AutoGT: Automated Graph Transformer Architecture Search
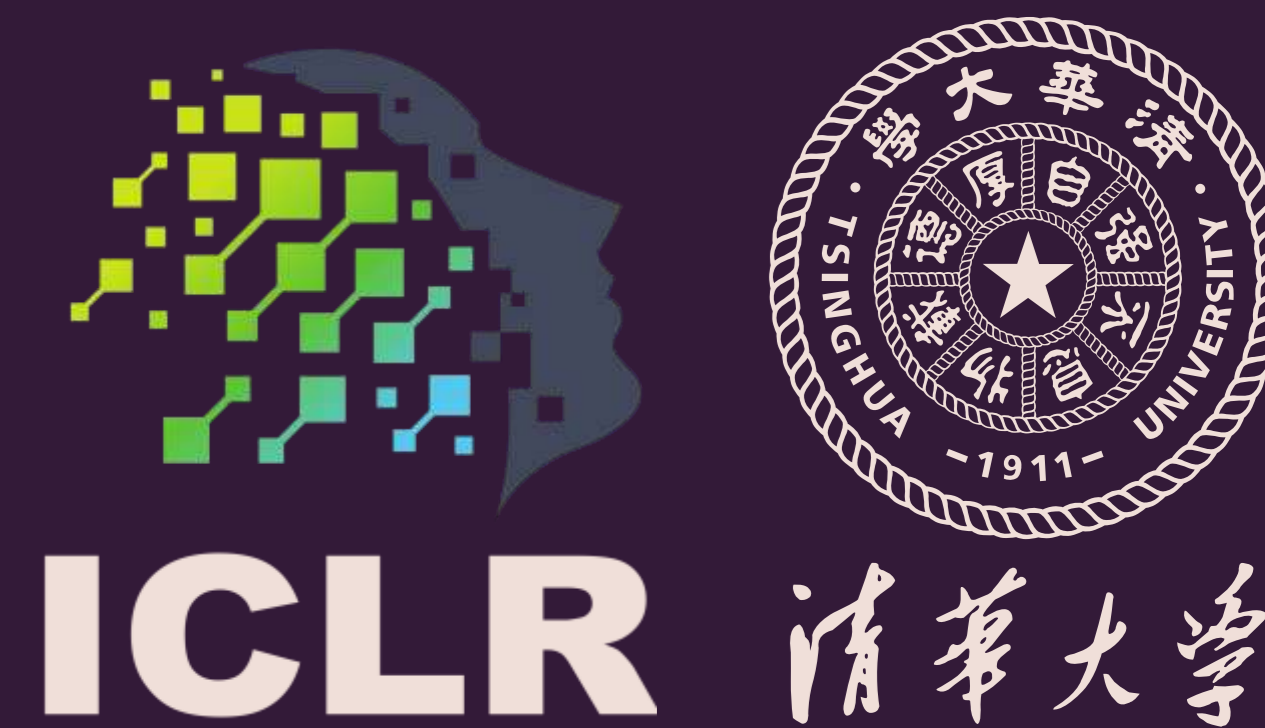
Zizhao Zhang, Xin Wang, Chaoyu Guan, Ziwei Zhang, Haoyang Li, Wenwu Zhu

Tsinghua University

ICLR 清华大学

➤ Motivation

- Graph Transformer gained success.
- *Human labor and expert knowledge* is needed for designing proper *Transformer architecture* and *graph specific encoding*.
- However, current Transformer automatic design works focus on non-graph data *without considering graph encoding*.
- How to design graph Transformer automatically?

