Learning to Search Feasible and Infeasible Regions of Routing Problems with Flexible Neural k-Opt

Yining Ma[1], Zhiguang Cao[2], Yeow Meng Chee[1]

[1]National University of Singapore
[2]Singapore Management University

Background - Learning to Optimize Vehicle Routing Problems (VRPs)

Can we better learn L2S solvers for VRPs under various constraints?

NEURAL INFORMATION PROCESSING SYSTEMS

L2C
Learning-to-Construct

Pros: Fast solution construction

Cons: Prone to local optima

L2S
Learning-to-Search

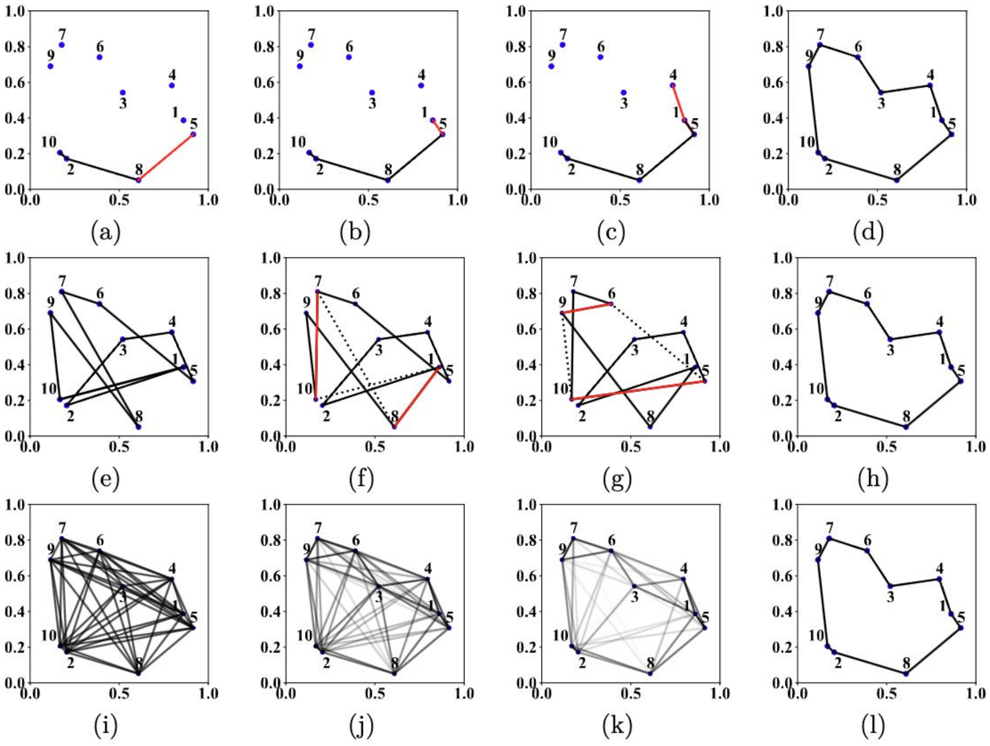
Pros: Aligns with VRP search process

Cons: Less efficient than SOTA peers

L2P
Learning-to-Predict

Pros: Efficient for large instances

Cons: Efficiency limited toTSP only



Contribution 1: Neural k-Opt (NeuOpt) - Action Factorization

The first flexible L2S solver capable of handling k-opt for any $k \geq 2$

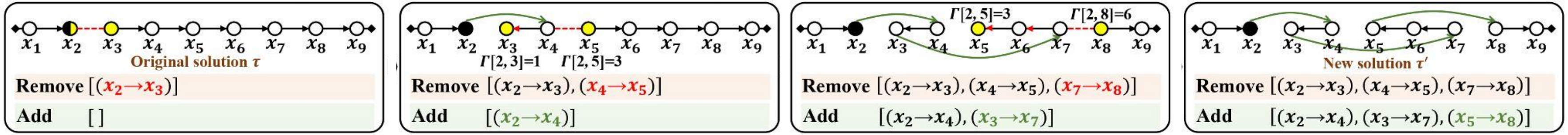
Issue in existing L2S solvers: Simplistic action space designs (fixed 2-opt or 3-opt only)!

