Progressive Memory Transformer

Memory-Aware Attention for Time-Series

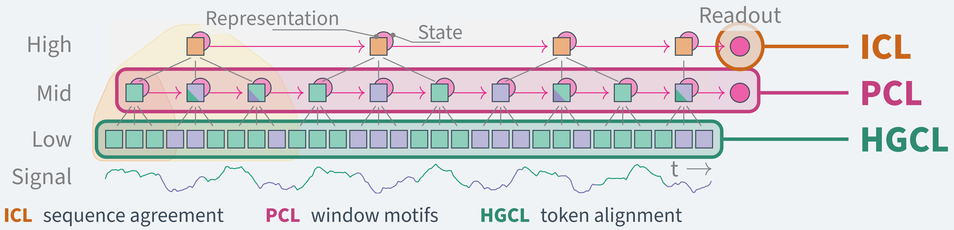
Tord Sture Stangeland[1,2] Andreas Köhler[2,3] Steffen Mæland[2,4] Adín Ramírez Rivera[1]

1 University of Oslo · 2 NORSAR Listening to the Earth · 3 UiT The Arctic University of Norway · 4 Western Norway University of Applied Sciences

Paper Code

DATA VIEW

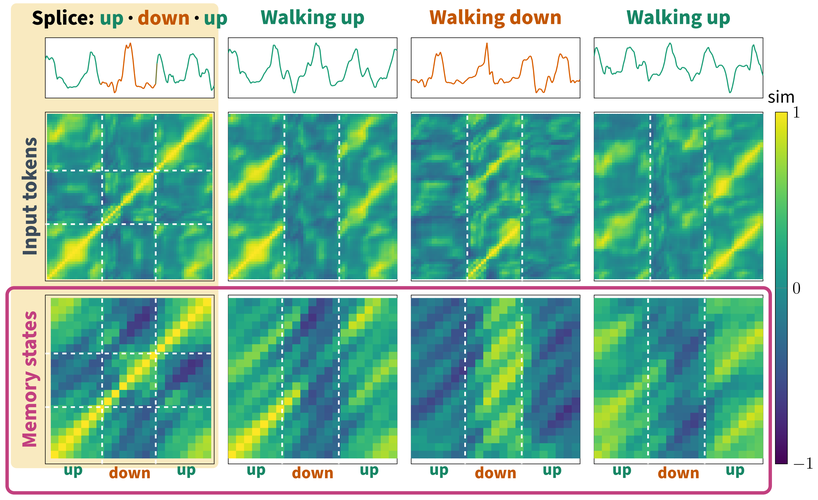
Three scales, one loss each



ICL sequence agreement **PCL** window motifs **HGCL** token alignment

Mid-range: motifs in memory

Memory states show the class blocks more sharply than **input tokens**



Every map: x = the splice, y = the sequence above it; time runs → and ↑; dashed lines mark the segments

Takeaway: against itself (first map), the splice forms a checkerboard in memory: the two **up** segments match each other across the **down** segment between them, and **down** stays dark against both.

Cue retention end-of-sequence probe

FordA	TS2Vec	T-XL	xLSTM	w/o PCL	PMT
AUC	.639	.967	.986	.962	**.995**
Top-1	.620	.905	.946	.904	**.966**

Window size share of the sequence

HAR	0.1	0.2	0.3	0.5	1.0
Top-1	91.9	**92.9**	92.7	92.6	92.1
Macro-F1	91.9	**93.0**	92.8	92.7	92.2

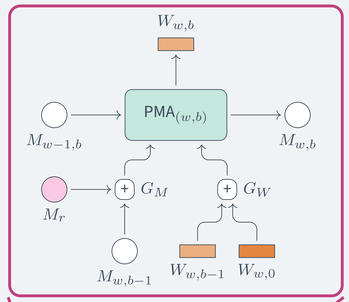
FUNCTIONAL VIEW

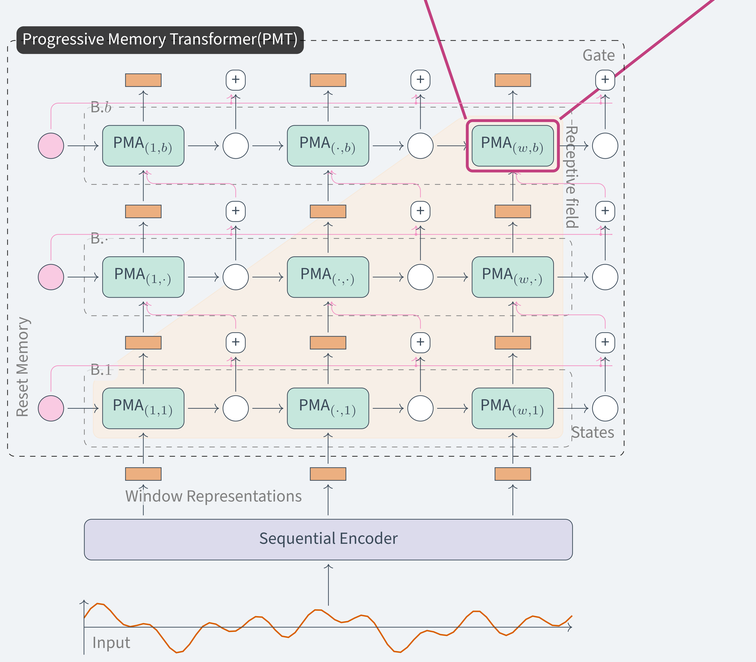
Progressive Memory Attention

Slots. Each window gets a few memory slots, read and written with its tokens in one attention pass (green).

Flow. Memory moves on to the next window and up the stack, so the receptive field (orange) grows with time and depth.

Gates. A reset gate mixes in a learned reset state (pink), so stale memory can be overwritten; a token gate keeps local detail.





Takeaway: a writable, window-aligned memory gives a transformer a **mid-range** representation, so each scale gets its own loss.

Probing every scale

Global low-label classification, Top-1 (%)

	1% labels			5% labels		
	TS2Vec	TS-TCC	**PMT**	TS2Vec	TS-TCC	**PMT**
HAR	91.0	82.9	**92.0**	92.1	92.6	**93.4**
Epilepsy	96.3	95.6	**96.6**	96.7	96.2	**96.9**
Wafer	95.3	96.5	**98.0**	**98.8**	98.2	**98.8**
FordA	87.1	81.5	**87.8**	92.5	**93.2**	89.5
FordB	67.9	74.8	**75.6**	78.8	**88.0**	76.9
POC	63.6	65.4	**65.9**	70.9	69.4	**72.7**
ElectricDevices	62.0	**64.6**	60.2	62.4	65.1	**65.9**
StarLightCurves	85.6	86.0	**92.6**	92.3	86.2	**96.9**
Average	81.1	80.9	**83.6**	85.6	86.1	**86.4**

TS2Vec and TS-TCC trained with SoftCLT; SVM on frozen features

Local forecasting: lower error in **31 of 40** comparisons

Average MSE	TS2Vec+SoftCLT	PMT
ETTh1, univariate	0.155	**0.114**
ETTh1, multivariate	**0.774**	0.801
Electricity, univariate	0.485	**0.477**
Electricity, multivariate	0.452	**0.399**

Backbone swap same losses and probe, 7 datasets

	Transformer-XL	xLSTM	PMT
Top-1	83.0	83.0	**83.6**
Macro-F1	80.8	80.8	**81.6**

One region per segment

t-SNE of the heat maps' splice; shade marks the position within each segment (light = start)

