

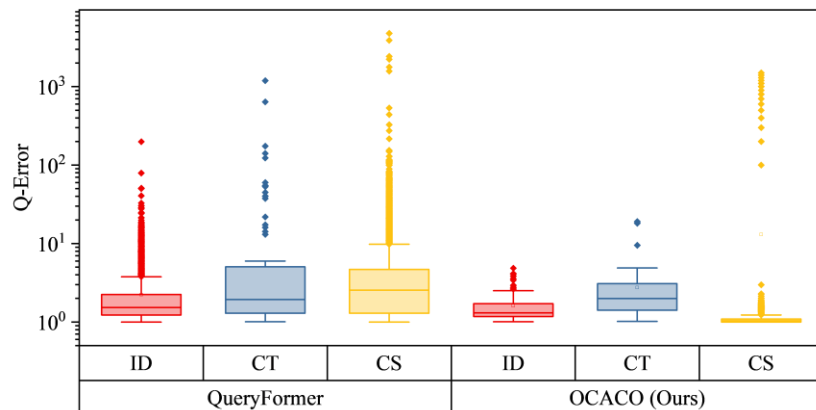
# OCACO: An Operator-level Cardinality and Cost Joint Estimator

**Tao JI, Haoyang LI, Kai ZHONG, Jing ZHANG, Cuiping LI,  
Hong CHEN**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-50348-6](https://doi.org/10.1007/s11704-025-50348-6)

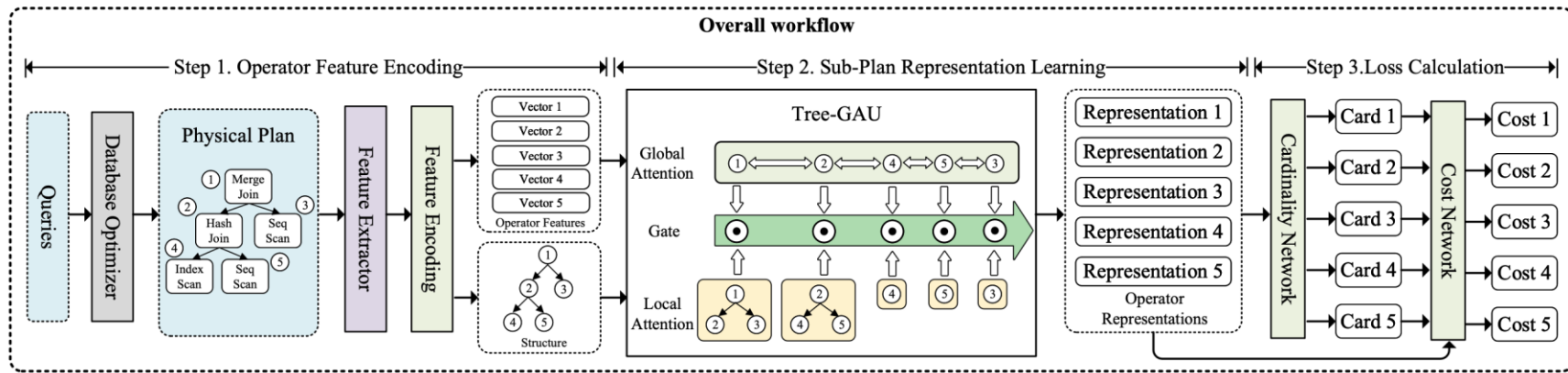
# Problems & Ideas

- Challenges:
  - Constraints in Training Data Quantity and Quality.
  - Poor Generalization and Unavailability of Predication Result.
    - Poor Generalization of Cross Sub-plan Predictions.
    - Unavailability of Plan-level CAE for COE.
  - Naive Feature Encoding.
- Ideas:
  - LLM for training queries generation.
  - A novel tree-structure-aware neural network for sub-plan representation.
  - Cardinality as input for cost estimation and operator-level training target.
  - Semantic-based operator feature encoding.