➤ Method

- We propose **Auto**mated **G**raph **T**ransformer (AutoGT) for *Graph Transformer Neural Architecture Search*.
- 1) **Graph Transformer Search Space**: based on a unified graph Transformer formulation, including both candidate Transformer architectures and various graph encodings
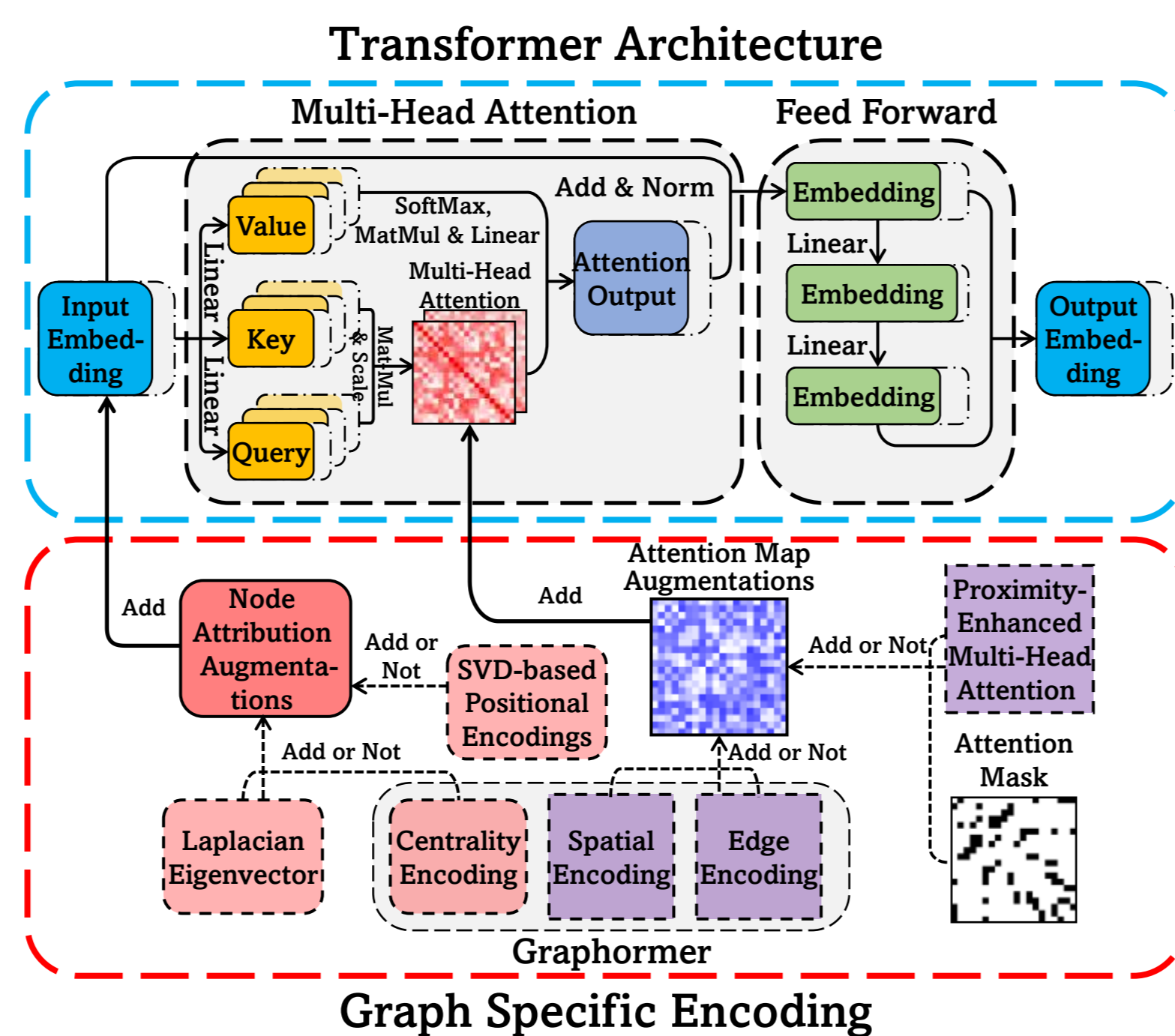
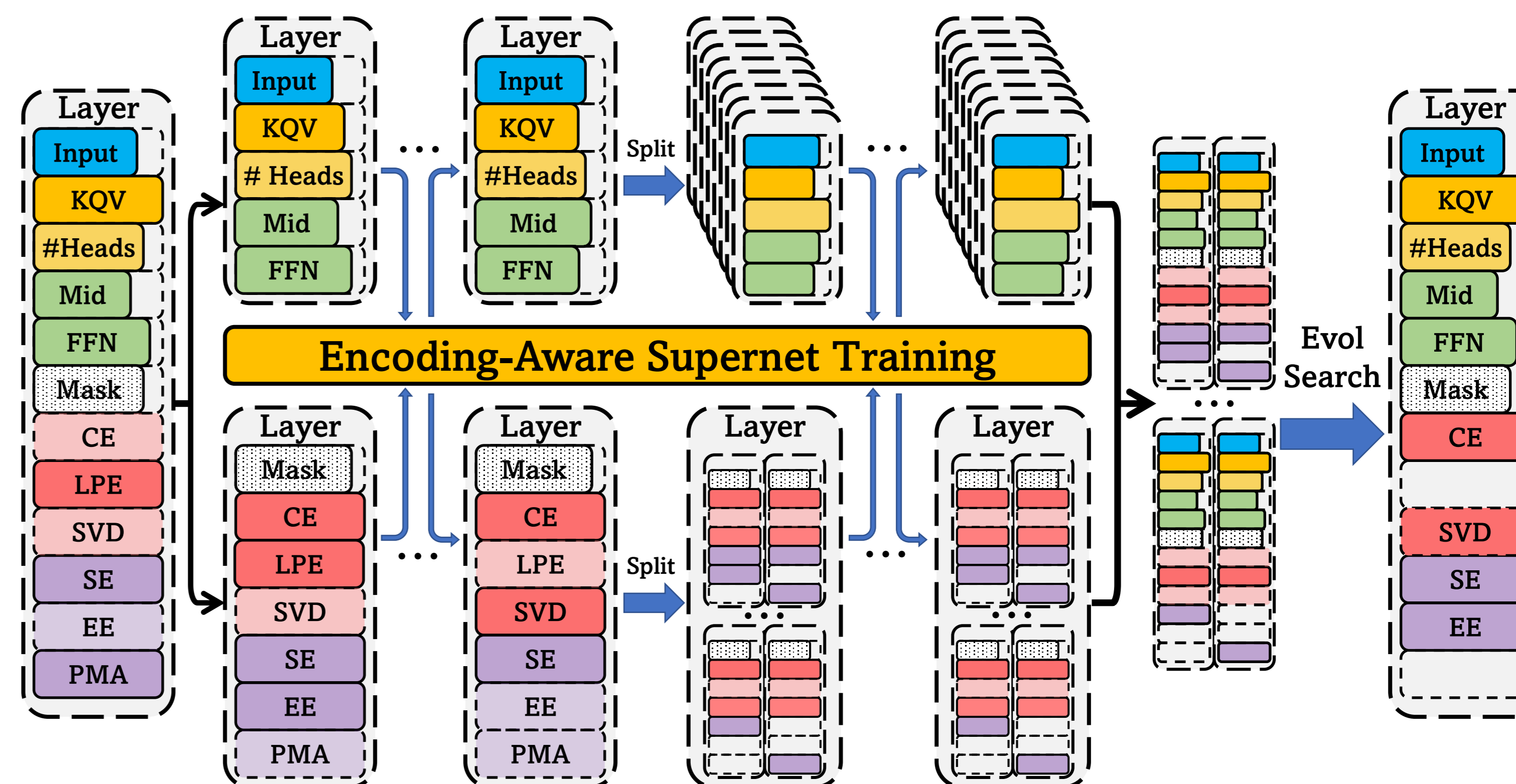- Unified graph Transformer formulation:
- Node Attribution Augmentation
- $\mathbf{H}_{aug}^{(l)} = \mathbf{H}^{(l)} + Enc_{node}(G)$
- Attention Map Augmentation
- $\mathbf{A}_{h,aug}^{(l)} = \text{softmax}\left(\dfrac{\mathbf{Q}_h^{(l)}\mathbf{K}_h^{(l)T}}{\sqrt{d}} + Enc_{map}(G)\right)$

