Enhancing Sparse Event Detection in Healthcare Time-Series via Adaptive Gate of Context–Detail Interaction

ICLR 2026

Beomjun Bark[1] Yun Kwan Kim[1, 2]

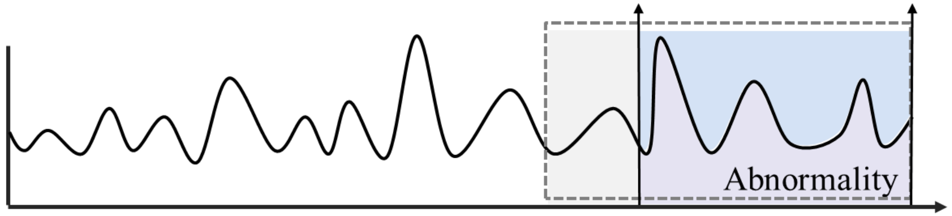
1 seers

2 KOREA UNIVERSITY

ICLR International Conference On Learning Representations

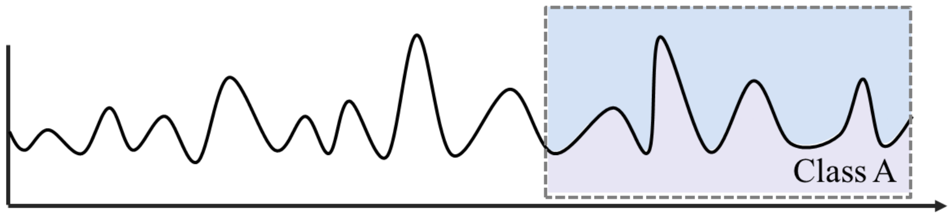
1. Introduction & 2. Related work

(a) Anomaly Detection in Time Series



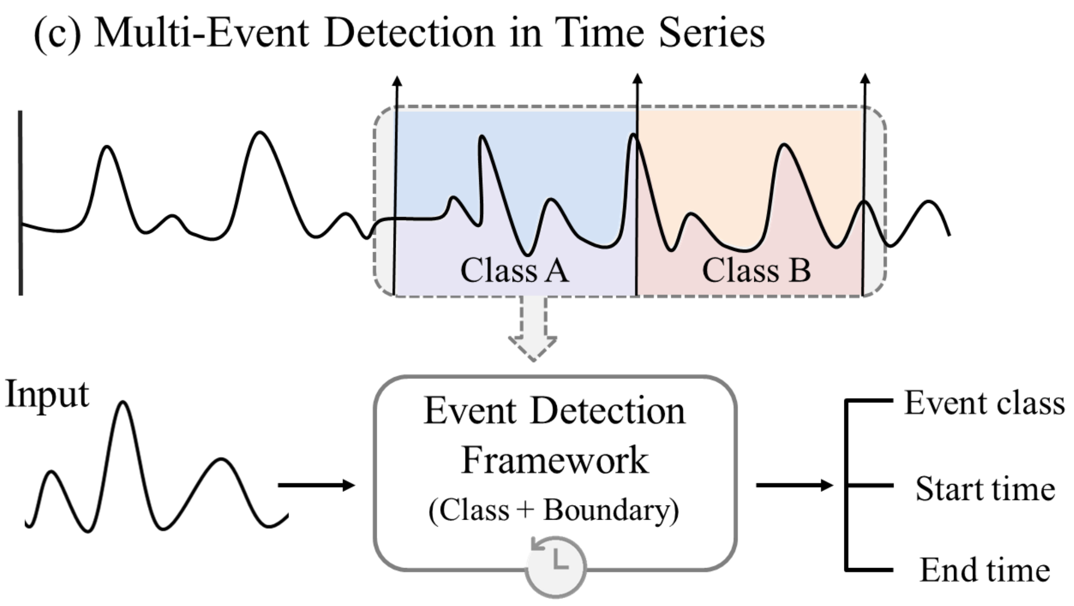
Identify only the abnormal start and end points

(b) Classification in Time Series



Within the fixed segments,
only the type and presence of
specific events can be identified

These two prior approaches indicate a lack of reliability of interpretation and clinical actionability from a healthcare perspective.

