Automatic Relation-aware Graph Network Proliferation

Shaofei Cai[1,2], Liang Li[1], Xinzhe Han[1,2], Jiebo Luo[3], Zhengjun Zha[4], Qingming Huang[1,2,5]

1. Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, China
2. University of Chinese Academy of Sciences, Beijing, China, 3. University of Rochester
4. University of Science and Technology of China, China, 5. Peng Cheng Laboratory, Shenzhen, China

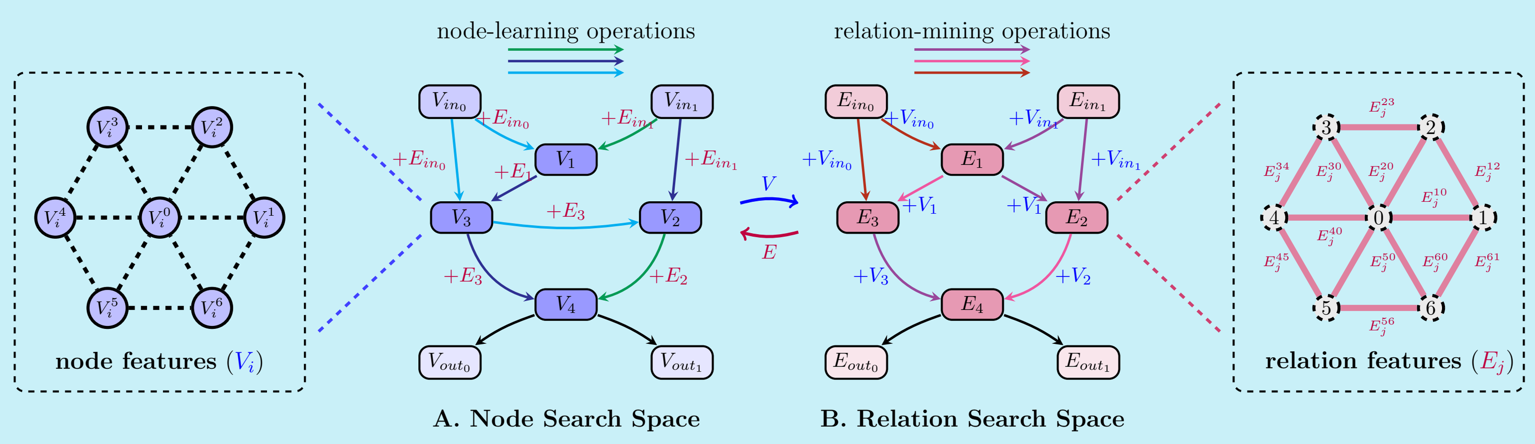
1. Abstract

We propose **Automatic Relation-aware Graph Network Proliferation** (ARGNP) for efficiently searching GNNs with a relation-guided message passing mechanism.

The main contributions are summarized as follows:

1. We devise a novel dual **Relation-aware GNN Search Space** that comprises both node and relation learning operations. These operations can extract hierarchical node (relational) information and provide anisotropic guidance for message passing on a graph.
2. We design a **Network Proliferation Search Paradigm** to progressively determine the GNN architectures by iteratively performing network division and differentiation.