Graph Specific Encoding

- 2) **Encoding-Aware Supernet Training**: train a single supernet first and split into several subnets according to subspaces.

➤ Experiment

Dataset	COX2_MD	BZR_MD	PTC_FM	DHFR_MD	PROTEINS	DBLP
GIN	$45.82_{14.35}$	$59.68_{14.65}$	$57.87_{8.86}$	$62.88_{8.26}$	$73.76_{4.61}$	$91.18_{0.42}$
DGCNN	$54.81_{18.51}$	$62.74_{20.59}$	$62.17_{3.62}$	$63.89_{5.91}$	$72.68_{3.75}$	$91.57_{0.54}$
DiffPool	$51.45_{14.28}$	$65.01_{14.74}$	$60.16_{5.87}$	$61.06_{9.42}$	$73.31_{3.75}$	OOT
GraphSAGE	$49.59_{12.80}$	$57.43_{13.50}$	$64.17_{3.28}$	$66.92_{2.35}$	$67.19_{6.97}$	$51.01_{0.02}$
Graphormer	$56.39_{15.03}$	$63.94_{12.58}$	$64.88_{7.58}$	$64.88_{7.58}$	$75.29_{3.10}$	$89.36_{2.31}$
GT(ours)	$54.44_{16.84}$	$63.33_{11.67}$	$64.18_{2.60}$	$65.68_{5.64}$	$73.94_{3.78}$	$90.67_{1.01}$
AutoGT(ours)	$\mathbf{59.72}_{23.26}$	$\mathbf{65.92}_{10.00}$	$\mathbf{65.60}_{3.71}$	$\mathbf{68.22}_{5.02}$	$\mathbf{77.17}_{3.40}$	$\mathbf{91.66}_{0.79}$

Dataset	OGBG-MolHIV	OGBG-MolBACE	OGBG-MolBBBP
GIN	$71.11_{2.57}$	$70.42_{4.78}$	$63.37_{1.81}$
DGCNN	$69.97_{2.16}$	$75.62_{2.64}$	$60.92_{1.78}$
DiffPool	$74.58_{1.71}$	$73.87_{4.50}$	$66.68_{6.08}$
GraphSAGE	$67.82_{3.67}$	$72.91_{1.24}$	$64.19_{3.50}$
Graphormer	$71.89_{2.66}$	$76.42_{1.67}$	$66.52_{0.74}$
AutoGT(ours)	$\mathbf{74.95}_{1.02}$	$\mathbf{76.70}_{1.42}$	$\mathbf{67.29}_{1.46}$