Tailored Action Factorization (S-move, I-move, E-move)

S-move
1. Remove one edge

I-move
1. Add a new edge (starting from the first end-point)
2. Remove the corresponding conflicting edge
3. Reverse the edge directions in between

E-move
1. Fix the loop



Advantages of Tailored Action Factorization

1. Breaks down complex k-opt into manageable step-by-step constructions.
2. Adapts *k* throughout the search, balancing coarse-grained (larger k) and fine-grained (smaller k) searches

Contribution 1: Neural k-Opt (NeuOpt) - RDS Decoder

The first flexible L2S solver capable of handling k-opt for any k ≥ 2

Recurrent Dual-Stream (RDS) decoder

Designs

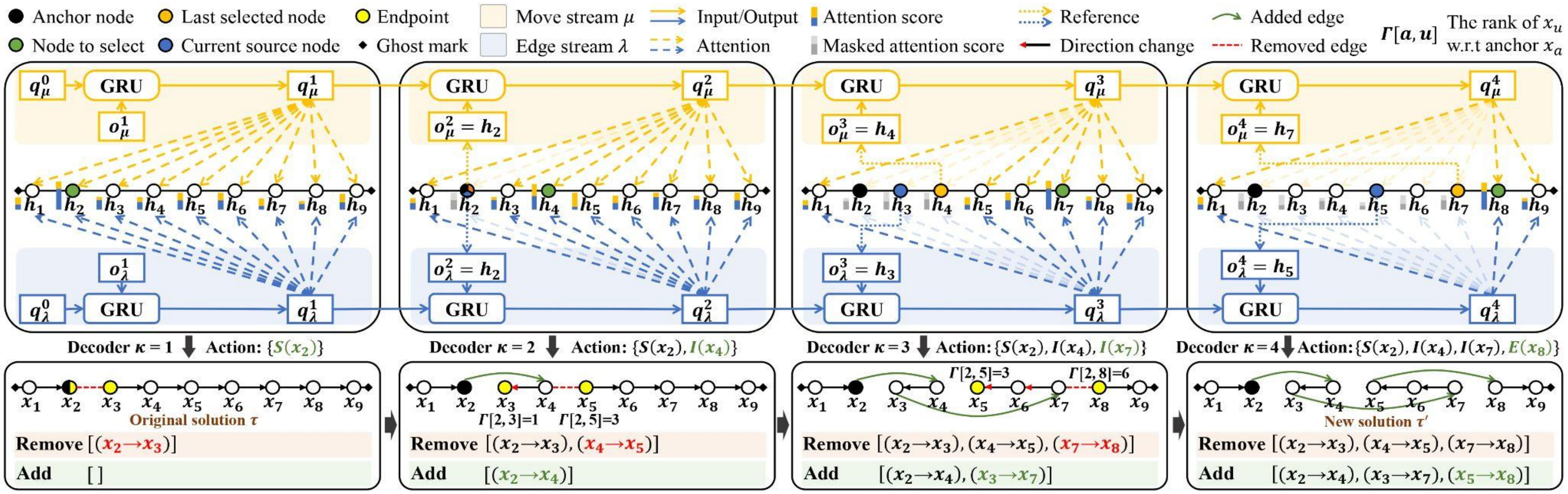
- **GRUs for Action Factorization**
- **Dual-Stream Contextual Modeling**
  - Move stream μ - past decisions
  - Edge stream λ - edge proposals

Advantage

Flexible: One united decoder for decoding k-opt exchanges with any k ≥ 2

Ablation of GRUs, μ, λ

Methods	TSP-100		CVRP-20	
	Size(M)	Obj.↓	Size(M)	Obj.↓
w/o-GRUs	0.468	7.804	0.470	6.165
w/o-μ	0.617	7.806	0.620	6.165
w/o-λ	0.617	7.799	0.620	6.164
Ours	0.683	7.798	0.685	6.163

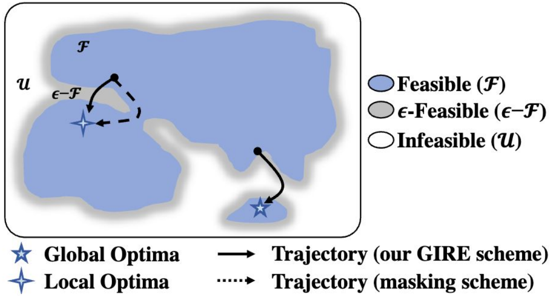


Contribution 2: Guided Infeasible Region Exploration (GIRE)

The first constraint handling scheme that explores both feasible and infeasible regions

NEURAL INFORMATION PROCESSING SYSTEMS

A search example of GIRE



Motivations and benefits of our GIRE

- Avoids non-trivial calculations of ground-truth action masks
- Fosters searches at the more promising feasibility boundaries
- Bridges (possibly isolated) feasible regions, helping escape local optima and discover shortcuts to better solutions
- Forces explicit awareness of constraints and VRP landscape

