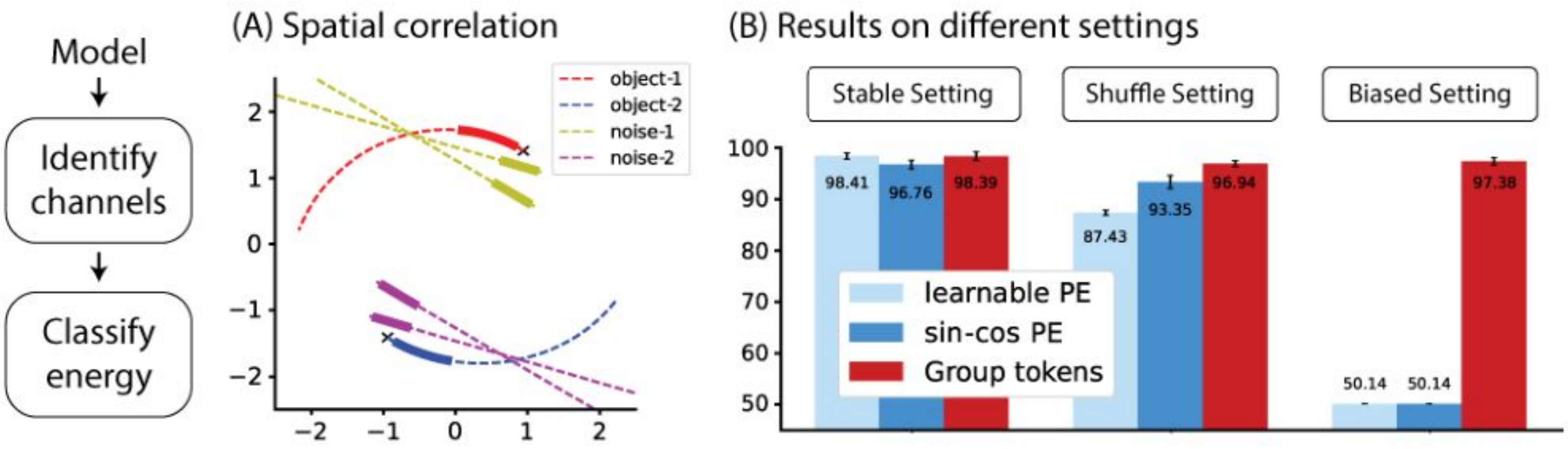


Channel-independent encoder

Add **Spatial** GEs!

Add **Temporal** GEs!

Results - Synthetic Data



Results - Time-series classification

	InlineSkate	Earthquakes	Adiac
GRU(Dey & Salem, 2017)	28.00	74.82	37.08
TCN(Lea et al., 2017)	22.55	74.28	58.06
MVTS(Zerveas et al., 2021)	22.18	74.82	57.54
MVTS + TGE	**34.73**	76.26	61.64
Δ	↑12.55	↑1.44	↑4.10
AutoTrans(Ren et al., 2022)	33.09	75.54	67.02
AutoTrans + TGE	**34.73**	**76.98**	**75.45**
Δ	↑1.64	↑1.44	↑8.43

Univariate time-series classification results

Results - Time-series classification

	SelfRegSCP2 (c=7)	FaceDetect (c=144)	Ethanol (c=3)	MotorImagery (c=64)	**Avg.**
NN	48.30	51.90	29.30	51.0	45.13
DTW_I	53.30	51.30	30.40	39.0	43.5
DTW_D	53.90	52.90	32.30	50.0	47.28
GRU(Dey & Salem, 2017)	51.11	56.56	34.60	51.0	48.32
TCN(Lea et al., 2017)	53.89	66.60	30.04	50.0	50.13
MVTS(Zerveas et al., 2021)	51.11	55.82	25.10	50.0	45.51
MVTS + TGE	51.67	61.75	30.42	55.0	49.71
Δ	↑0.56	↑5.93	↑5.32	↑5.0	↑4.20
AutoTrans(Ren et al., 2022)	44.78	65.12	27.76	53.0	47.67
AutoTrans + TGE	52.78	**68.05**	27.00	56.0	50.96
Δ	↑8.00	↑2.93	↓0.76	↑3.0	↑3.29
PatchTST(Nie et al., 2022)	50.56	54.99	25.86	54.0	46.35
GAFormer	**56.11**	67.99	**41.44**	**61.0**	**56.64**

