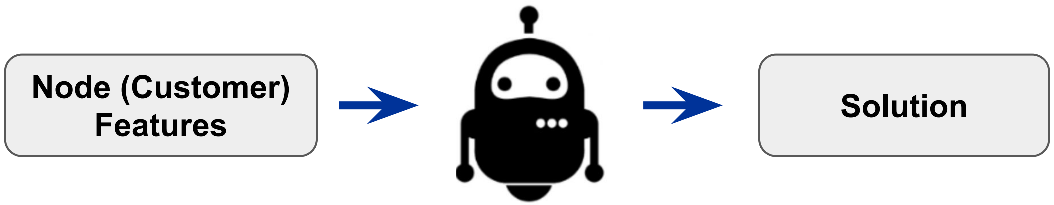
Learning to Iteratively Solve Routing Problems with Dual-Aspect Collaborative Transformer

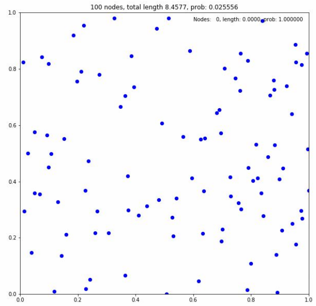
Yining Ma[1], Jingwen Li[1], Zhiguang Cao[2], Wen Song[3], Le Zhang[4], Zhenghua Chen[5], Jing Tang[6]

[1]National University of Singapore
[2]Singapore Institute of Manufacturing Technology, A*STAR
[3]Institute of Marine Science and Technology, Shandong University
[4]University of Electronic Science and Technology of China
[5]Institute for Infocomm Research, A*STAR
[6]The Hong Kong University of Science and Technology

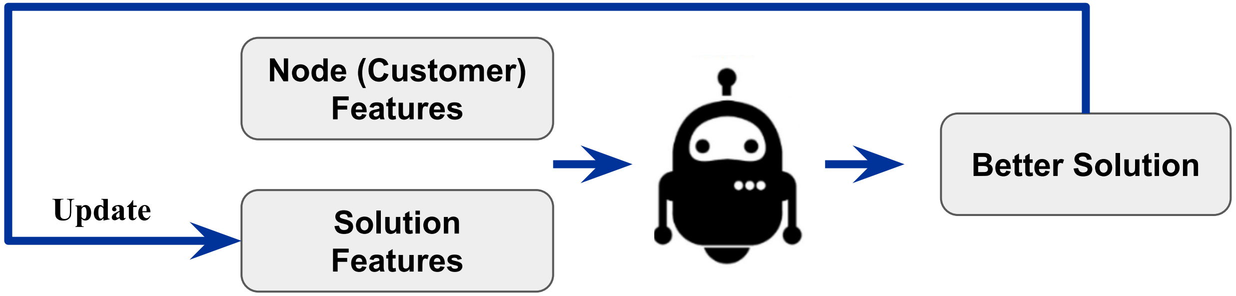
1 Types of Neural Solver for VRPs: Construction / Improvement

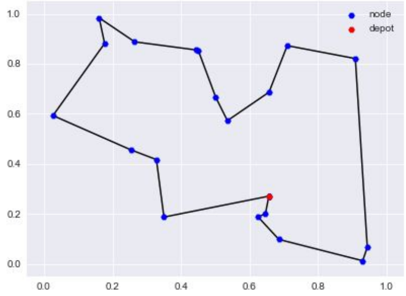
Different from construction solvers, improvement solvers need to encode VRP solutions properly.





Neural Construction Solver

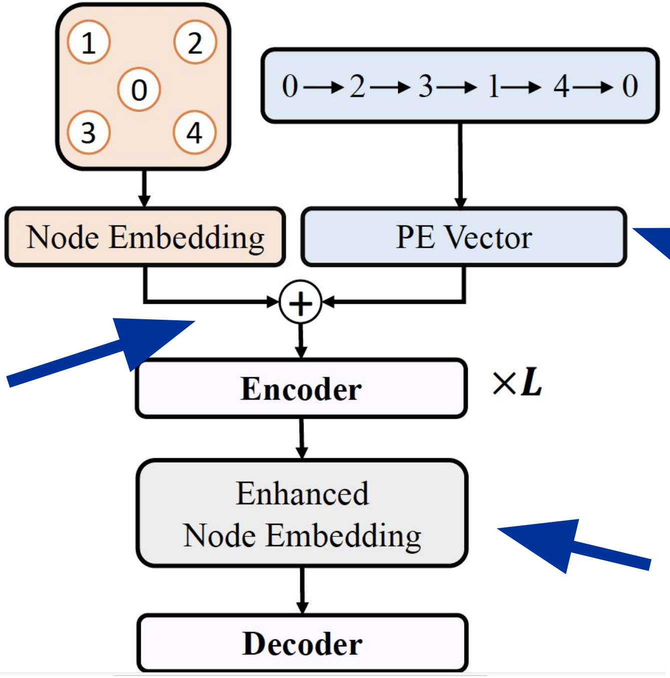




Neural Improvement Solver (Our focus)

2 Although Transformer has been shown effective for processing sequence data, its positional encoding (PE) may not be optimal for encoding VRP solutions.

