ICLR
International Conference On
Learning Representations

AutoGT: Automated Graph Transformer Architecture Search

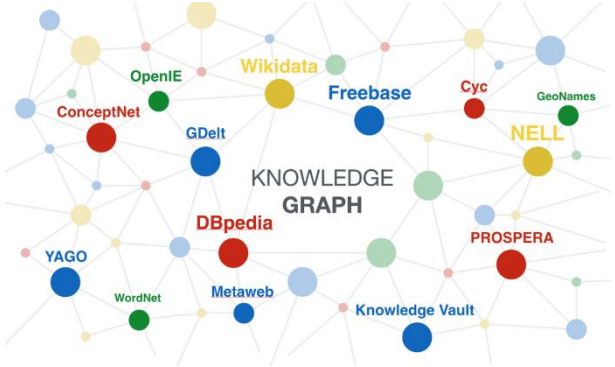
Zizhao Zhang, Xin Wang, Chaoyu Guan, Ziwei Zhang,
Haoyang Li, Wenwu Zhu

Tsinghua University

Graph Structured Data is Ubiquitous

Social Network

Knowledge Graph

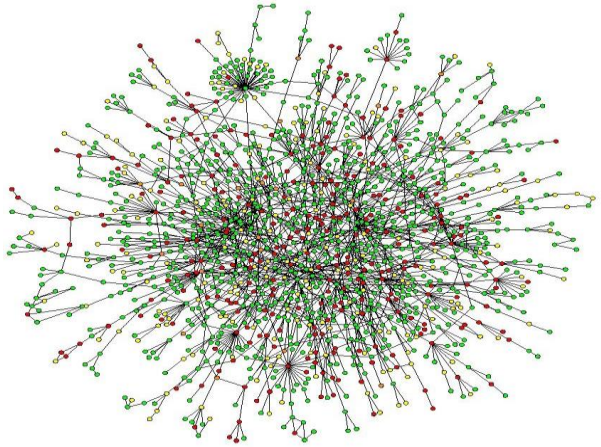


Traffic Network

Internet of Things

Information Network

Biology Network



Graph Transformer

- Graph Transformer uses graph encoding to introduce graph information to Transformer

Graph Transformer

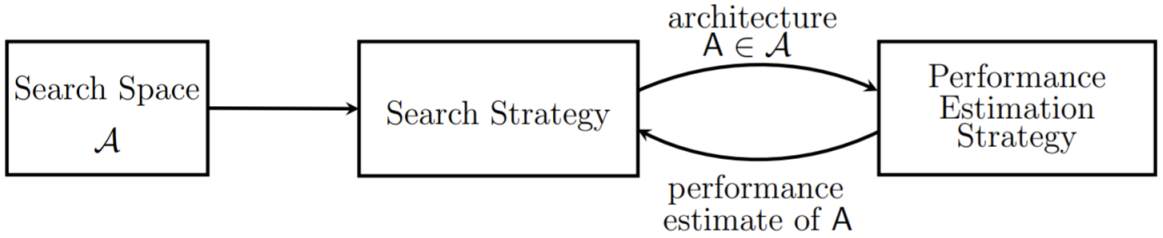
- Graph Transformer uses graph encoding to introduce graph information to Transformer
- It has attracted intensive research interests
  - UniMP achieves new state-of-the-art results on OGB
  - Graphormer wins the first place in OGB-LSC 2021
  - GPS++ wins the first place in OGB-LSC 2022

Graph Transformer

- Graph Transformer uses graph encoding to introduce graph information to Transformer
- It has attracted intensive research interests
  - UniMP achieves new state-of-the-art results on OGB
  - Graphormer wins the first place in OGB-LSC 2021
  - GPS++ wins the first place in OGB-LSC 2022
- Limitation
  - Rely on human labor and expertise knowledge to design
  - May introduce bias, which leads to sub-optimal solution

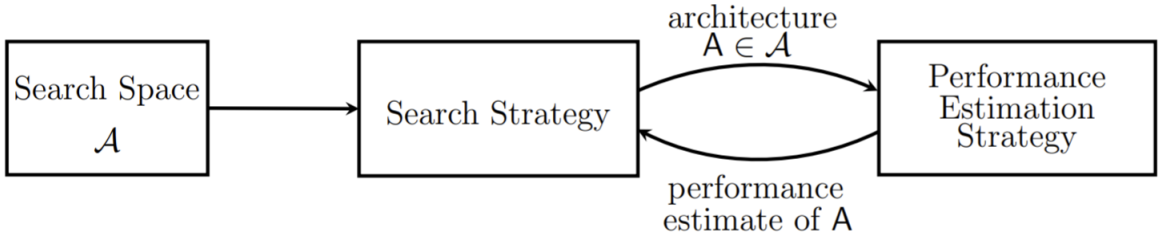
Neural Architecture Search (NAS)

- Goal: Automatically find the best neural architecture



Neural Architecture Search (NAS)

- Goal: Automatically find the best neural architecture



- Two challenges:
- How to design a unified search space for graph Transformer?
- How to tackle the coupling relations between Transformer architectures and graph encoding?

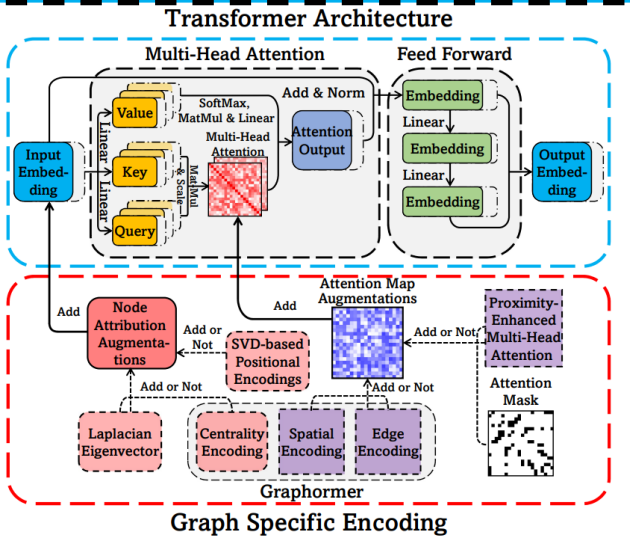
Neural Architecture Search (NAS)

- Two challenges:
- How to design a unified search space for graph Transformer?
- How to tackle the coupling relations between Transformer architectures and graph encoding?