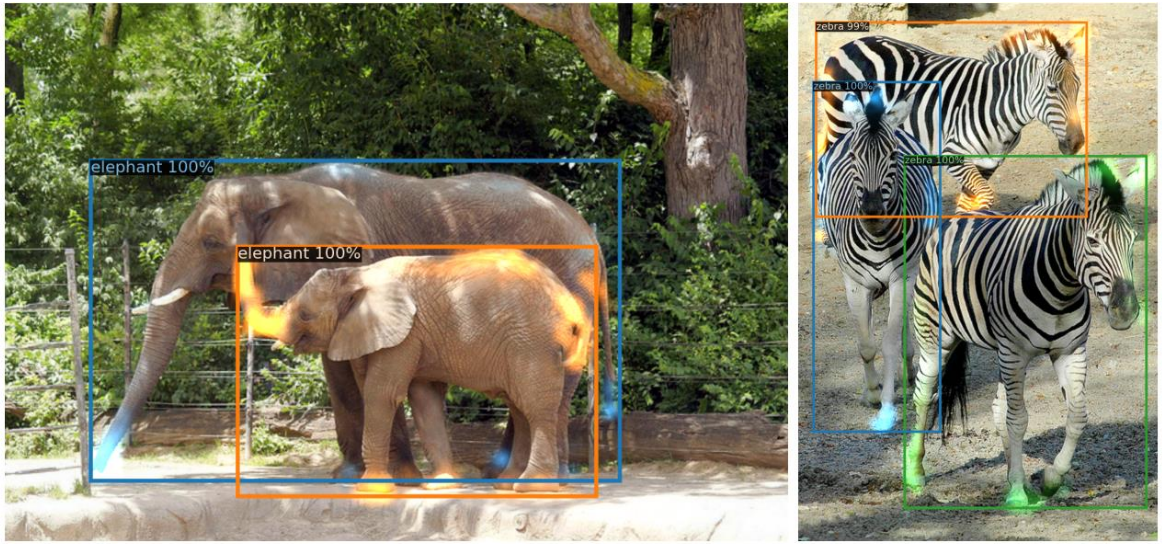


Identify the start and end times of events, as well as their classes

→ Serve as important insights for diagnosis

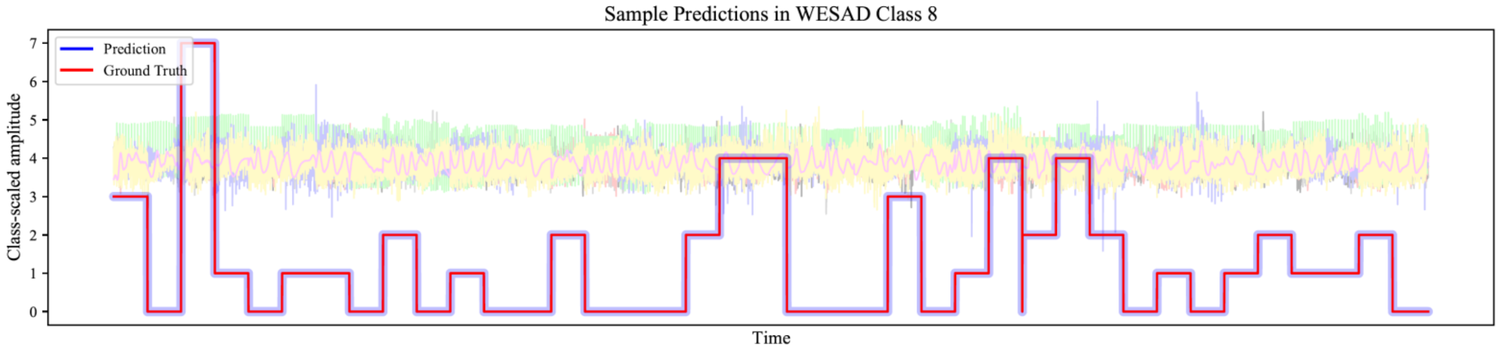
How?

[1]



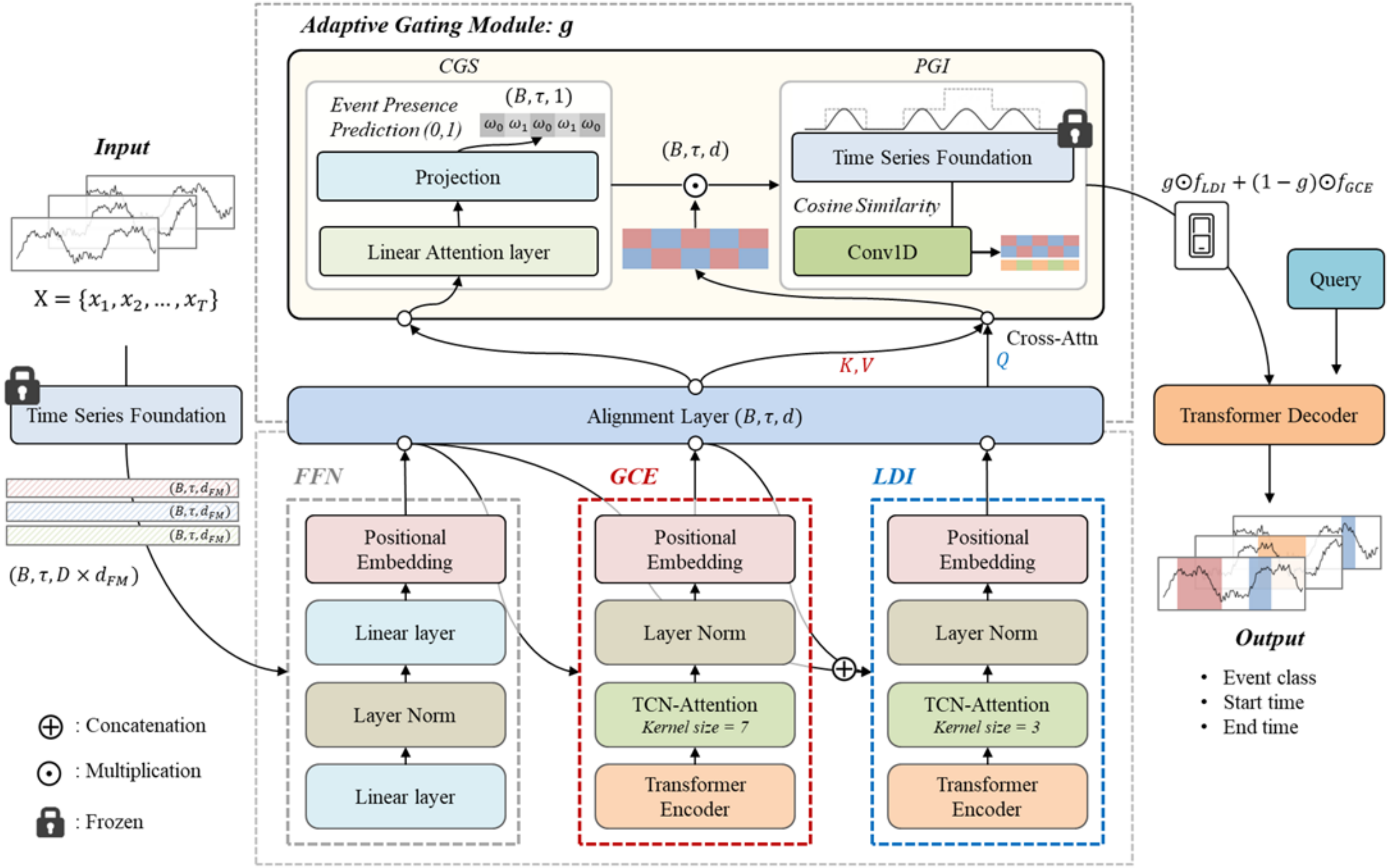
Concept:

Apply a DEtection TRansformer-based approach to time-series data

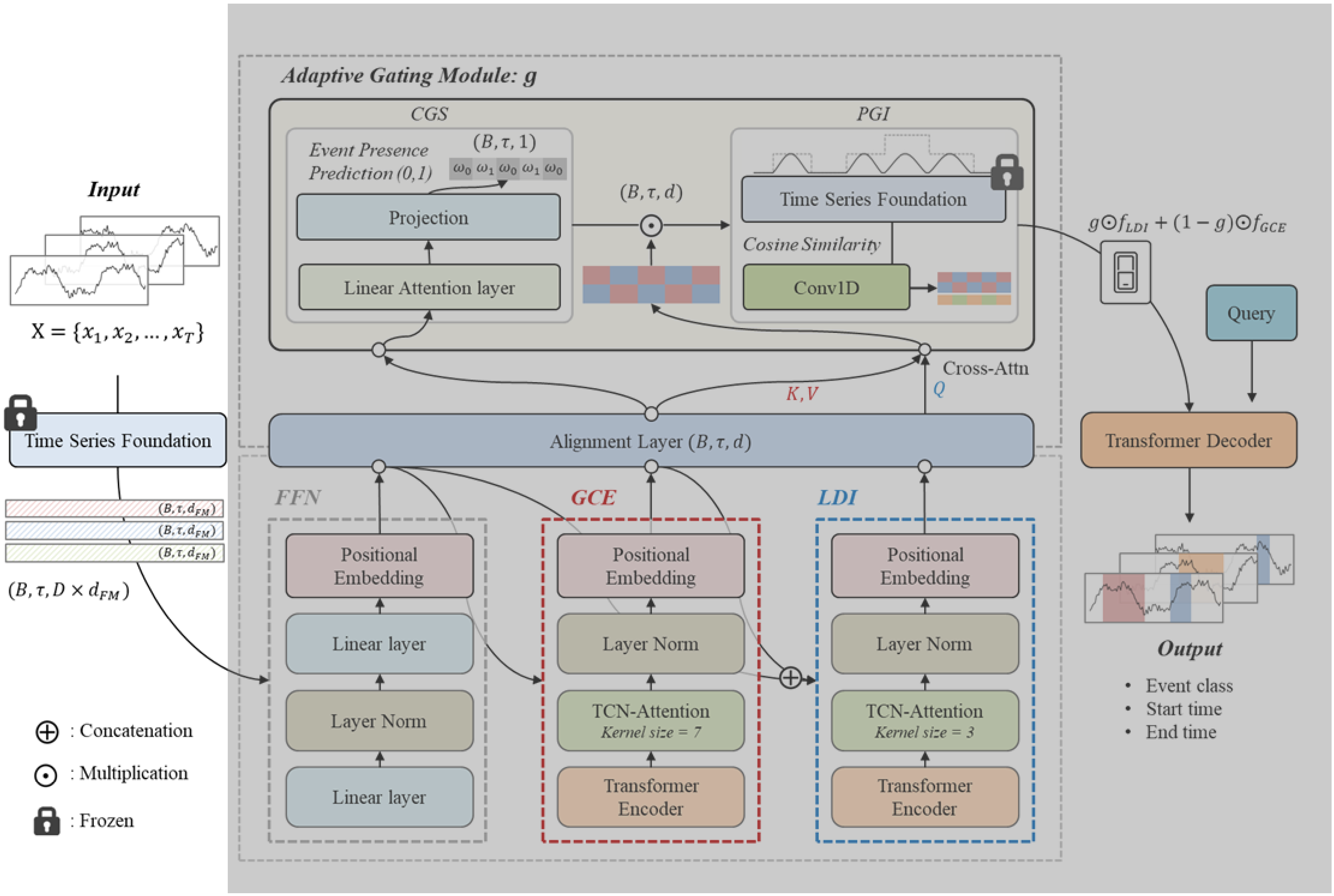


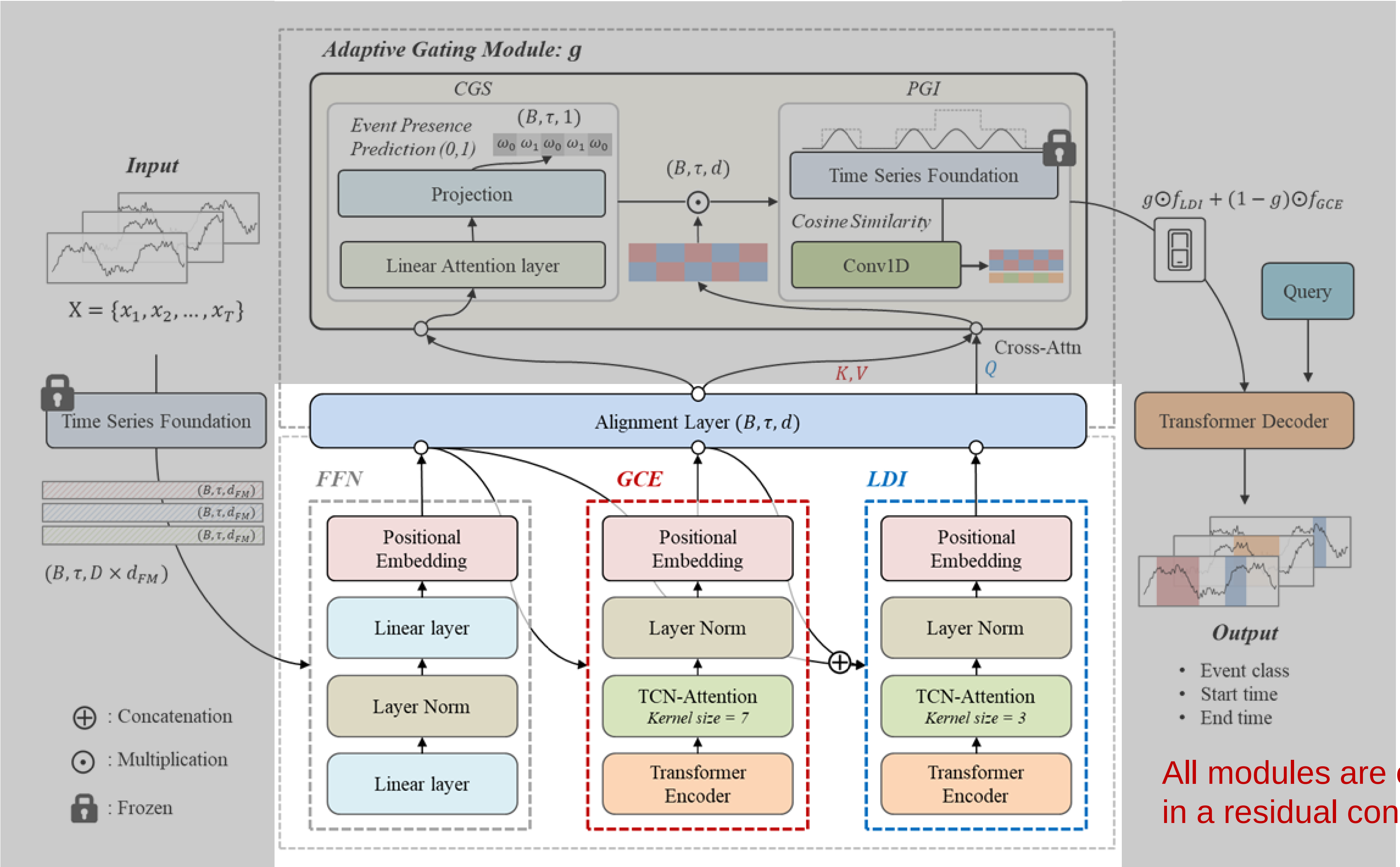
[1] Carion, Nicolas, et al. "End-to-end object detection with transformers." *European conference on computer vision*. Cham: Springer International Publishing, 2020.