Node Features:
Coordinates, demand, etc.



Node Positional Features:
Position of node in the solution sequence

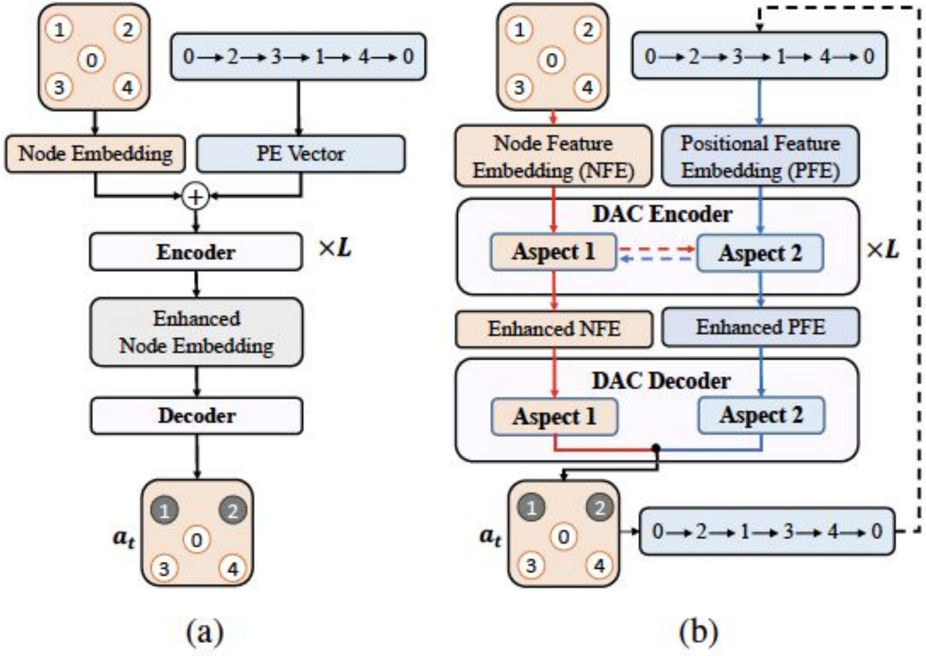
1. Embedding Fusion through an Addition Operator
(Limitation: mixed correlations can bring unreasonable noises and random biases to the encoder)

2. Learning a Unified Set of Embeddings
(Limitation: may cause disharmony or disturbance for VRP tasks)

3. For Encoding Linear Sequences Only
(Limitation: cannot capture the circularity and symmetry of VRP solution, significantly hurt generalization performance)

3. Our Proposed Method and Main Contribution

a) Dual-Aspect Representation - Better Transformer-style encoder for VRPs



- **Learning Two Sets of Embeddings**
- **DAC Encoder**
  - **DAC-Att**
  - **Cross-Aspect Referential Attention**

Table 3: Dual v.s. single aspect representation

Steps	Method	# Params	N=50	N=100
T=1k	SA-T	0.37M	0.35% (1m)	3.49% (3m)
	DACT	0.29M	**0.14%** (1m)	**1.62%** (4m)
T=5k	SA-T	0.37M	0.05% (5m)	1.55% (16m)
	DACT	0.29M	**0.02%** (6m)	**0.61%** (18m)