Neural Architecture Search (NAS)

- Two challenges:
- How to design a unified search space for graph Transformer?
- How to tackle the coupling relations between Transformer architectures and graph encoding?

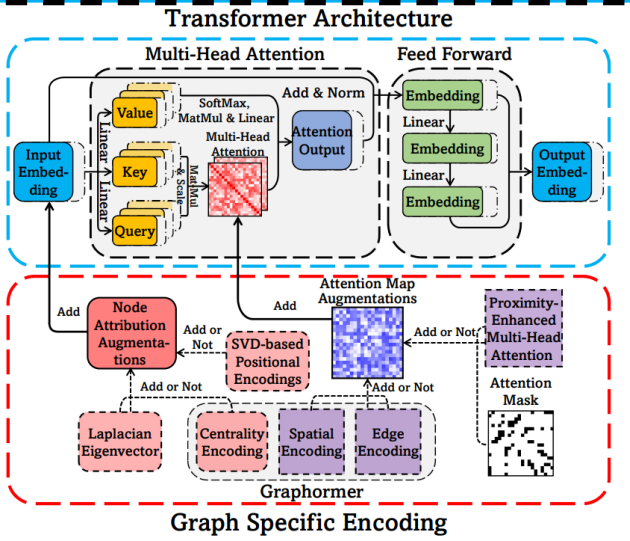
Graph Transformer Search Space



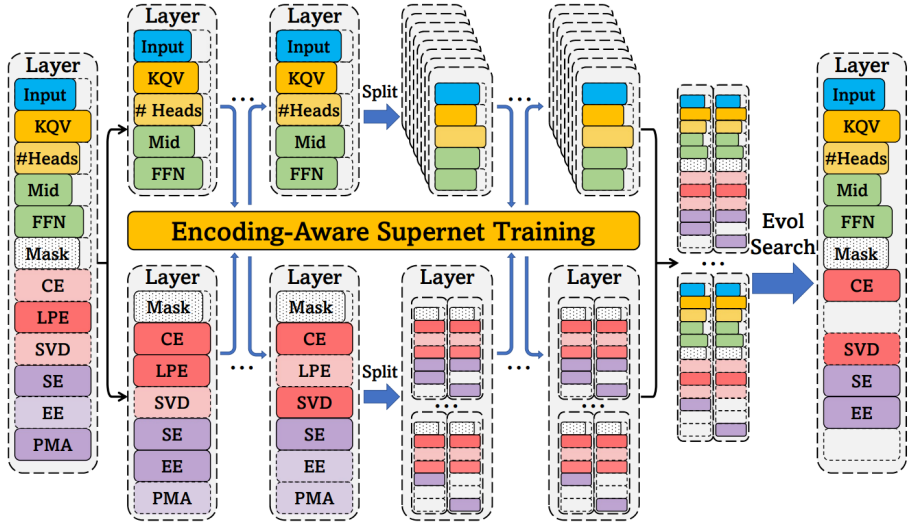
Neural Architecture Search (NAS)

- Two challenges:
- How to design a unified search space for graph Transformer?
- How to tackle the coupling relations between Transformer architectures and graph encoding?