Feature Supplement

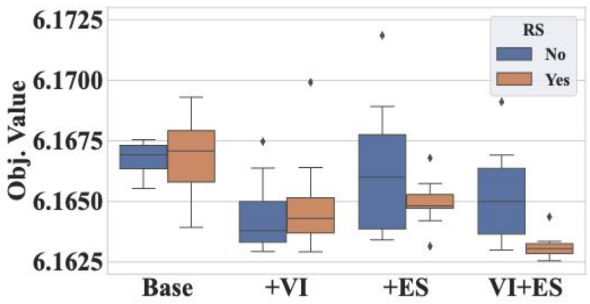
- **FI features via node embedding** (in encoder)
  - Indicate feasibility within the current solution
- **ES features via hyper-networks** (in decoder)
  - Provide historical exploration behaviour statistics

Reward Shaping

$$r_t^{\text{GIRE}} = r_t + \alpha \cdot r_t^{\text{reg}} + \beta \cdot r_t^{\text{bonus}}$$

- **Regulation:** imposes penalties when exploration is only focused on one region (extreme exploration behaviour)
- **Bonus:** encourages the search at the ε-feasible regions (boundaries of feasible and infeasible regions)

Effects of GIRE



Contribution 3: Dynamic Data Augmentation (D2A)

Help to enhance the search diversity and escape local optima during inference

NEURAL INFORMATION PROCESSING SYSTEMS

Pseudocode of D2A Algorithm

Algorithm 3 Dynamic Data Augmentation (D2A)

Input: Instance $\mathcal{G}$, policy network π_θ, inference step T, number of augments $D2A$, maximum number of consecutive steps allowed before considering the search trapped in local optima T_{D2A}

Output: Best solution found during solving all the augmented instances $\mathcal{G}_i$

for $i = 1, \cdots, D2A$ **do**
  Get an augmented instance: $\mathcal{G}_i \leftarrow$ **Augmentation**$(\mathcal{G})$;
  Get a random solution $\tau_{i,0}$ and set it as the best-so-far solution for $\mathcal{G}_i$: $\tau_i^{\text{bsf}} \leftarrow \tau_{i,0}$;
  Set counter: $T_i^{\text{stall}} \leftarrow 0$;
end for
for $t = 1, \cdots, T$ **do**
  for $i = 1, \cdots, D2A$ **do**
    Run one inference step to get a new solution $\tau_{i,t}$ for $\mathcal{G}_i$ using policy network π_θ;
    if new solution $\tau_{i,t}$ is a new best-so-far solution for instance $\mathcal{G}_i$ **then**
      Update the best-so-far solution: $\tau_i^{\text{bsf}} \leftarrow \tau_{i,t}$;
      Reset counter: $T_i^{\text{stall}} \leftarrow 0$;
    else
      Increment counter: $T_i^{\text{stall}} = T_i^{\text{stall}} + 1$;
    end if
    if $T_i^{\text{stall}} \geq T_{\text{D2A}}$ **then**
      Get a new augmented instance: $\mathcal{G}_i \leftarrow$ **Augmentation**$(\mathcal{G})$;
      Reset counter $T_i^{\text{stall}} \leftarrow 0$
    end if
  end for
end for

Effects of D2A

Inference Type	TSP-100 Gap↓	CVRP-100 Gap↓
w/o-D2A (T=5k)	0.09%	1.00%
w-D2A (T=5k)	0.05%	0.87%
w/o-D2A (T=10k)	0.04%	0.71%
w-D2A (T=10k)	0.02%	0.60%

Main Results (on TSP)