2. Relation-aware GNN Search Space

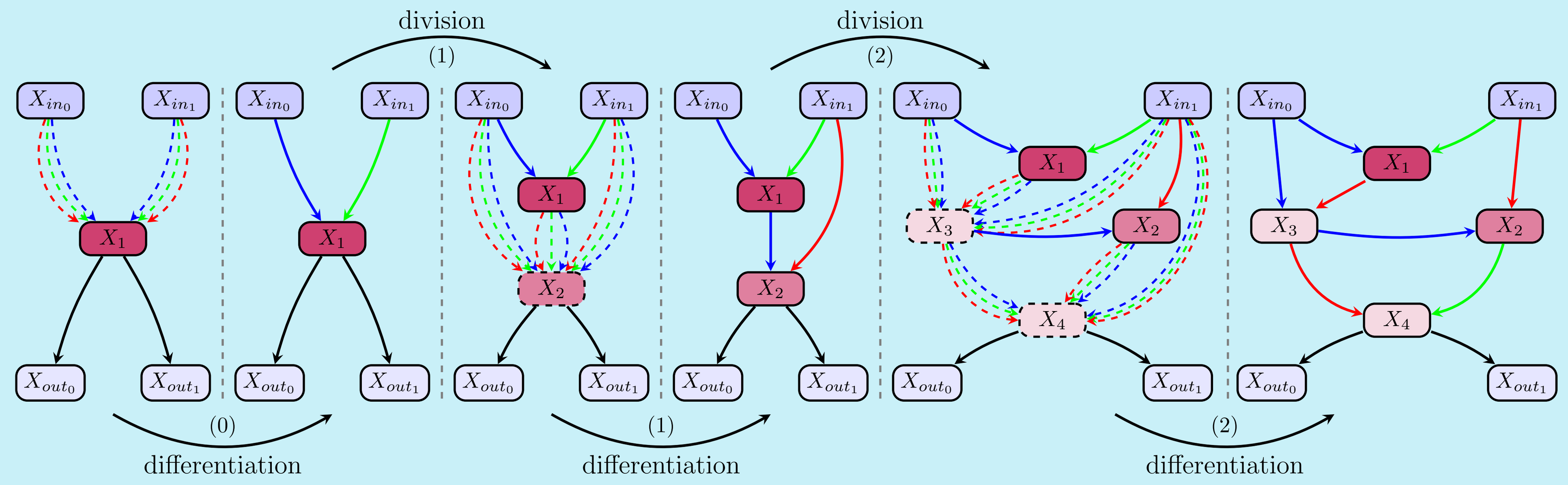


The proposed dual relation-aware gnn search space comprises:

1. **Node-learning operations** implement the anisotropic message aggregation under the guidance of relation features.

2. **Relation-mining operations** extract relational information hidden in each pair of edge-connected nodes.

3. Network Proliferation Search Paradigm



The network proliferation is an iterative process, which consists of:

1. Network division: divides each feature vertex into two parts and constructs a series of local super-networks.

2. Network differentiation: aims to differentiate each local super-network into a specific sub-network.

The spatial-temporal complexity is reduced from $O(n^2)$ to $O(n)$.

Algorithm 1 Network Proliferation Search Paradigm

Input: a search algorithm $\mathcal{A}$, architecture size $\mathcal{S}$
Output: a graph neural architecture defined by $\{\mathbb{V}, \mathbb{L}\}$
Define: $e(X_s, X_t, O)$ is a link from X_s to X_t with an operation O, where $O \in \{\boldsymbol{o}, \bar{\boldsymbol{o}}\}$, $\bar{\boldsymbol{o}}$ is the mixture operation

$\mathbb{V} \leftarrow \{X_1\}$
$\mathbb{L} \leftarrow \{e(X_{in_0}, X_1, \bar{\boldsymbol{o}}), e(X_{in_1}, X_1, \bar{\boldsymbol{o}})\}$
while True **do**
  // network differentiation
  $\mathbb{V}_r \leftarrow \mathbb{V} \cup \{X_{in_0}, X_{in_1}\}$
  Create a graph neural architecture $\mathcal{G}$ from $\{\mathbb{V}_r, \mathbb{L}\}$
  Initialize $\mathcal{G}$ with new parameters
  Perform search algorithm: $\mathbb{L} \leftarrow \mathcal{A}(\mathcal{G})$
  if $len(\mathbb{V}) \geqslant \mathcal{S}$ **then**
    return $\{\mathbb{V}_r, \mathbb{L}\}$
  // network division
  $\mathbb{V}_{tmp} \leftarrow \mathbb{V}$, $\mathbb{L}_{tmp} \leftarrow \mathbb{L}$, $l \leftarrow len(\mathbb{V})$
  for X_i in $\mathbb{V}_{tmp}$ **do**
    $\mathbb{V} \leftarrow \mathbb{V} \cup \{X_{i+l}\}$
    $\mathbb{L} \leftarrow \mathbb{L} \cup \{e(X_i, X_{i+l}, \bar{\boldsymbol{o}})\}$
    for $e(X_s, X_t, O)|_{t=i}$ in $\mathbb{L}_{tmp}$ **do**
      $\mathbb{L} \leftarrow \mathbb{L} \cup \{e(X_s, X_{i+l}, \bar{\boldsymbol{o}})\}$
  $\mathbb{L}_{tmp} \leftarrow \mathbb{L}$
  for $e(X_s, X_t, O)|_{s\in\mathbb{V}_{tmp}, s+l\neq t}$ in $\mathbb{L}_{tmp}$ **do**
    $\mathbb{L} \leftarrow \mathbb{L} \cup \{e(X_{s+l}, X_t, O)\}/\{e(X_s, X_t, O)\}$