Graph Transformer Search Space



Encoding-Aware Supernet Training



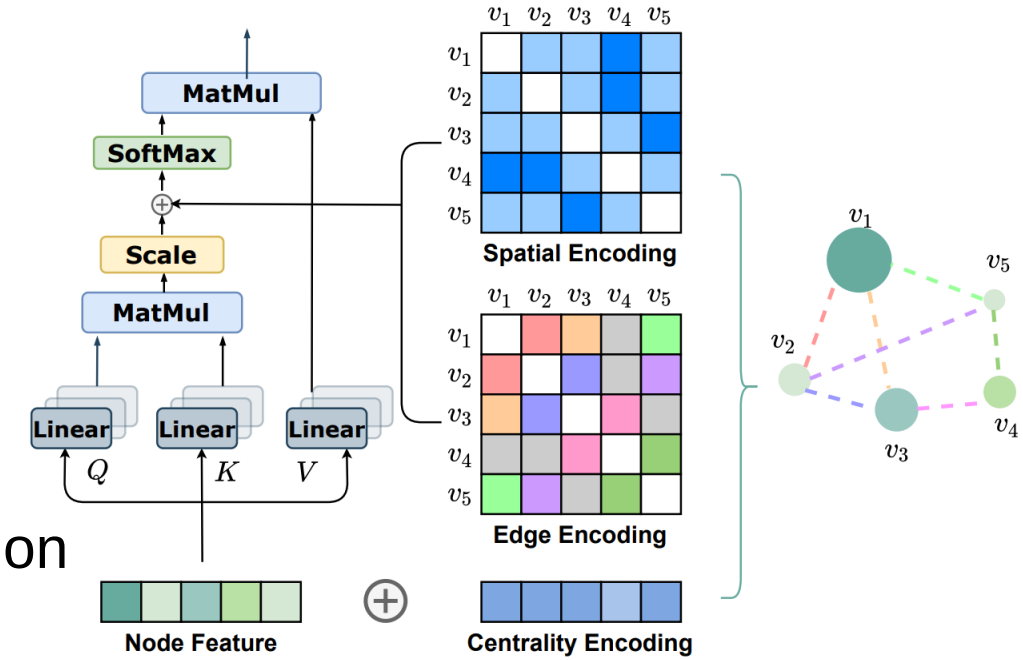
Graph Encoding

- Transformer Input: Sequential Data
- Language Data: Words in Sentence

- Visual Data: Image Patches in Image
- Graph Data: Nodes in Graph

Graph Encoding

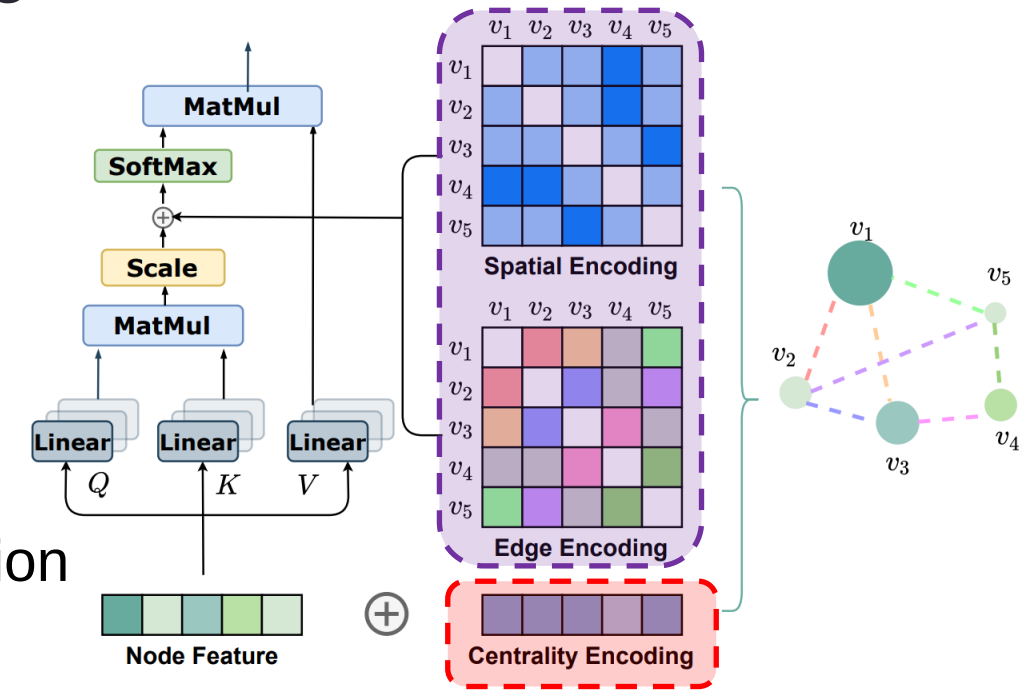
- There are two kinds of graph encoding
- Node Attribution Augmentations
  - Pre-calculate node positional encoding
  - Added to the node feature
- Attention Map Augmentations
  - Manually designed graph structural information
  - Added to the attention matrix



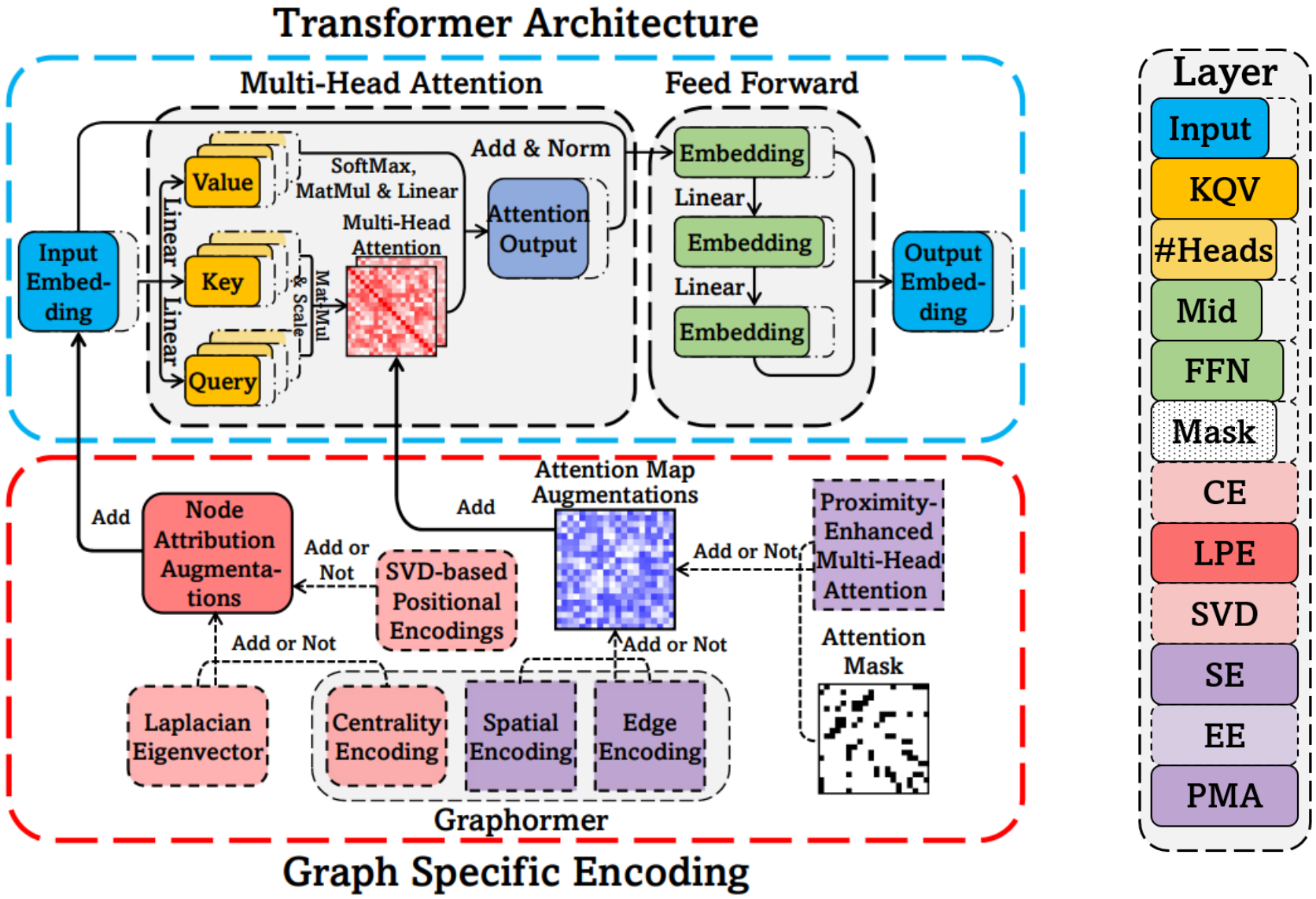
Do Transformers Really Perform Bad for Graph Representation? *NeurIPS 2021*