3. Methodology



Extracting features from multivariate time-series data on a per-dimension basis using the Chronos T5-bolt-tiny foundation model



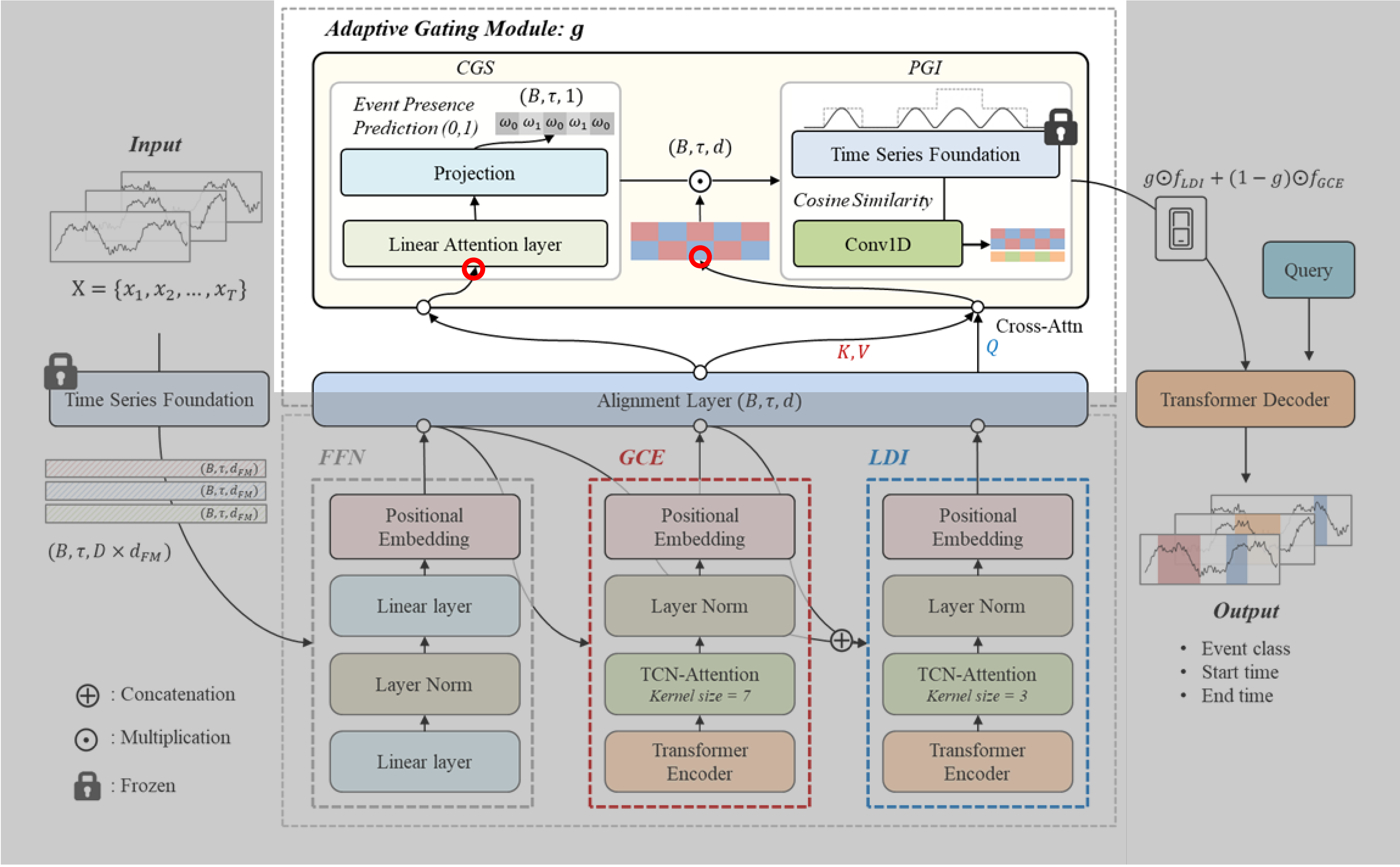


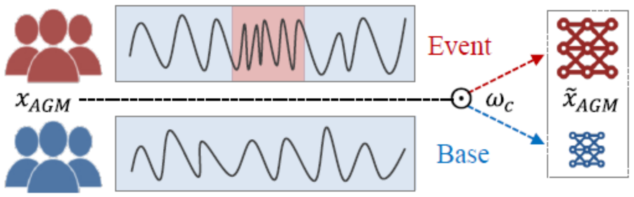
All modules are connected in a residual connection.