4. Ablation Study on Search Space

Architecture	E ☑	Node Level: CLUSTER Metric (AA %) ↑	CLUSTER Params (M)	CLUSTER Search (day)	Graph Level: ZINC Metric (MAE) ↓	ZINC Params (M)	ZINC Search (day)	Graph Level: CIFAR10 Metric (OA %) ↑	CIFAR10 Params (M)	CIFAR10 Search (day)	Edge Level: TSP Metric (F1) ↑	TSP Params (M)	TSP Search (day)
GCN [29]	×	$68.50_{\pm0.98}$	0.50	ⓜ	$0.367_{\pm0.011}$	0.50	ⓜ	$56.34_{\pm0.38}$	0.10	ⓜ	$0.630_{\pm0.001}$	0.10	ⓜ
GIN [59]	×	$64.72_{\pm1.55}$	0.52	ⓜ	$0.526_{\pm0.051}$	0.51	ⓜ	$55.26_{\pm1.53}$	0.10	ⓜ	$0.656_{\pm0.003}$	0.10	ⓜ
GraphSage [21]	×	$63.84_{\pm0.11}$	0.50	ⓜ	$0.398_{\pm0.002}$	0.51	ⓜ	$65.77_{\pm0.31}$	0.10	ⓜ	$0.665_{\pm0.003}$	0.10	ⓜ
GAT [53]	×	$70.59_{\pm0.45}$	0.53	ⓜ	$0.384_{\pm0.007}$	0.53	ⓜ	$64.22_{\pm0.46}$	0.11	ⓜ	$0.671_{\pm0.002}$	0.10	ⓜ
GatedGCN [9]	✓	$76.08_{\pm0.34}$	0.50	ⓜ	$0.214_{\pm0.013}$	0.51	ⓜ	$67.31_{\pm0.31}$	0.10	ⓜ	$0.838_{\pm0.002}$	0.53	ⓜ
PNA [15]	×	N/A	N/A	N/A	$0.320_{\pm0.032}$	0.39	ⓜ	$70.46_{\pm0.44}$	0.11	ⓜ	N/A	N/A	N/A
PNA [15]	✓	N/A	N/A	N/A	$0.188_{\pm0.004}$	0.39	ⓜ	$70.47_{\pm0.72}$	0.11	ⓜ	N/A	N/A	N/A
DGN [5]	×	N/A	N/A	N/A	$0.219_{\pm0.010}$	0.39	ⓜ	$72.70_{\pm0.54}$	0.11	ⓜ	N/A	N/A	N/A
DGN [5]	✓	N/A	N/A	N/A	$0.168_{\pm0.003}$	0.39	ⓜ	$72.84_{\pm0.42}$	0.11	ⓜ	N/A	N/A	N/A
GNAS-MP [12]	×	$74.77_{\pm0.15}$	1.61	0.80	$0.242_{\pm0.005}$	1.20	0.40	$70.10_{\pm0.44}$	0.43	3.20	$0.742_{\pm0.002}$	1.20	2.10
ARGNP (2)	×	$61.61_{\pm0.27}$	0.07	0.04	$0.430_{\pm0.003}$	0.09	0.01	$66.55_{\pm0.13}$	0.10	0.11	$0.655_{\pm0.003}$	0.09	0.05
ARGNP (4)	×	$64.06_{\pm0.45}$	0.14	0.07	$0.303_{\pm0.013}$	0.14	0.01	$66.65_{\pm0.39}$	0.18	0.14	$0.668_{\pm0.003}$	0.17	0.06
ARGNP (8)	×	$68.73_{\pm0.12}$	0.25	0.20	$0.239_{\pm0.009}$	0.27	0.02	$67.37_{\pm0.32}$	0.33	0.48	$0.674_{\pm0.002}$	0.29	0.21
ARGNP (16)	×	$71.92_{\pm0.29}$	0.53	0.71	$0.221_{\pm0.004}$	0.51	0.06	$67.10_{\pm0.51}$	0.58	1.77	$0.684_{\pm0.002}$	0.56	0.76
ARGNP (2)	✓	$64.99_{\pm0.31}$	0.08	0.06	$0.318_{\pm0.009}$	0.08	0.01	$69.14_{\pm0.30}$	0.10	0.17	$0.773_{\pm0.001}$	0.08	0.08
ARGNP (4)	✓	$74.75_{\pm0.25}$	0.15	0.09	$0.197_{\pm0.006}$	0.15	0.01	$71.83_{\pm0.32}$	0.17	0.23	$0.821_{\pm0.001}$	0.14	0.10
ARGNP (8)	✓	$76.32_{\pm0.03}$	0.29	0.31	$0.155_{\pm0.003}$	0.28	0.04	$73.72_{\pm0.32}$	0.33	0.84	$0.841_{\pm0.001}$	0.30	0.39
ARGNP (16)	✓	**$77.35_{\pm0.05}$**	0.52	1.10	**$0.136_{\pm0.002}$**	0.52	0.15	**$73.90_{\pm0.15}$**	0.64	2.95	**$0.855_{\pm0.001}$**	0.62	1.23