Graph Encoding

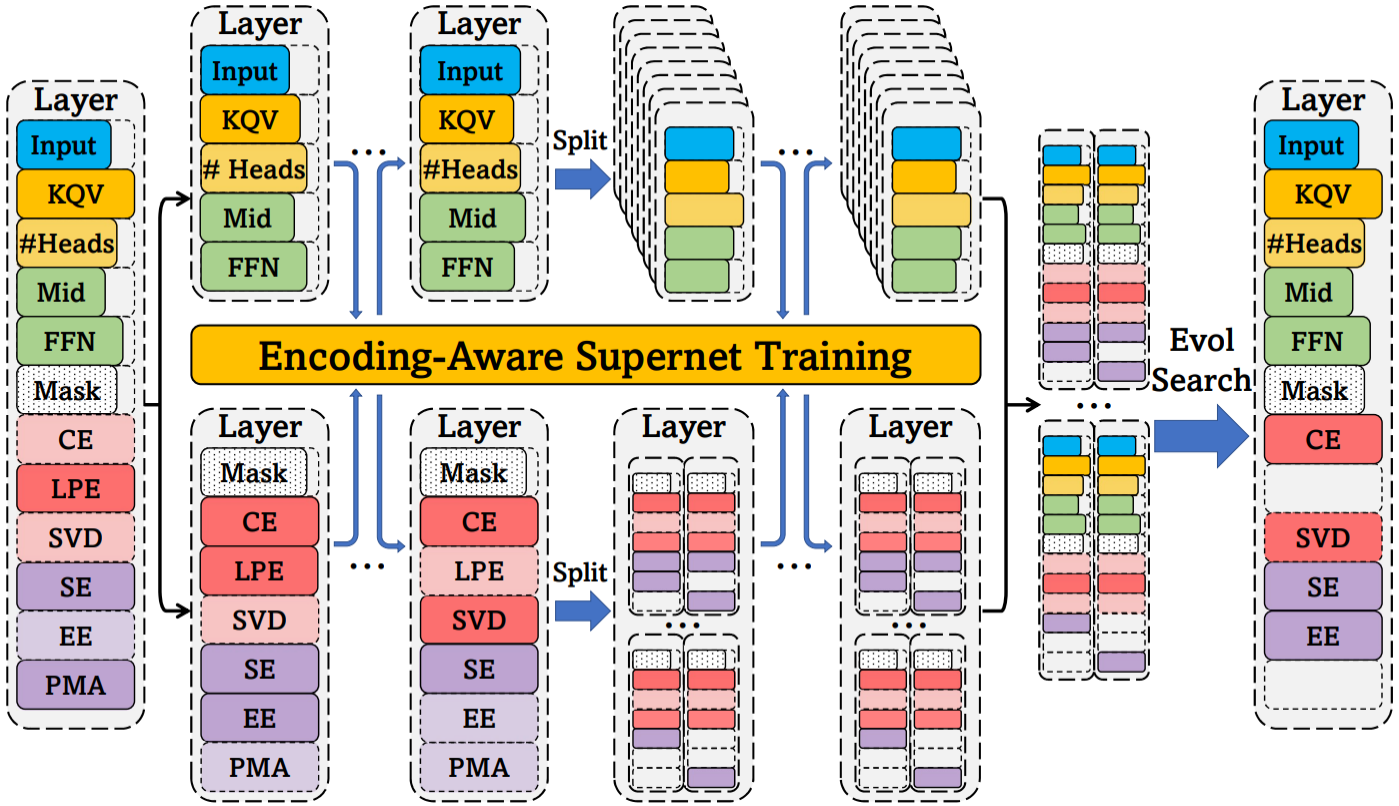
- There are two kinds of graph encoding
- Node Attribution Augmentations
  - Pre-calculate node positional encoding
  - Added to the node feature
- Attention Map Augmentations
  - Manually designed graph structural information
  - Added to the attention matrix