- FFN (Feed-Forward Network): A module that effectively embeds features into Feature Map
- GCE (Global Context Explorer): A module that extracts global features from the feature representation
- LDI (Local Detail Inspector): A module that extracts local features from the feature representation
- Alignment Layer: A layer that aligns potentially mismatched representations across each module

AGM Input:

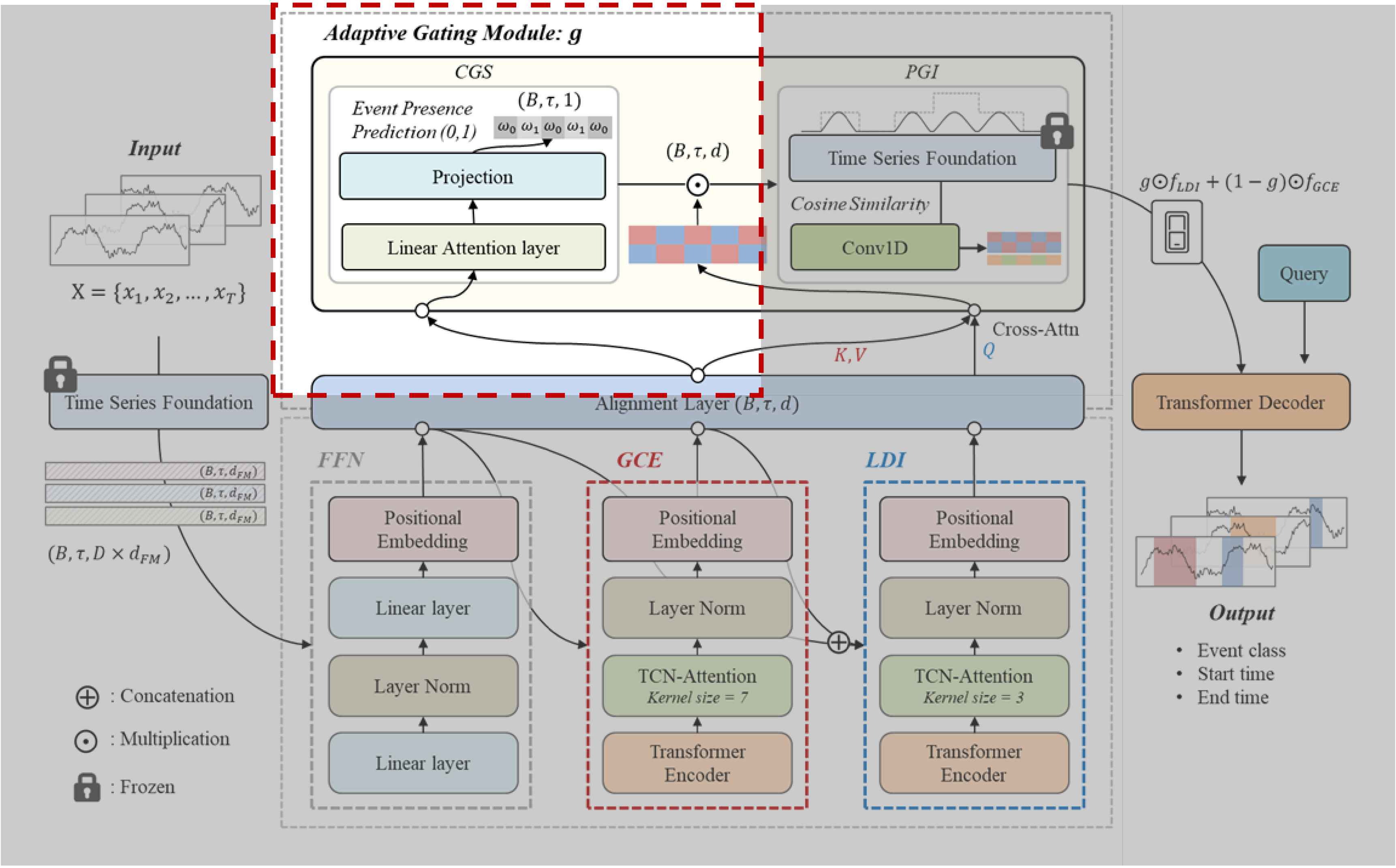
1. GCE Output → CGS Input
2. *CrssAttn*(GCE Output, LDI Output) → Component of PGI Input

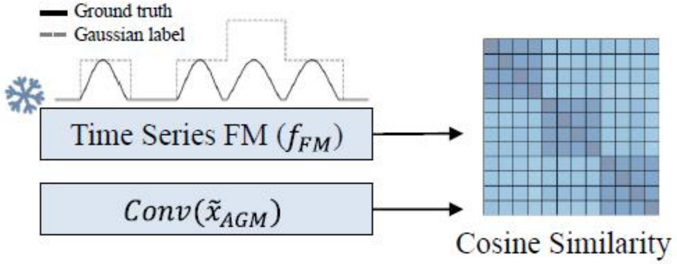




This module is responsible for classifying whether an event exists for a given input and amplifying the features according to the class ratio (Y/N).