We achieve SOTA performance on TSP benchmark

	Method	Model Type	Post (Per-Ins.) Proc.	$N=20$ Obj.↓	$N=20$ Gap↓	$N=20$ Time↓	$N=50$ Obj.↓	$N=50$ Gap↓	$N=50$ Time↓	$N=100$ Obj.↓	$N=100$ Gap↓	$N=100$ Time↓
TSP	Concorde [54]	Exact	-	3.827	-	2m	5.696	-	9m	7.765	-	43m
	LKH-2 [51]	H	-	3.827	0.00%	6m	5.696	0.00%	1.3h	7.765	0.00%	5.7h
	GCN+BS [14]$^{\#}$	L2P/SL	BS+H	3.827	0.00%	15m	5.698	0.04%	23m	7.869	1.35%	46m
	Att-GCN+MCTS [6]$^{\ddagger,\#}$	L2P/SL	MCTS	(≈3.830)	(≈0.00%)	≈2m	(≈5.691)	(≈0.01%)	≈8m	(≈7.764)	(≈0.04%)	≈15m
	GNN+GLS [40] (relocate+2-opt)‡	L2P/SL	GLS	-	≈0.00%	≈2.8h	-	≈0.00%	≈2.8h	-	≈0.58%	≈2.8h
	CVAE-Opt-DE [43]‡	L2P/UL	DE	-	≈0.00%	≈1.2d	-	≈0.02%	≈2.5d	-	≈0.34%	≈1.8d
	DPDP [42] (100k)	L2P/SL	DP		-			-		7.765	0.00%	1.9h
	DIMES [7] (T=10)$^{\ddagger,\#}$	L2P/RL	AS+M+M		-			-		(≈7.762)	(≈0.01%)	-
	DIFUSCO [15] (T=50, S=16)$^{\#}$	L2P/SL	2-opt		-		5.696	0.01%	5.8h	7.766	0.02%	21.7h
	AM+LCP* [33] ({1280, 45})	L2C/RL	-	3.828	0.01%	2.1h	5.699	0.05%	4.9h	7.811	0.60%	10.9h
	Pointerformer [32] (A=8, T=200)	L2C/RL	-	3.827	0.00%	13m	5.697	0.02%	1.1h	7.773	0.11%	5.6h
	Sym-NCO [13] (A=8, T=200)	L2C/RL	-		-			-		7.771	0.08%	5.6h
	POMO [4] (A=8, T=200)	L2C/RL	-	3.827	0.00%	13m	5.696	0.00%	1.1h	7.770	0.07%	5.6h
	POMO+EAS [5] (A=8, T=200)	L2C/RL	AS	3.827	0.00%	24m	5.696	0.00%	2h	7.769	0.05%	10.9h
	POMO+EAS+SGBS [34] (short)	L2C/RL	AS+BS		-			-		7.767	0.04%	6.5h
	POMO+EAS+SGBS [34] (long)	L2C/RL	AS+BS		-			-		7.767	0.03%	1.1d
	Costa et al. [16] (2-opt, T=2k)	L2S/RL	-	3.827	0.00%	31m	5.703	0.12%	40m	7.824	0.77%	1.1h
	Sui et al. [17] (3-opt, T=2k)‡	L2S/RL	-	≈3.84	≈0.00%	≈32m	≈5.70	≈0.08%	≈48m	≈7.82	≈0.74%	≈1.3h
	Wu et al. [39] (2-opt, T=5k)	L2S/RL	-		-		5.709	0.23%	1.3h	7.884	1.54%	2h
	DACT [9] (2-opt, A=4, T=10k)	L2S/RL	-	3.827	0.00%	1.5h	5.696	0.00%	4.1h	7.772	0.10%	13.5h
	NeuOpt (D2A=1, T=1k)	L2S/RL	-	3.827	0.00%	2m	5.697	0.02%	6m	7.790	0.33%	17m
	NeuOpt (D2A=1, T=5k)	L2S/RL	-	3.827	0.00%	12m	5.696	0.00%	32m	7.768	0.05%	1.4h
	NeuOpt (D2A=1, T=10k)	L2S/RL	-	3.827	0.00%	23m	5.696	0.00%	1.1h	7.766	0.02%	2.8h
	NeuOpt (D2A=5, T=1k)	L2S/RL	-	3.827	0.00%	12m	5.696	0.00%	32m	7.767	0.04%	1.4h
	NeuOpt (D2A=5, T=3k)	L2S/RL	-	3.827	0.00%	35m	5.696	0.00%	1.6h	7.765	0.01%	4.2h
	NeuOpt (D2A=5, T=5k)	L2S/RL	-	3.827	0.00%	1h	5.696	0.00%	2.7h	7.765	0.00%	7h

Main Results (on CVRP)