3. Our Proposed Method and Main Contribution

b) Cyclic Positional Encoding

- **We enable Transformer for encoding cyclic sequences**

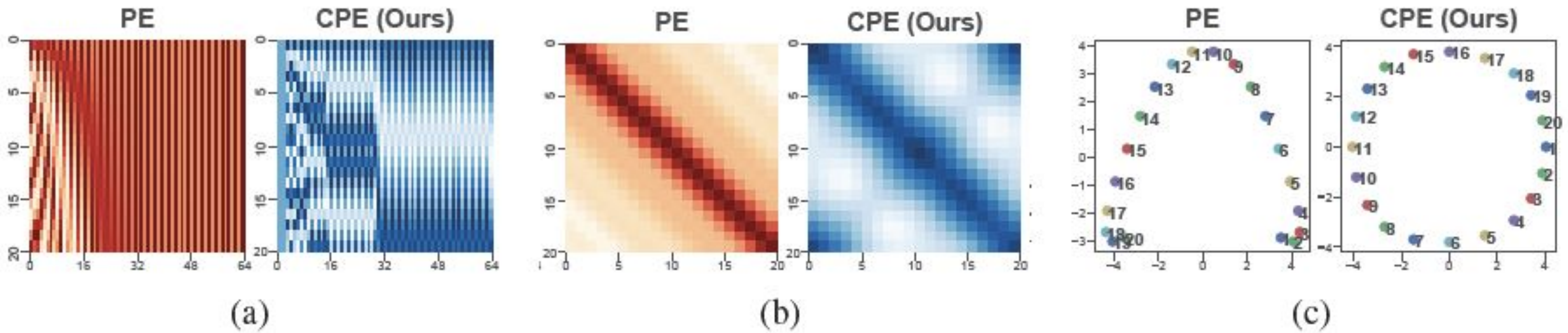


Figure 5: Comparison of our CPE method with absolute PE method on a TSP instance with 20 nodes. (a) the embedding vectors, (b) the correlations (dot products) between every two embeddings, and (c) the top two principal components after PCA (principal component analysis) projection.

3. Our Proposed Method and Main Contribution

b) Cyclic Positional Encoding -> Better Generalization on synthetic and benchmark instances

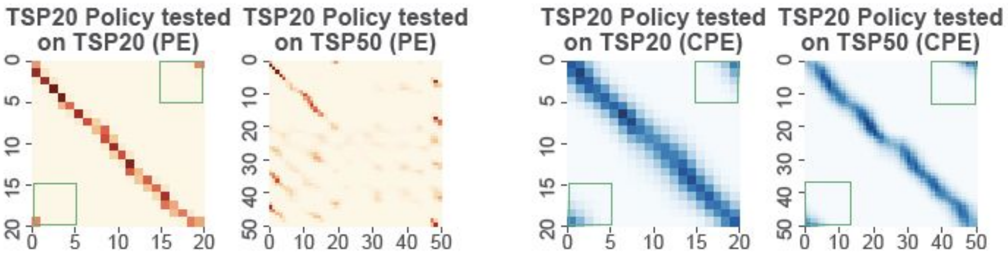
Table 2: Generalization performance. (a) DACT v.s. baselines on benchmark datasets (up to 200 customers, see Appendix E.4 for detailed results and discussion); (b) PE v.s. CPE on different sizes.

Method	TSPLIB	CVRPLIB
OR-Tools [37]	3.34%	8.06%
AM-sampling [5]	22.83%	26.66%
POMO [8]	10.06%	6.10%
Wu et al. [11]	4.17%	5.20%
DACT	**2.07%**	**3.41%**

(a)

Method	N=20		N=100	
	Obj.	Gap	Obj.	Gap
DACT-PE (T=5k)	3.84	0.21%	8.38	7.93%
DACT-CPE (T=5k)	**3.83**	**0.10%**	**7.99**	**2.98%**
Wu et al. [11] (T=5k)	3.91	2.14%	9.03	16.37%
OR-Tools [37]	3.83	0.00%	8.06	3.87%

(b)



(a) (b)

Figure 6: Visualization of the attention scores for the encoder when a trained model is used to solve instances with a larger size. (a) using PE method; (b) using CPE method (ours).

3. Our Proposed Method and Main Contribution

c) Curriculum Learning strategy - Better RL algorithm for Neural Combinatorial Optimization

Training with n-step PPO and a Curriculum Learning Strategy:

- Our CL strategy:
  - gradually prescribes higher-quality solutions as the initial states for training.

- Benefits:
  - Better sample efficiency while reducing the variance of training

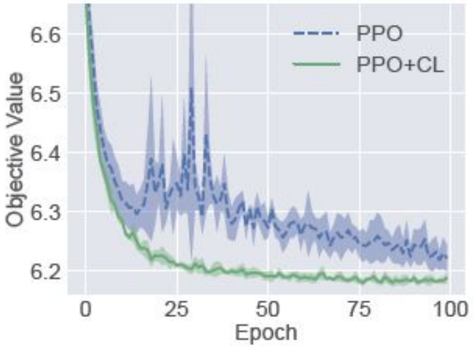


Figure 7: Training curves of PPO with and without CL on CVRP20 (random seeds 1-5).

4. Results on TSP and CVRP

Our DACT advances the current SOTA on learning purely data-driven improvement VRP solver.