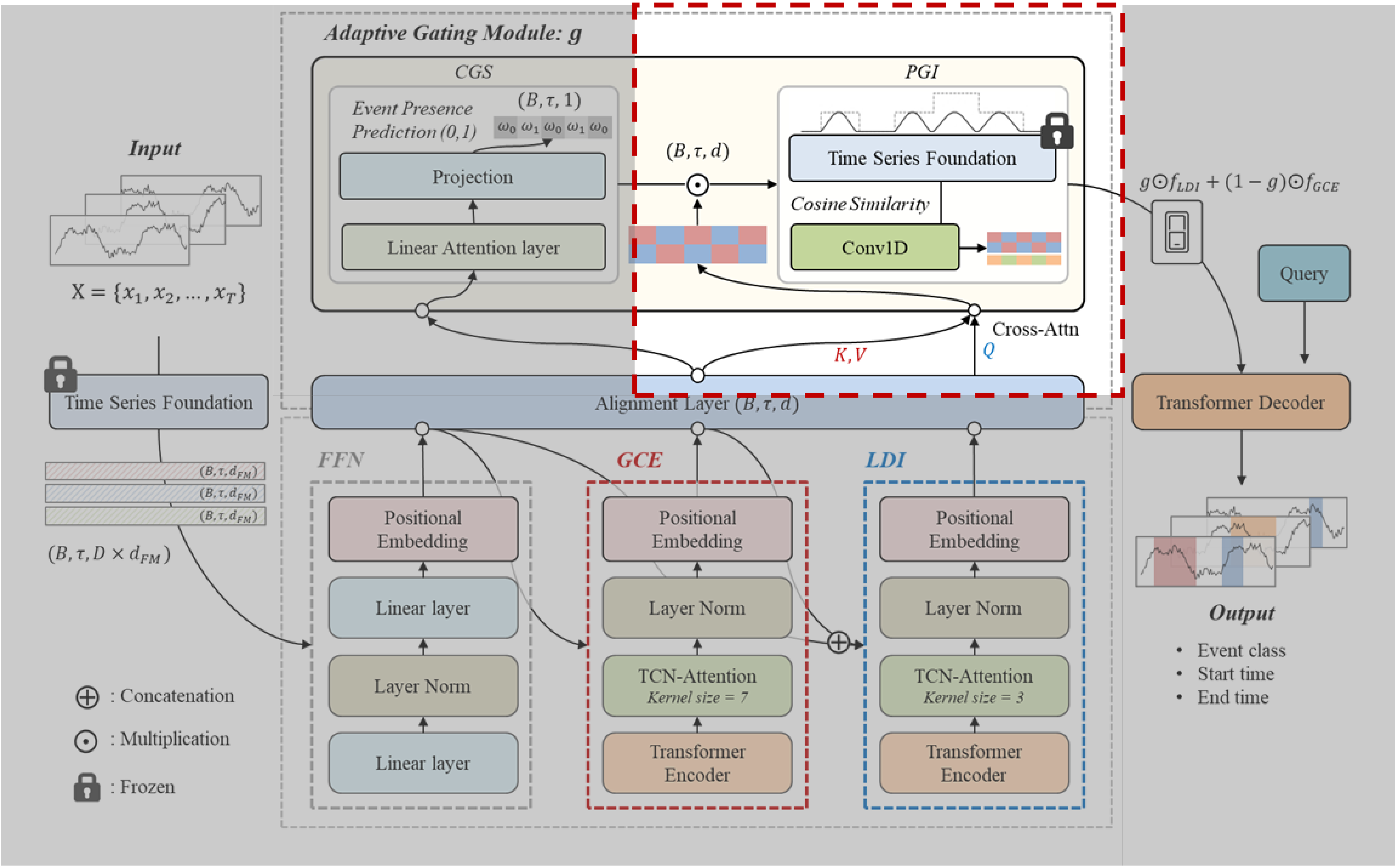
Purpose: To enable faster convergence through event presence/absence detection, and to selectively focus only on occurrence samples (sparse events).

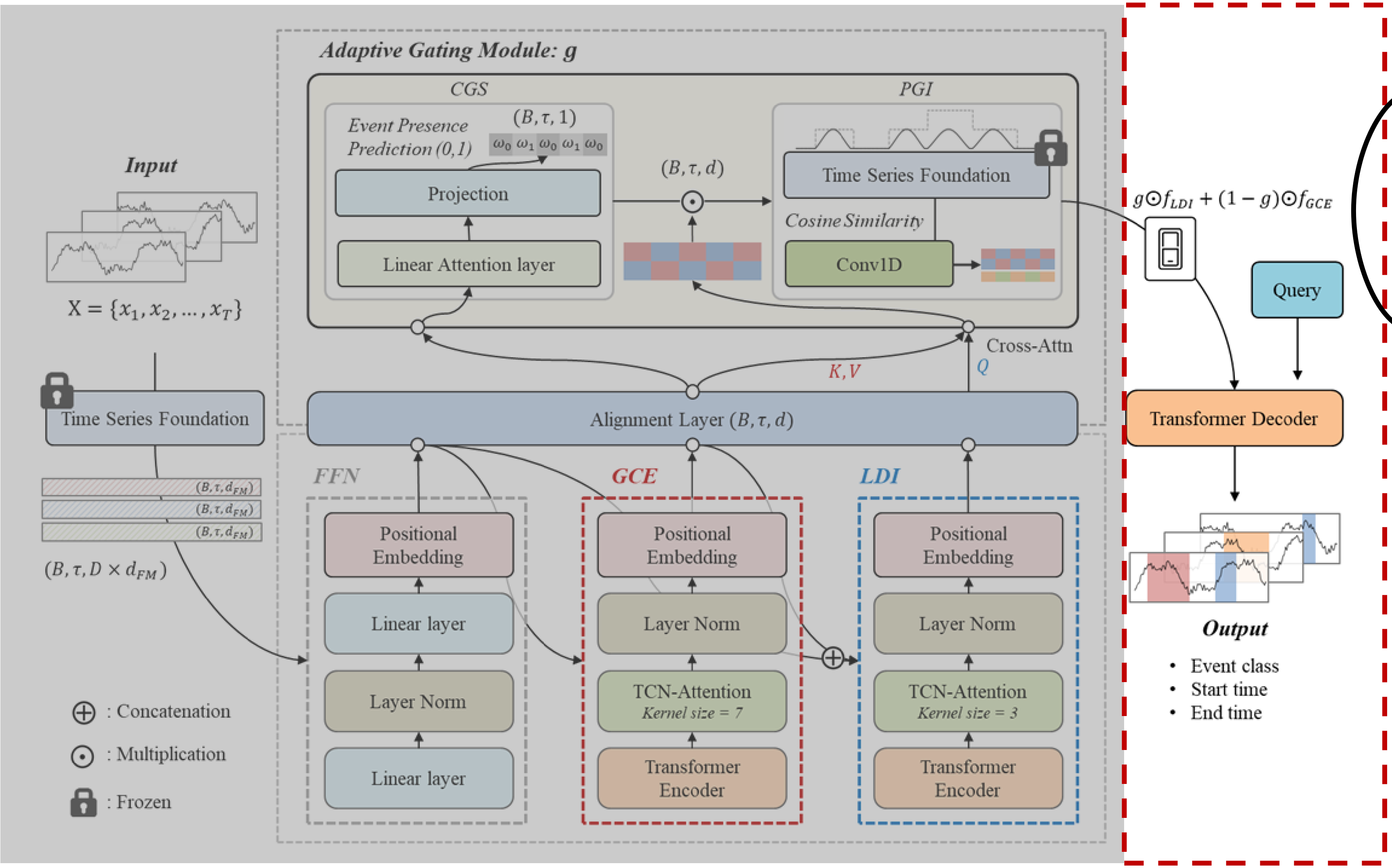




This module transforms the positional information of labels into a Gaussian form, enabling the *Convolution layer* to effectively learn positional information based on similarity.