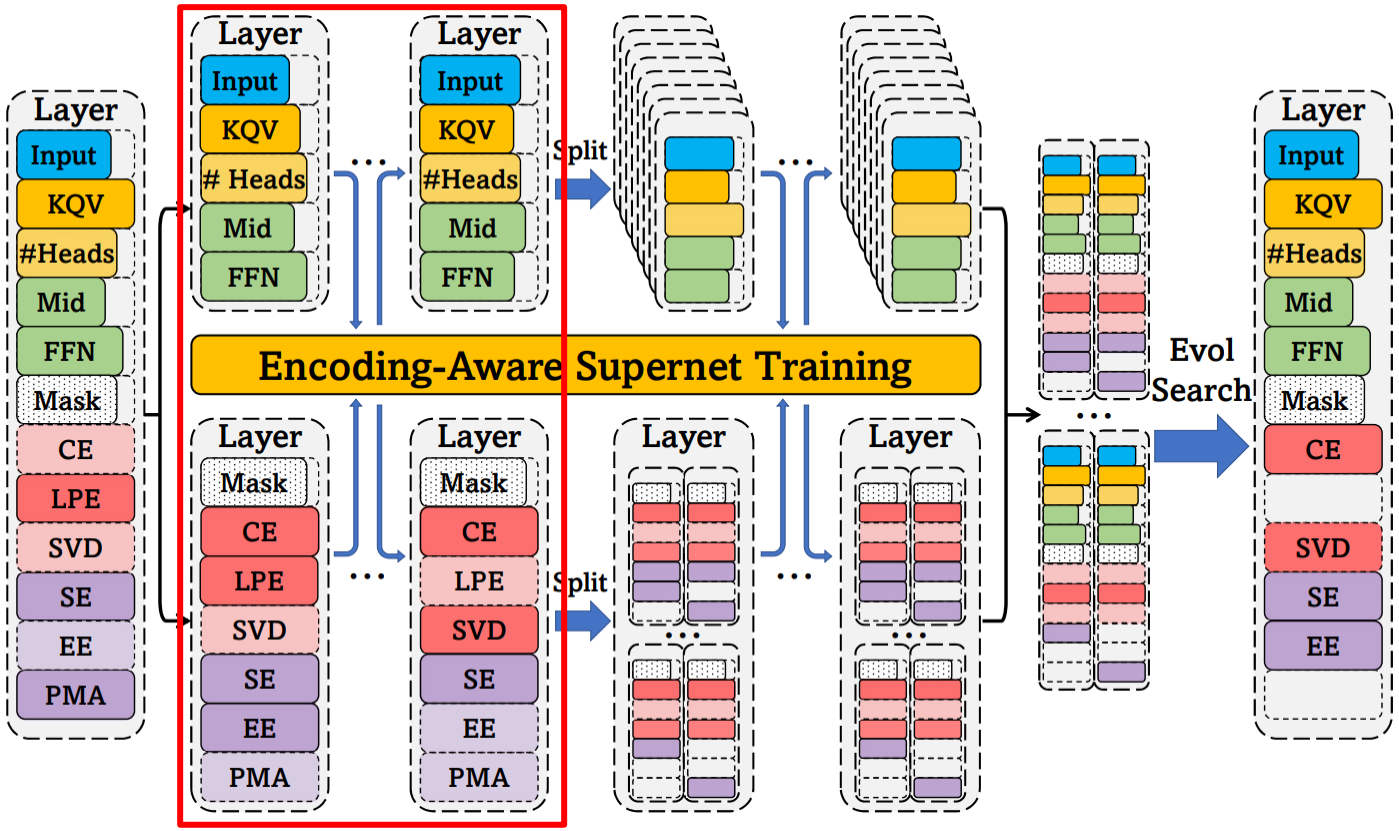
Graph Transformer Search Space



Encoding-Aware Supernet Training

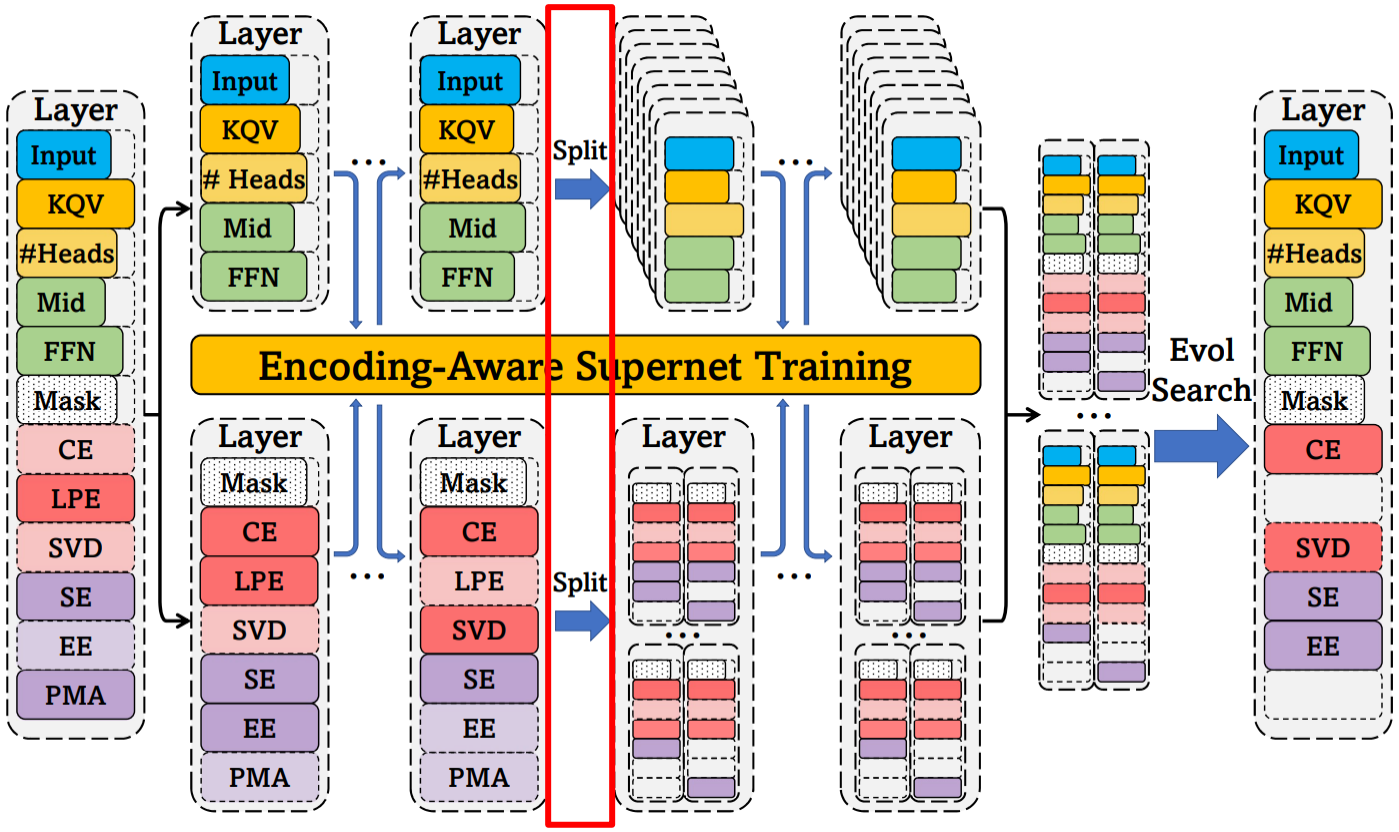


Encoding-Aware Supernet Training



① Train Supernet

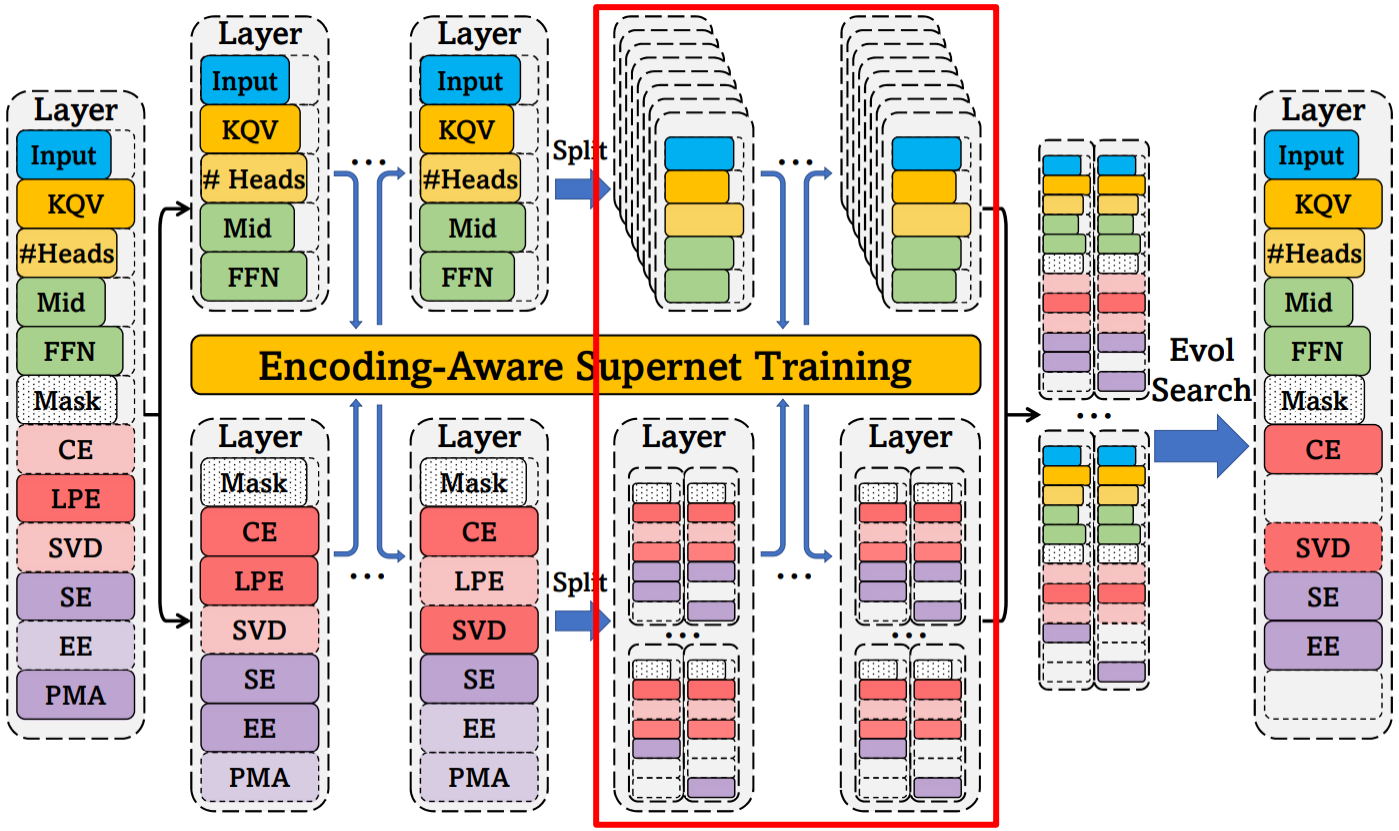
Encoding-Aware Supernet Training



① Train Supernet

② Split Supernet into Subnets

Encoding-Aware Supernet Training

