

Table 1: Results on TSPLIB [Reinelt, 1991] instances. Models are only trained on $n = 100$.

Instance	Opt.	POMO		POMO_AT		POMO_HAC		POMO_DivTrain		CNF		CNF+EAS	
		Obj.	Gap	Obj.	Gap	Obj.	Gap	Obj.	Gap	Obj.	Gap	Obj.	Gap
kroA100	21282	21420	0.65%	21347	0.31%	21308	0.12%	21370	0.41%	21308	0.12%	21282	0.00%
kroB100	22141	22200	0.27%	22211	0.32%	22200	0.27%	22199	0.26%	22216	0.34%	22199	0.26%
kroC100	20749	20799	0.24%	20768	0.09%	20753	0.02%	20768	0.09%	20758	0.04%	20749	0.00%
kroD100	21294	21446	0.71%	21391	0.46%	21407	0.53%	21435	0.66%	21353	0.28%	21294	0.00%
kroE100	22068	22259	0.87%	22288	1.00%	22167	0.45%	22213	0.66%	22121	0.24%	22106	0.17%
eil101	629	630	0.16%	630	0.16%	629	0.00%	631	0.32%	630	0.16%	629	0.00%
lin105	14379	14477	0.68%	14426	0.33%	14408	0.20%	14402	0.16%	14403	0.17%	14379	0.00%
pr107	44303	44678	0.85%	47819	7.94%	44596	0.66%	46285	4.47%	44719	0.94%	44303	0.00%
pr124	59030	59389	0.61%	59257	0.38%	59385	0.60%	59558	0.89%	59076	0.08%	59030	0.00%
bier127	118282	133042	12.48%	118606	0.27%	118608	0.28%	118337	0.05%	118841	0.47%	118282	0.00%
ch130	6110	6119	0.15%	6130	0.33%	6115	0.08%	6125	0.25%	6111	0.02%	6110	0.00%
pr136	96772	97983	1.25%	100225	3.57%	97617	0.87%	100145	3.49%	97567	0.82%	97198	0.44%
pr144	58537	58935	0.68%	59544	1.72%	58913	0.64%	59265	1.24%	58868	0.57%	58537	0.00%
ch150	6528	6554	0.40%	6582	0.83%	6556	0.43%	6578	0.77%	6550	0.34%	6554	0.40%
kroA150	26524	26755	0.87%	26898	1.41%	26736	0.80%	26813	1.09%	26722	0.75%	26524	0.00%
kroB150	26130	26405	1.05%	26506	1.44%	26379	0.95%	26467	1.29%	26494	1.39%	26143	0.05%
pr152	73682	74249	0.77%	77537	5.23%	75291	2.18%	77127	4.68%	74876	1.62%	73682	0.00%
rat195	2323	2486	7.02%	2500	7.62%	2461	5.94%	2467	6.20%	2449	5.42%	2338	0.65%
kroA200	29368	29992	2.12%	30222	2.91%	29771	1.37%	30143	2.64%	29755	1.32%	29435	0.23%
kroB200	29437	30298	2.92%	30157	2.45%	29890	1.54%	30267	2.82%	29862	1.44%	29508	0.24%

Table 2: Robustness study of DIMES [Qiu et al., 2022] on 1000 TSP100 instances. G, S, MCTS, and AS denote greedy, sample, Monte Carlo tree search, and active search, respectively.

	DIMES (G)		DIMES (AS+G)		DIMES (S)		DIMES (AS+S)		DIMES (MCTS)		DIMES (AS+MCTS)	
	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
Clean	14.65%	1.31m	5.21%	2.42h	13.50%	1.40m	5.11%	2.44h	0.05%	2.50m	0.03%	2.47h
Fixed Adv.	19.29%	1.32m	12.12%	2.45h	18.24%	1.42m	11.87%	2.45h	0.19%	2.50m	0.16%	2.49h

Table 3: Robustness study of DeepACO [Ye et al., 2023] on 1000 TSP100 instances. Note that only DeepACO (NLS) uses local search. We use 100 ants and $T \in \{1, 10, 50, 100\}$ ACO iterations. The total inference time of each method with $T = 100$ is reported.

	ACO - 1.7h				DeepACO - 1.7h				DeepACO (NLS) - 5.6h			
	T=1	T=10	T=50	T=100	T=1	T=10	T=50	T=100	T=1	T=10	T=50	T=100
Clean	98.65%	46.15%	25.47%	20.31%	13.65%	7.21%	5.54%	5.08%	0.46%	0.08%	0.03%	0.02%
Fixed Adv.	40.87%	19.09%	12.61%	10.92%	25.50%	14.60%	10.75%	9.71%	0.31%	0.08%	0.00%	0.00%

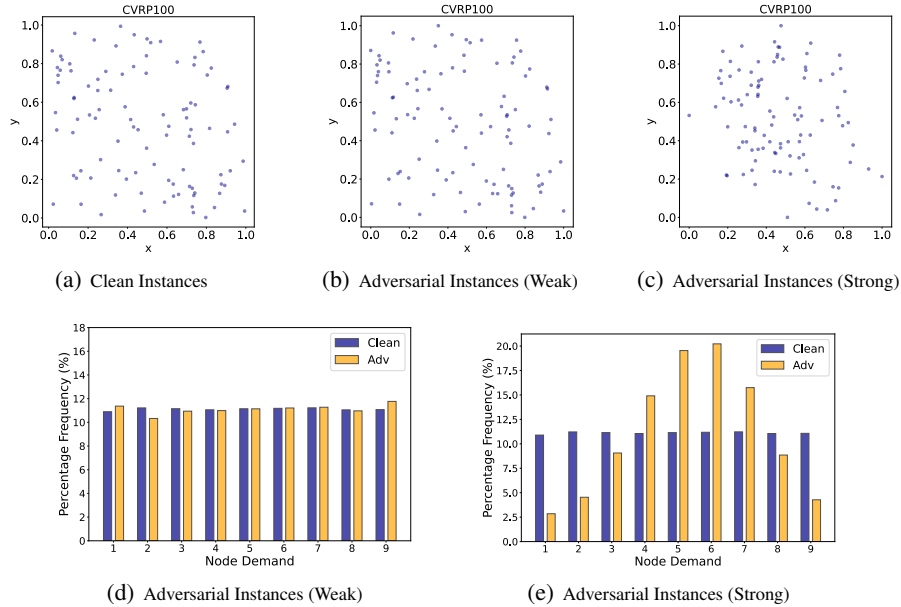


Figure 1: Visualization of clean instances and corresponding adversarial instances generated by weak and strong attack. (a-c) shows the spatial distribution of node locations, and (d-e) shows the percentage frequency distribution of node demands over the entire CVRP100 test dataset.