Purpose: To explicitly emphasize the positions of sparse events while allowing for a smoother representation.





AGM provides gated information that selectively highlights relevant signals for sparse event detection

Applied to both local(LDI) and Global(GCE) feature (final feature)

And … same as vanilla DETR

Training

by PGI   by CGS

$$L_{\text{total}} = \boxed{\alpha L_{\cos}} + \boxed{\beta L_{\text{BCE}}} + \gamma L_{\text{Detection}}$$

4. Experiment & Result

Datasets

- MIT-BIH: Arrhythmia detection
- SHDB-AF: Arrhythmia detection
- WESAD: Emotion recognition
- OPP: Activity monitoring

Baseline

- DETR
- Multi-scale DETR
- Deformable DETR
- DAB-DETR
- DN-DETR
- Deformable-DINO

Metrics

- macro Point-wise F1 score
- macro Affiliation F1 score
- mean Average Precision

Model	Metric	MIT-BIH			SHDB-AF		WESAD	OPP
		Class 3	Class 6	Class 15	Class 3	Class 5	Class 8	Class 5
DETR	PW-F1	77.22	64.41	41.32	93.24	60.19	59.53	61.30
	AF-F1	82.72	58.29	43.35	93.47	60.83	62.29	61.89
	mAP	51.45	53.33	45.23	95.78	77.77	53.14	50.85
Multi-scale DETR	PW-F1	75.60	73.36	65.52	91.99	63.49	63.88	58.05
	AF-F1	76.00	**62.39**	49.34	92.13	64.67	65.65	58.38
	mAP	66.03	55.60	50.97	97.76	76.85	68.24	48.28
Deformable DETR	PW-F1	85.39	68.68	58.96	90.86	53.33	62.69	61.05
	AF-F1	84.71	60.81	45.67	91.10	53.99	65.45	60.92
	mAP	63.17	52.59	**52.16**	97.79	74.18	66.59	51.86
DAB-DETR	PW-F1	83.13	71.05	53.50	93.11	51.52	55.59	60.35
	AF-F1	83.70	60.77	46.03	93.23	52.14	58.58	60.88
	mAP	65.57	54.19	46.85	**98.96**	**86.16**	**72.75**	50.30
DN-DETR	PW-F1	77.82	66.82	62.17	91.59	71.92	66.44	57.00
	AF-F1	77.71	58.18	47.37	86.26	70.13	66.32	56.05
	mAP	66.28	51.42	48.80	96.69	80.06	55.12	49.29
Deformable-DINO	PW-F1	86.41	74.11	63.49	89.57	66.03	69.22	56.80
	AF-F1	83.22	60.00	48.59	88.82	65.98	70.73	55.80
	mAP	72.05	52.20	46.92	96.00	83.65	69.96	50.68
Ours	PW-F1	**90.63**	**83.37**	**74.86**	**96.23**	**83.41**	**73.59**	**64.98**
	AF-F1	**85.96**	60.87	**52.85**	**96.09**	**83.78**	**74.29**	**62.81**
	mAP	**77.66**	**57.54**	44.55	97.38	85.83	65.19	**60.07**

< Overall performance comparison across multiple datasets >

Table 2: Summary of notable results for specific dataset events. Each row reports the performance for a specific event within each dataset and provides the corresponding ratio. **Bold** indicates the best performance, while underlined values indicate the second-best performance. Full results for all classes are provided in *Appendix F*.

Dataset (Event)	Ratio (%)	Metric	DETR	Multi-scale DETR	Deformable DETR	DAB-DETR	DN-DETR	DINO	Ours
MIT-BIH Class 3 (AFL)	0.91	PW-F1	64.02	58.11	77.17	72.30	66.65	78.01	**84.58**
		AF-F1	77.39	60.51	81.90	77.39	71.04	79.13	**84.30**
MIT-BIH Class 15 (T)	1.32	PW-F1	0.60	38.31	34.19	28.06	58.97	60.94	**80.17**
		AF-F1	33.33	44.44	30.00	32.26	52.94	58.33	**78.57**
SHDB-AF Class 5 (PAT&NOD)	0.03	PW-F1	29.12	37.80	22.51	12.50	43.84	25.85	**66.41**
		AF-F1	29.75	38.65	22.88	12.50	37.69	24.54	**67.26**
WESAD Class 8 (Task 1)	0.95	PW-F1	30.11	31.36	10.41	37.46	41.56	49.59	**60.00**
		AF-F1	34.99	31.69	12.57	44.40	41.51	50.97	**57.12**
OPP Class 5 (Stand)	42.36	PW-F1	46.75	40.02	36.45	44.28	35.35	35.73	**55.61**
		AF-F1	45.91	40.61	32.97	44.53	27.37	29.18	**48.83**
MIT-BIH Class 6 (Ventricular arrhythmia)	0.59	PW-F1	20.26	44.59	29.88	44.51	24.53	54.71	**79.04**
		AF-F1	28.27	**36.36**	30.83	25.27	21.75	25.60	19.52

Consistently outperforms strong baselines with notable F1 gains.

Both sparse event & non-sparse event

Table 3: Summary of model performance by event length. Event classes are positioned as short or long based on their class-specific mean event length (CMEL, in seconds) relative to the global mean event length (GMEL, in seconds), considering a fixed window length. This table provides a detailed comparison of performance metrics for cases with short or long events across multiple datasets, based on the GMEL. **Bold** indicates the best performance, while underlined values indicate the second-best performance. Full results for all classes are provided in *Appendix F*.