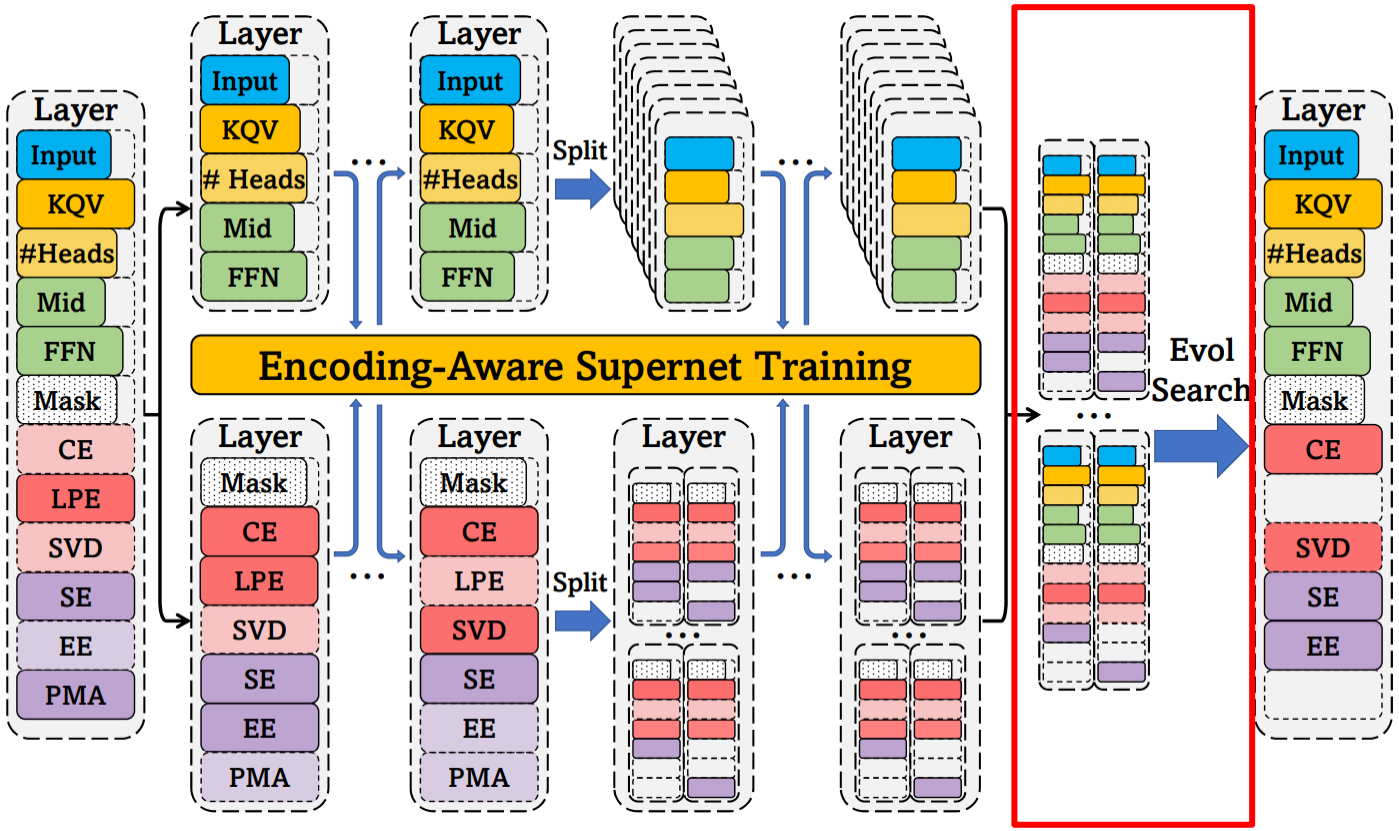


① Train Supernet

② Split Supernet into Subnets

③ Train Subnets

Encoding-Aware Supernet Training



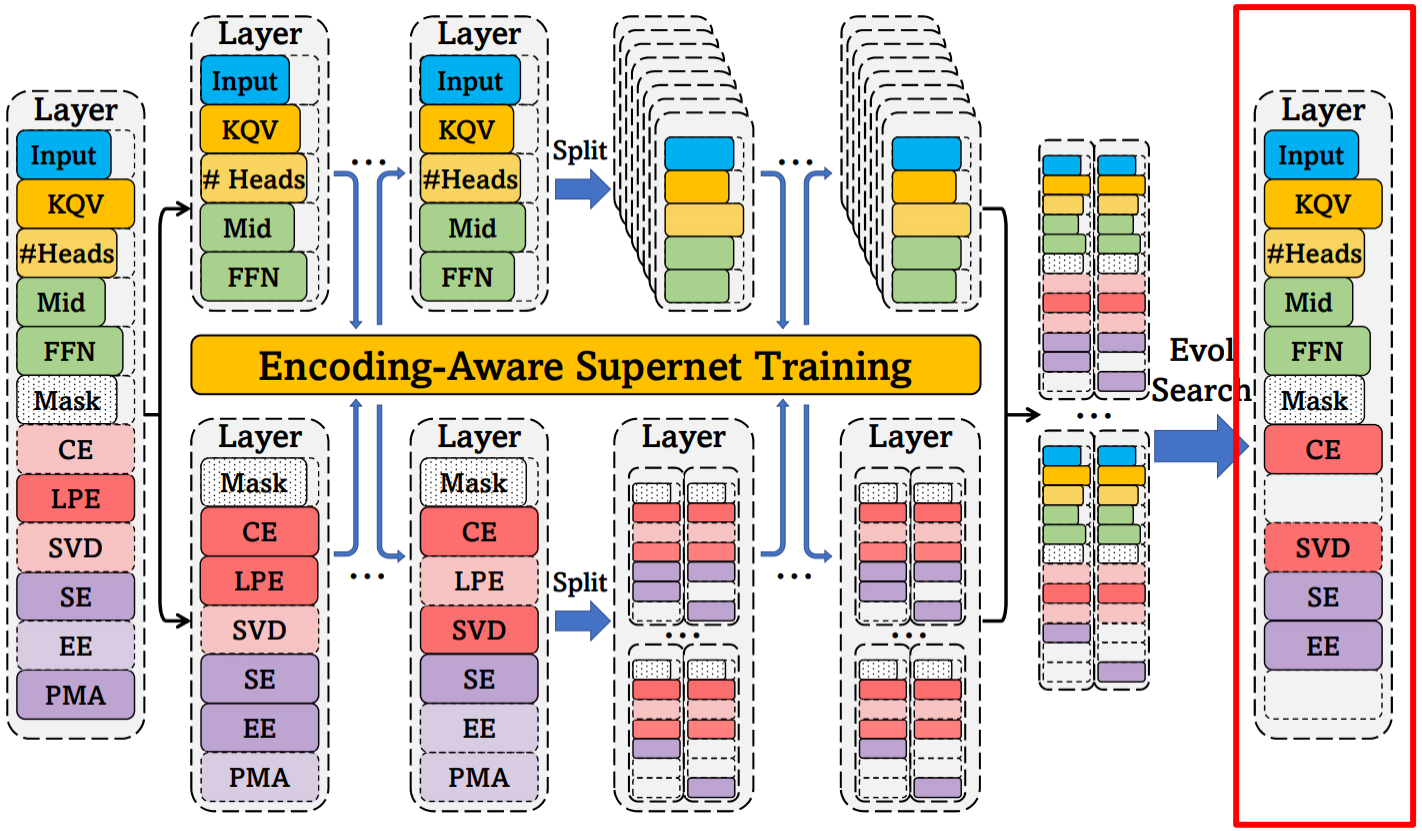
① Train Supernet

② Split Supernet into Subnets

③ Train Subnets

④ Evaluate performance using Subnets

Encoding-Aware Supernet Training



① Train Supernet

② Split Supernet into Subnets

③ Train Subnets

④ Evaluate performance using Subnets

⑤ Find the best architecture

Dataset Statistics

Dataset	#Graph	#Class	#Avg. Nodes	#Avg. Edges	# Node Feature	# Edge Feature
COX2_MD	303	2	26.28	335.12	7	5
BZR_MD	306	2	21.3	225.06	8	5
PTC_FM	349	2	14.11	14.48	18	4
DHFR_MD	393	2	23.87	283.01	7	5
PROTEINS	1,133	2	39.06	72.82	3	0
DBLP	19,456	2	10.48	19.65	41,325	3