1. Mining relational information can significantly improve the GNN's reasoning ability.
2. Searching with relation-aware GNN search space achieves higher performance with fewer parameters.

5. Ablation Study on Search Paradigm

#	Method	L (#)	Search Strategy	Cell ☑	NPSP ☑	ZINC Metric (MAE) ↓	Params (M)	Search (Day)
1	R-space	8	Random	×	×	$0.303_{\pm0.058}$	0.27	0.
2	R-space	8	DARTS	✓	×	$0.160_{\pm0.005}$	0.28	0.17
3	R-space	8	DARTS	×	×	$0.157_{\pm0.008}$	0.28	0.30
4	R-space	8	DARTS	×	✓	**$0.150_{\pm0.006}$**	0.29	**0.08**
5	R-space	8	SGAS	✓	×	$0.165_{\pm0.008}$	0.30	0.13
6	R-space	8	SGAS	×	×	$0.161_{\pm0.008}$	0.30	0.25
7	R-space	8	SGAS	×	✓	**$0.155_{\pm0.003}$**	0.28	**0.06**
8	R-space	16	Random	×	×	$0.185_{\pm0.024}$	0.51	0.
9	R-space	16	DARTS	✓	×	$0.144_{\pm0.004}$	0.57	0.38
10	R-space	16	DARTS	×	×	N/A	N/A	OOM
11	R-space	16	DARTS	×	✓	**$0.139_{\pm0.005}$**	0.56	**0.24**
12	R-space	16	SGAS	✓	×	$0.140_{\pm0.003}$	0.60	0.32
13	R-space	16	SGAS	×	×	N/A	N/A	OOM
14	R-space	16	SGAS	×	✓	**$0.136_{\pm0.002}$**	0.52	**0.21**

a. L denotes the size of the searched network.
b. Cell indicates whether to use cell-sharing trick.
c. NPSP indicates whether to use the network proliferation search paradigm.