Dataset (Event)	CMEL (GMEL)	Metric	DETR	Multi-scale DETR	Deformable DETR	DAB-DETR	DN-DETR	DINO	Ours
MIT-BIH Class 3 (AFIB)	8.52 (8.32)	PW-F1	90.42	93.08	93.61	93.95	88.99	94.81	**96.67**
		AF-F1	88.04	**91.49**	87.51	90.00	84.44	87.31	87.91
MIT-BIH Class 15 (SVTA)	7.75 (5.19)	PW-F1	81.21	**96.71**	94.72	96.36	94.24	96.43	96.34
		AF-F1	84.53	85.71	85.42	88.42	85.22	87.11	**88.45**
MIT-BIH Class 15 (P)	2.95 (5.19)	PW-F1	6.63	69.51	71.55	63.31	70.26	70.60	**80.05**
		AF-F1	37.60	44.64	50.83	51.91	50.25	51.15	**55.79**
MIT-BIH Class 15 (VFL)	0.88 (5.19)	PW-F1	3.36	31.54	31.53	22.68	16.55	32.16	**69.03**
		AF-F1	15.37	17.43	13.27	**22.93**	9.38	11.69	10.27
SHDB-AF Class 5 (AT)	9.61 (9.90)	PW-F1	42.69	58.20	38.28	32.81	65.82	55.68	**75.20**
		AF-F1	43.59	59.29	38.81	33.85	66.23	57.64	**76.16**
WESAD Class 8 (Stress)	9.45 (9.41)	PW-F1	75.93	84.41	85.44	83.63	77.02	85.28	**86.58**
		AF-F1	77.45	84.94	85.79	83.70	73.76	85.33	**86.57**
OPP Class 5 (Lie)	6.80 (3.93)	PW-F1	83.46	76.22	84.78	79.69	81.31	82.58	**86.67**
		AF-F1	85.58	79.13	87.03	82.54	85.03	85.27	**87.86**

Achieves strong performance across varying event lengths.

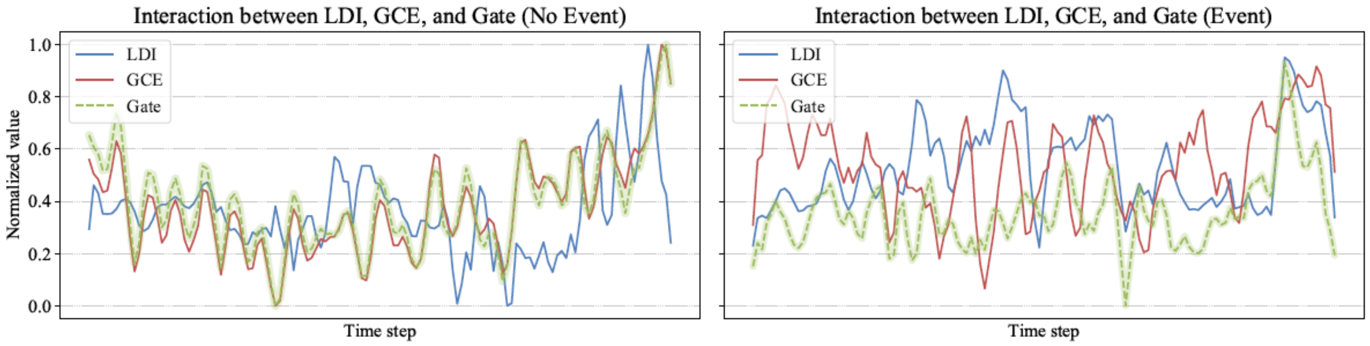


Figure 4: Adaptive gating module selectively prioritizes feature scales. The normalized outputs of LDI, GCE, and AGM are shown over time. In non-event periods (left), the gate closely follows the low-variability GCE, emphasizing stable global information. During events (right), the gate increases the contribution of high-variability LDI, capturing subtle changes and improving robustness.

- *When no event is present*: GCE predominantly governs the decision.
- *When an event is present*: The gate dynamically controls GCE and LDI to capture sparse events.

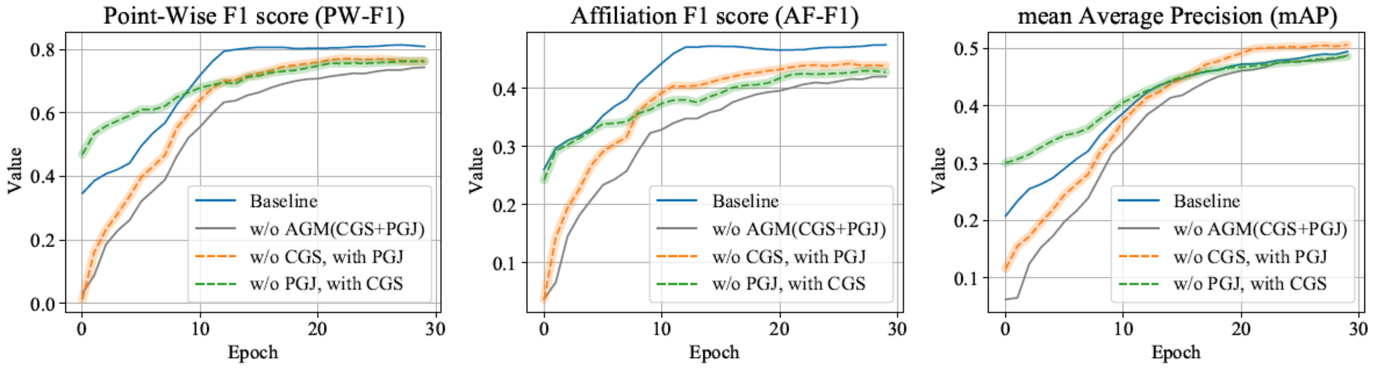


Figure 5: Epoch-wise performance of the model on three metrics (PW-F1, AF-F1, mAP), illustrating the contributions of CGS and PGI. Early in training (epochs 0–10), the CGS-only model (”w/o PGI, with CGS”, green dashed) converges faster, emphasizing event presence. After 15 epochs, the PGI-only model (”w/o CGS, with PGI”, orange dashed) surpasses CGS, achieving higher final performance by refining predictions based on positional information. This demonstrates the complementary roles of CGS and PGI in accelerating early learning and improving final outcomes.

- Using CGS alone has a strong impact on early convergence.
- PGI influences convergence from around epoch 15 onward.
- Consequently, the AGM combining both generally achieves the best overall performance.

Thank You

Beomjun Bark / pbj7698@gmail.com
Yun Kwan Kim / ykwin@korea.ac.kr