Dataset	#Graph	#Class	#Avg. Nodes	#Avg. Edges	# Node Feature	# Edge Feature
OGBG-MolBACE	1,513	2	25.51	27.47	9	3
OGBG-MolBBBP	2,039	2	34.09	36.86	9	3
OGBG-MolHIV	41,127	2	24.06	25.95	9	3

Experimental Results

- Our method shows the best performance on various datasets

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)	$\mathbf{59.72}_{23.26}$	$\mathbf{65.92}_{10.00}$	$\mathbf{65.60}_{3.71}$	$\mathbf{68.22}_{5.02}$	$\mathbf{77.17}_{3.40}$	$\mathbf{91.66}_{0.79}$

Experimental Results

- Our method shows the best performance on various datasets
- Our encoding-aware supernet training strategy indeed improves performance

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)	$\mathbf{74.95}_{1.02}$	$\mathbf{76.70}_{1.42}$	$\mathbf{67.29}_{1.46}$

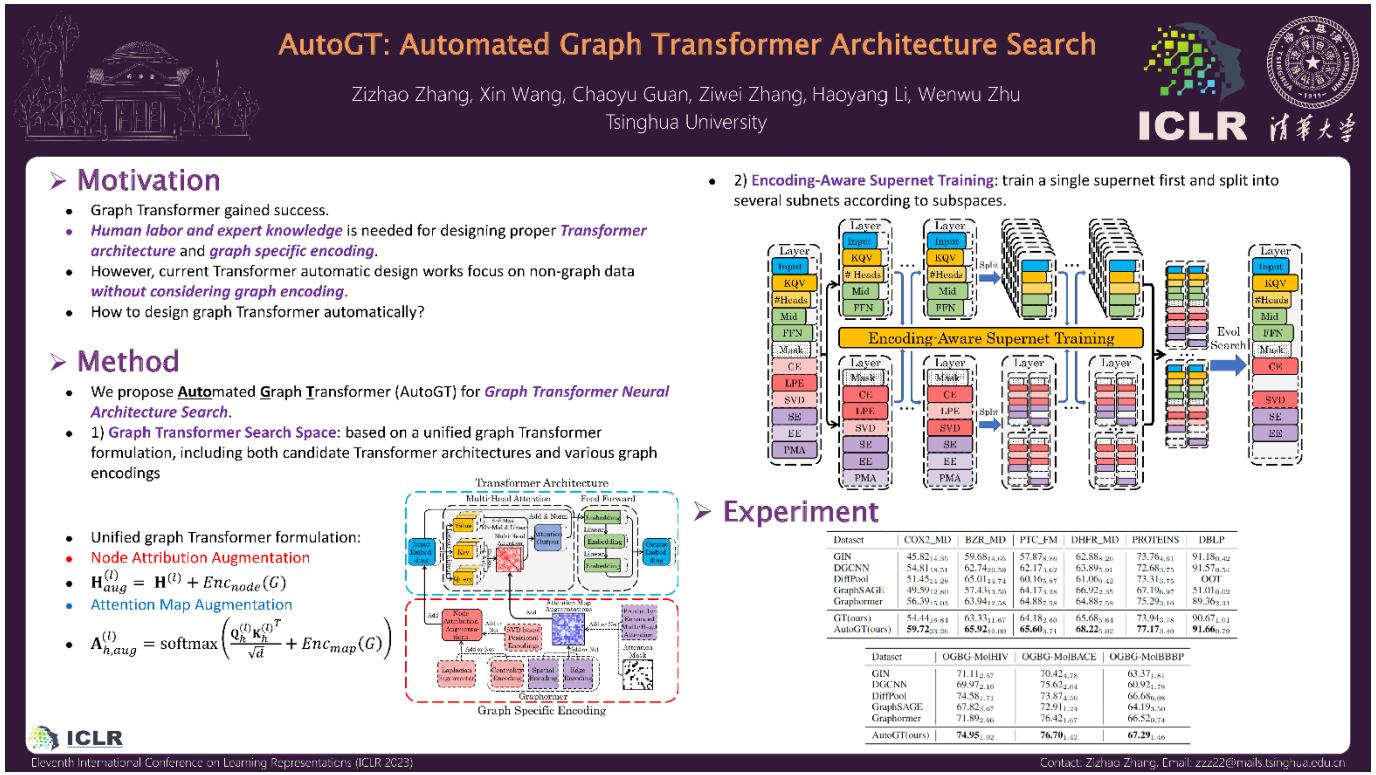
Method	Accuracy
One-Shot	$75.92_{3.10}$
Positional-Aware	$76.19_{3.42}$
AutoGT	$77.17_{3.40}$

Conclusions

- We propose AutoGT, the first neural architecture search framework for graph Transformer
- We design a unified search space for graph Transformer, and propose a novel encoding-aware supernet training strategy
- Extensive experiments demonstrate its effectiveness and wide applicability

Poster

- Time: 4:30 p.m. ~ 6:30 p.m.
- Place: MH1-2-3-4 #122



Thanks!

Zizhao Zhang, Tsinghua University

zzz22@mails.tsinghua.edu.cn

AutoGT: Automated Graph Transformer Architecture Search

Poster Location: MH1-2-3-4 #122

ICLR
International Conference On Learning Representations