We achieve SOTA performance on CVRP benchmark

NEURAL INFORMATION PROCESSING SYSTEMS

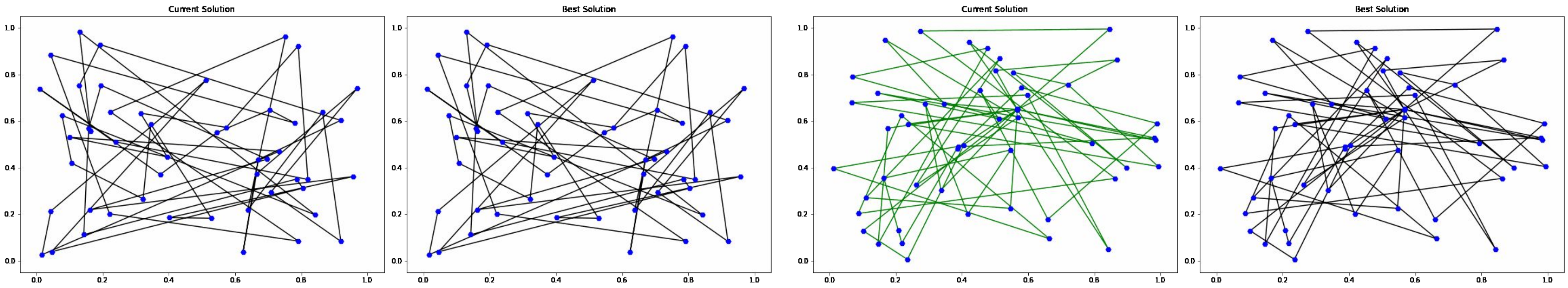
	Method	Model Type	Post (Per-Ins.) Proc.	$N=20$ Obj.↓	$N=20$ Gap↓	$N=20$ Time↓	$N=50$ Obj.↓	$N=50$ Gap↓	$N=50$ Time↓	$N=100$ Obj.↓	$N=100$ Gap↓	$N=100$ Time↓
CVRP	HGS [21]	H	-	6.130	-	10.7h	10.366	-	1.2d	15.563	-	2.5d
	LKH-3 [20]	H	-	6.135	0.08%	17.9h	10.375	0.09%	2.8d	15.647	0.54%	5.7d
	CVAE-Opt-DE [43]‡	L2P/UL	DE	≈6.14	-	≈2.4d	≈10.40	-	≈4.7d	≈15.75	-	≈11d
	DPDP [42] (1000k)	L2P/SL	DP		-			-		15.627	0.41%	1.2d
	AM+LCP [33] ({2560, 1})‡	L2C/RL	-	≈6.15	≈0.33%	≈23m	≈10.52	≈1.48%	≈52m	≈16.00	≈2.81%	≈2.1h
	Sym-NCO [13] (A=8, T=200)	L2C/RL	-		-			-		15.702	0.89%	7.2h
	POMO [4] (A=8, T=200)	L2C/RL	-	6.136	0.09%	11m	10.397	0.30%	1.4h	15.672	0.70%	7.2h
	POMO+EAS [5] (A=8, T=200)	L2C/RL	AS	6.132	0.04%	38m	10.379	0.13%	3.1h	15.610	0.30%	16h
	POMO+EAS+SGBS [34] (short)	L2C/RL	AS+BS		-			-		15.587	0.15%	1d
	POMO+EAS+SGBS [34] (long)	L2C/RL	AS+BS		-			-		15.579	0.10%	4.1d
	NLNS [8] (Ruin-Repair, T=5k)	L2S/RL	-	6.175	0.73%	48m	10.506	1.35%	1.4h	15.915	2.26%	2.4h
	NCE [37] (CROSS exchange)‡	L2S/SL	-	≈6.13	≈0.00%	≈11h	≈10.41	≈0.42%	≈2.3d	≈15.81	≈1.59%	≈10.4d
	Wu et al. [39] (2-opt, T=5k)	L2S/RL	-		-		10.544	1.72%	4.2h	16.165	3.87%	5h
	DACT [9] (2-opt, A=6, T=10k)	L2S/RL	-	6.130	0.01%	4h	10.383	0.16%	16h	15.736	1.11%	1.7d
	NeuOpt-GIRE (D2A=1, T=1k)	L2S/RL	-	6.132	0.03%	4m	10.430	0.61%	12m	15.865	1.94%	28m
	NeuOpt-GIRE (D2A=1, T=5k)	L2S/RL	-	6.130	0.00%	20m	10.382	0.16%	59m	15.698	0.87%	2.3h
	NeuOpt-GIRE (D2A=1, T=10k)	L2S/RL	-	6.130	0.00%	41m	10.375	0.08%	2h	15.656	0.60%	4.6h
	NeuOpt-GIRE (D2A=5, T=6k)	L2S/RL	-	6.130	0.00%	2.1h	10.369	0.03%	5.9h	15.610	0.30%	13.8h
	NeuOpt-GIRE (D2A=5, T=20k)	L2S/RL	-	6.130	0.00%	6.8h	10.367	0.01%	19.7h	15.586	0.15%	1.9d
	NeuOpt-GIRE (D2A=5, T=40k)	L2S/RL	-	6.130	0.00%	13.7h	10.367	0.01%	1.6d	15.579	0.10%	3.8d

Demonstrations & GitHub Links

Thank you for listening and welcome to explore our GitHub!

GIF 1: NeuOpt Search for TSP

GIF2: NeuOpt-GIRE for CVRP



GitHub link: https://github.com/yining043/NeuOpt