	Method	N=20			N=50			N=100		
		Obj.	Gap	Time	Obj.	Gap	Time	Obj.	Gap	Time
TSP	Concorde	3.83	-	(3m)	5.70	-	(10m)	7.76	-	(1h)
	LKH	3.83	0.00%	(38s)	5.70	0.00%	(5m)	7.76	0.00%	(20m)
	OR-Tools	3.86	0.94%	(42s)	5.85	2.87%	(5m)	8.06	3.86%	(23m)
	Neural-2-Opt [23]	3.84‡	**0.00%**	(15m)	5.70	0.12%	(29m)	7.83	0.87%	(41m)
	Wu et al. [11] (T=5k)	3.83	**0.00%**	(1h)	5.70‡	0.20%	(1.5h)	7.87	1.42%	(2h)
	DACT (T=1k)	3.83	0.04%	{7s}(24s)	5.70	0.14%	{16s}(1m)	7.89	1.62%	{48s}(4m)
	DACT (T=5k)	3.83	**0.00%**	{32s}(2m)	5.70	0.02%	{2m}(6m)	7.81	0.61%	{4m}(18m)
	DACT (T=10k)	3.83	**0.00%**	{1m}(5m)	5.70	0.01%	{3m}(13m)	7.79	0.37%	{8m}(40m)
	DACT$\times$4 augment	3.83	**0.00%**	{3m}(10m)	5.70	**0.00%**	{10m}(1h)	7.77	**0.09%**	{29m}(2.5h)
	GCN-BS [6]	3.84‡	0.01%	(12m)	5.70	0.01%	(18m)	7.87	1.39%	(40m)
	AM-sampling [5]	3.84‡	0.08%	(5m)	5.73	0.52%	(24m)	7.94	2.26%	(1h)
	MDAM-BS [7]	3.84‡	**0.00%**	(3m)	5.70	0.03%	(14m)	7.79	0.38%	(44m)
	POMO [8]	3.83	0.04%	(1s)	5.70‡	0.21%	(2s)	7.80	0.46%	(11s)
	POMO$\times$8 augment [8]	3.83	**0.00%**	(3s)	5.69‡	0.03%	(16s)	7.78	0.15%	(1m)
	DPDP (100k) [26]	-	-	-	-	-	-	7.77‡	0.00%	(3h)
	CVAE-Opt-DE [13]	-	**0.00%**$^{\#}$	11m$^{\#}$	-	0.02%$^{\#}$	22m$^{\#}$	-	0.34%$^{\#}$	55m$^{\#}$
CVRP	LKH	6.14	0.00%	1h	10.38	0.00%	4h	15.68	0.00%	8h
	OR-Tools	6.46	5.68%	2m	11.27	8.61%	13m	17.12	9.54%	46m
	NeuRewriter [4]	6.15$^{\#}$	-	6m$^{\#}$	10.51$^{\#}$	-	11m$^{\#}$	16.10$^{\#}$	-	17m$^{\#}$
	NLNS [27]	6.19$^{\#}$	-	6m$^{\#}$	10.54$^{\#}$	-	11m$^{\#}$	15.99$^{\#}$	-	16m$^{\#}$
	Wu et al. [11] (T=5k)	6.12‡	0.39%	(2h)	10.45	0.70%	(4h)	16.03‡	2.47%	(5h)
	DACT (T=1k)	6.15	0.28%	{16s}(33s)	10.61	2.13%	{43s}(2m)	16.17	3.18%	{2m}(5m)
	DACT (T=5k)	6.13	-0.00%	{1m}(3m)	10.48	1.01%	{3m}(8m)	15.92	1.55%	{8m}(23m)
	DACT (T=10k)	6.13	-0.04%	{2m}(6m)	10.46	0.79%	{6m}(16m)	15.85	1.12%	{16m}(45m)
	DACT$\times$6 augment	6.13	**-0.08%**	{11m}(35m)	10.39	**0.14%**	{32m}(1.5h)	15.71	**0.19%**	{1.5h}(4.5h)
	AM-sampling [5]	6.25	1.87%	(6m)	10.62	2.40%	(28m)	16.23‡	3.72%	(2h)
	MDAM-BS [7]	6.14	0.18%	(5m)	10.48	0.98%	(15m)	15.99‡	2.23%	(1h)
	POMO [8]	6.17‡	0.82%	(1s)	10.49	1.14%	(4s)	15.83	0.98%	(19s)
	POMO$\times$8 augment [8]	6.14‡	0.21%	(5s)	10.42	0.45%	(26s)	15.73	0.32%	(2m)
	DPDP (100k) [26]	-	-	-	-	-	-	15.69‡	0.31%	(6h)
	CVAE-Opt-DE [13]	6.14$^{\#}$	-	21m$^{\#}$	10.40$^{\#}$	-	41m$^{\#}$	15.75$^{\#}$	-	1.5h$^{\#}$

$^{\#}$ the obj. values, gaps or time are obtained based on 2,000 instances in their original papers, and not directly comparable to ours.
‡ the obj. values obtained by Concorde or LKH may be slightly different from ours since the 10,000 instances are randomly generated. E.g., for TSP50, the optimal values according to our running of Concorde is 5.70, while 5.69 in POMO and Wu et al.. We thus focus more on gaps.