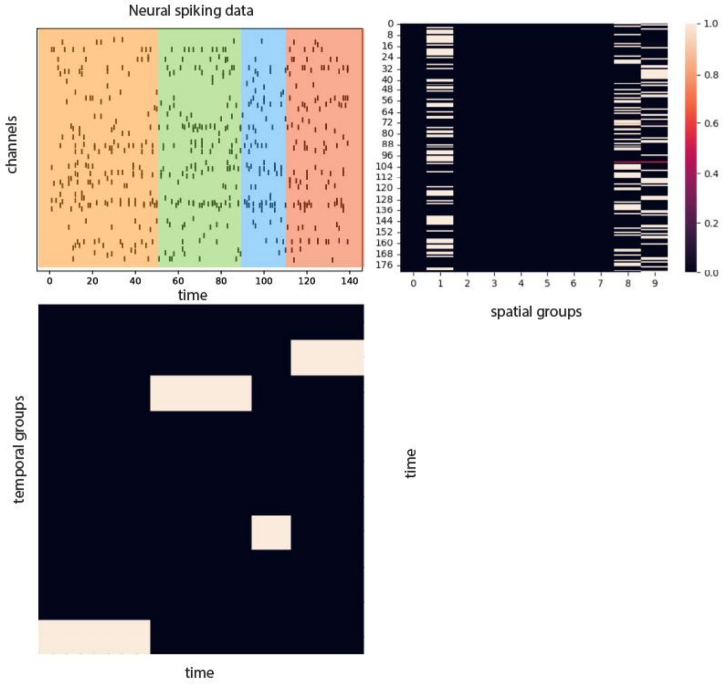
Multivariate time-series classification results

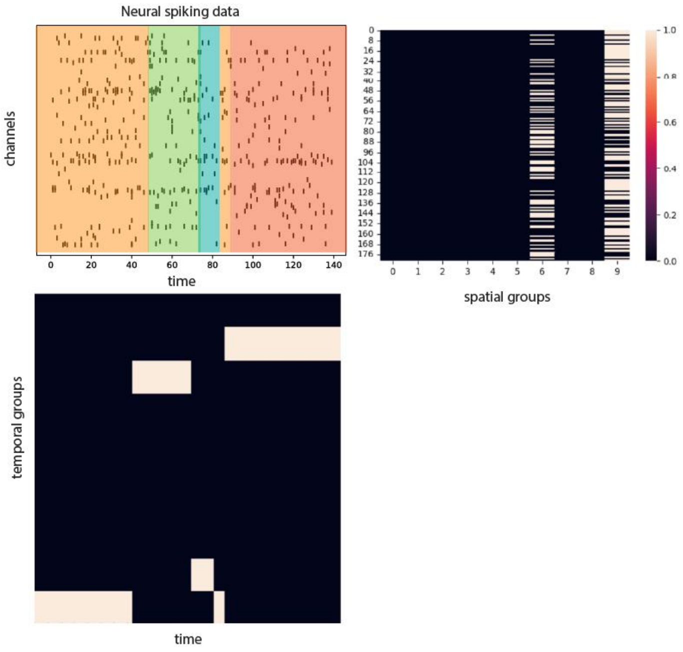
Results - Neural decoding

	Classification (Acc)				*Regression (R^2)*	
	Chewie-1	Chewie-2	Mihi-1	Mihi-2	NLB-Maze	NLB-RTT
GRU(Dey & Salem, 2017)	75.00	94.44	73.81	86.05	0.8887	**0.5951**
TCN(Lea et al., 2017)	78.13	91.67	90.48	81.40	0.8946	0.5407
NDT(Ye & Pandarinath, 2021)	81.06	88.89	88.10	**90.70**	0.8708	0.4621
EIT(Liu et al., 2022)	75.00	77.78	78.57	65.91	0.8791	0.4691
GAFormer	**81.25**	**94.44**	**92.86**	88.37	**0.9136**	0.5433

Results - Visualizations of GEs

GEs are computed for each input sequence, allowing for some visualization of how time and space are grouped!





Conclusion

- ***A novel data-adaptive group embedding (GE) technique***
  Learns grouping structures in multivariate timeseries datasets.

- ***A novel framework, Group-aware Transformer (GAFormer)***
  Provides a robust solution to learning of spatial and temporal patterns that leads to improved classification.

- ***Applications to multivariate time-series datasets and neural activity recordings***
  Offers meaningful interpretability and state-of-the-art performance in a variety of different types of timeseries datasets.

Thank you for your attention!

NerDS Lab

dyerlab.gatech.edu

https://github.com/nerdslab/GAFormer