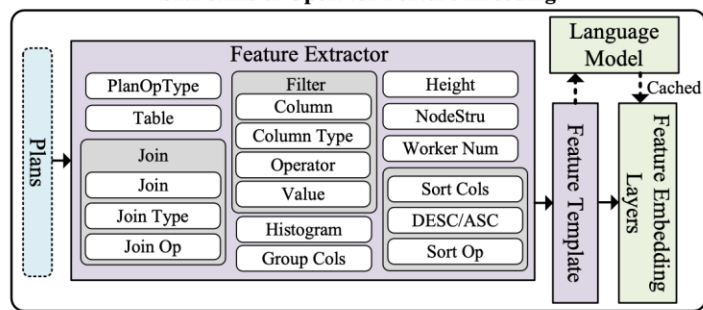


Boxplot analysis of Q-Error for COE in QueryFormer and our method on the IMDB dataset under three evaluation settings: Identical Distribution (ID), Cross Template (CT), and Cross Sub-plan (CS). In ID, evaluation plans share the same distribution as the training set. In CT, evaluation plans are generated from templates not seen during training. In CS, the model is evaluated on each sub-plans derived from the ID test set.

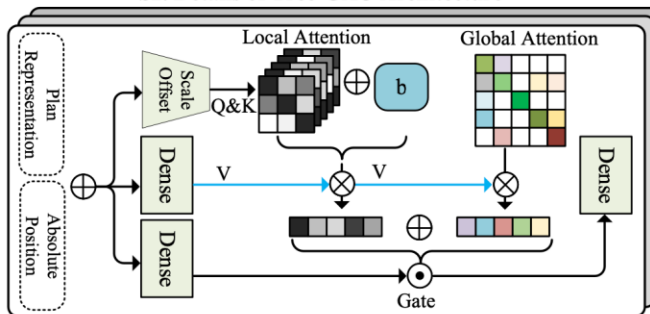
# Main Contributions



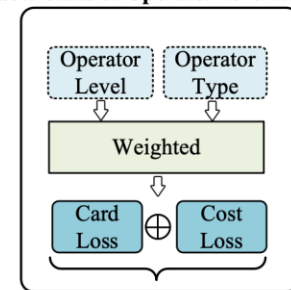
S1. Details of Operator Feature Encoding



S2. Details of Tree-GAU Architecture



S3. Details of Operator-level Loss



	PG	-	-	288	416	442	486	177	909	1140	2330	144	290	601	6730	277	7220	$1 \times 10^4$	$1 \times 10^4$
	End2End	Query	TreeLSTM	1.57	3.02	7.2	17.3	1.88	29.4	144	474	3.30	45.7	429	1280	2.87	3.43	22.3	43.3
	NEO	Query	TreeCNN	2.36	21.5	121	228	7.91	40.4	172	552	3.79	58.5	419	3720	2.00	2.91	3.54	5.61
	QueryFormer	Query	Transformer	3.14	10.1	16.6	27.5	1.93	53.3	133	812	1.36	3.53	6.30	46.9	1.35	3.06	4.77	10.5
COE	NeuroCard	Data	Auto-Regression	1.55	5.03	9.97	46.2	3.24	98.4	214	325	4.72	8.97	52.6	395	1.57	5.61	11.1	17.8
	ALECE	Hybrid	Transformer	1.14	5.64	12.9	72.7	2.75	38.3	76.8	112	1.53	7.08	14.4	52.8	1.79	6.21	12.8	28.5
	TS-OCACO	Query	TreeGAU	1.15	1.65	1.83	3.13	2.17	81.1	149	437	1.45	3.58	6.29	25.4	1.09	1.52	1.61	2.34
	ZS-OCACO	Query	TreeGAU	1.75	3.65	4.49	6.24	<b>1.40</b>	4.06	7.01	35.9	1.61	5.15	8.77	23.9	1.51	2.41	2.44	2.55
	OCACO	Query	TreeGAU	<b>1.08</b>	<b>1.29</b>	<b>1.40</b>	<b>2.71</b>	2.00	<b>3.86</b>	<b>6.81</b>	<b>18.4</b>	<b>1.31</b>	<b>3.10</b>	<b>4.88</b>	<b>17.2</b>	<b>1.07</b>	<b>1.25</b>	<b>1.33</b>	<b>1.47</b>

The proposed cardinality and cost estimation techniques significantly improve both accuracy and generalization performance. Moreover, the pre-trained model achieves strong accuracy in zero-shot deployment and few-shot fine-tuning scenarios.