1. The cell-sharing trick improves the search efficiency but seriously narrows the original search space and limits the final searched GNN's capability.

2. Our network proliferation search paradigm can both improves the search effect and search efficiency.

3. The proposed search paradigm works well with different search strategy (such as DARTS and SGAS).

6. Task-based Layer

Different from traditional GNNs whose global graph representation is only constructed on the readout of node features. Our method explicitly models relational information, so it naturally constructs global graph representation with both node and relation features.

The Global Node Feature:

$$V_g = \sigma(BN([V_1 \parallel \cdots \parallel V_L]))$$

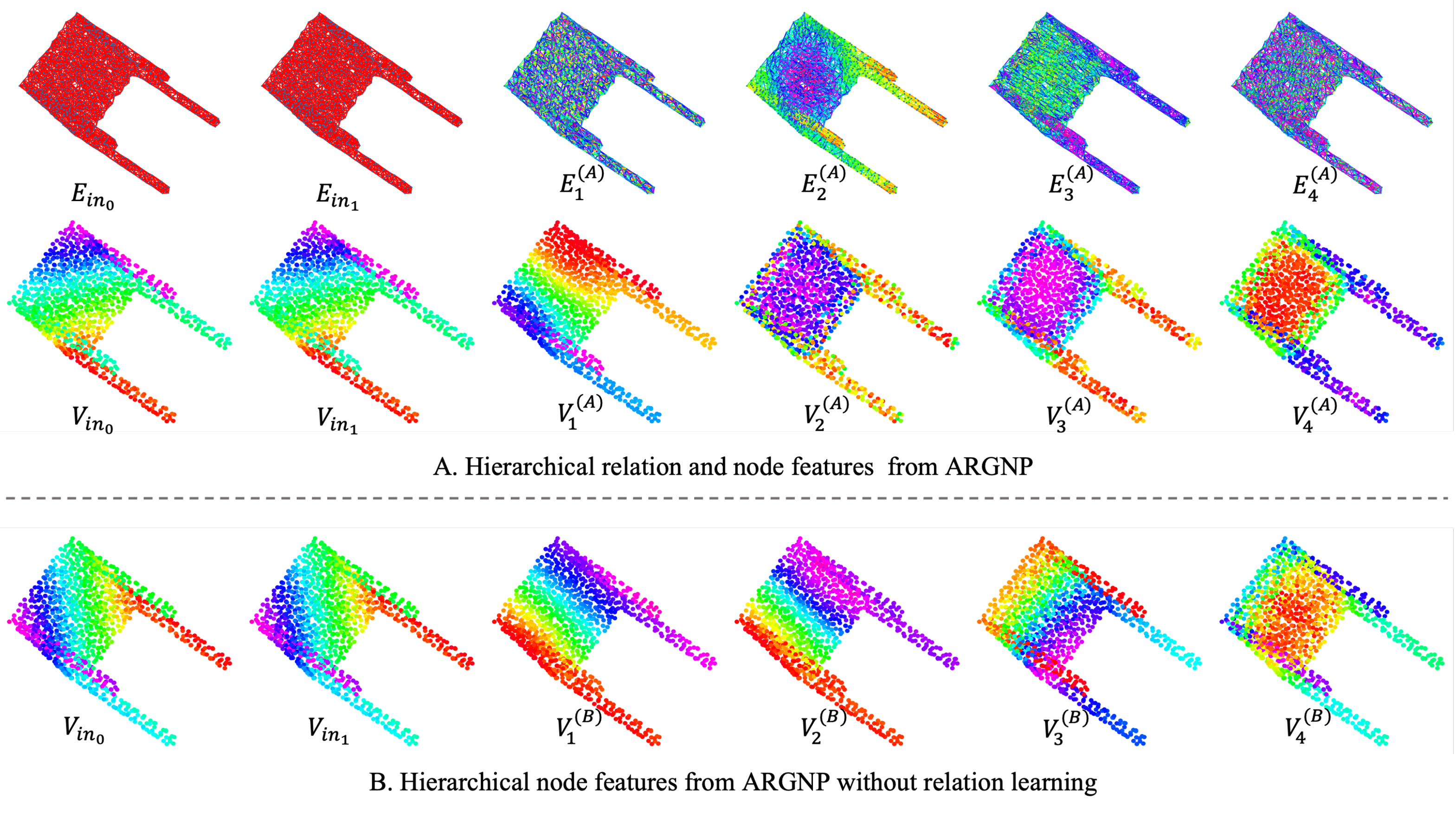
The Global Relation Feature:

$$E_g = \sigma(BN([E_1 \parallel \cdots \parallel E_L]))$$

The Global Graph Representation:

$$G_g = \left[\frac{1}{|V_g|}\sum_{i\in V_g} V_g^i \parallel \frac{1}{|E_g|}\sum_{j\in E_g} E_g^i\right]$$

7. Visualizing Hierarchical Features

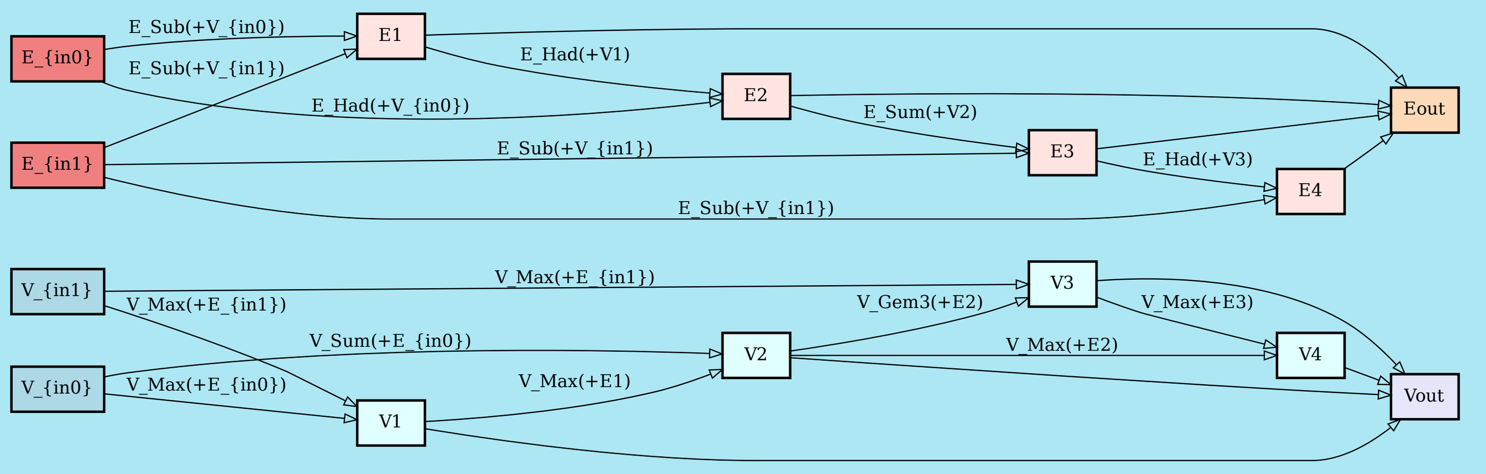


Visualization of the learned hierarchical features for 3D point cloud recognition

1. The learned hierarchical relation features represent different message passing preferences and can guide better message passing mechanisms to learn more effective node features.

2. ARGNP can capture the structural information and well discriminate different parts of the object, which is significantly better than traditional GNNs without relation learning architecture.

8. Searched GNN Results



The searched GNN architecture with the size of 4 on the ModelNet.

1. In the node search space, it prefers to choose V_Max as the node-learning operation.

2. In the relation search space, it prefers to select E_Sub and E_Had as the relation-mining operation.

Thanks for your attention.