Summary of Best Performing Parameters

This table highlights the configurations that yielded the best average evaluation tour length for each problem size (N). The "Avg. Eval Tour Length" in **bold** indicates the overall best performing model for that specific N value when comparing baseline and alternative loss types.

N	Loss Type	Best Run #	Key Lambda/C1_p Values (length, degree, subtour or C1_p)	Epochs	Initial LR	LR Gamma	Grad Clip	Avg. Eval Tour Length	Final Training Loss (approx.)
20	baseline	1	C1_p=15.0	50	0.005	0.1	N/A	**4.1690**	30.0
20	alternative	2	1.0, 50.0, 20.0	50	0.005	0.1	N/A	4.1812	25.69
50	baseline	15	C1_p=20.0	50	0.0005	0.3	Yes (1.0)	6.3447	~22.0
50	alternative	5	1.0, 50.0, 40.0	50	0.0005	0.1	Yes (1.0)	**6.3332**	55.10
100	baseline	16	C1_p=20.0	120	0.0005	0.3	Yes (1.0)	8.7508	~35.2
100	alternative	12	1.0, 50.0, 40.0	120	0.0005	0.3	Yes (1.0)	**8.7169**	86.48
200	baseline	17	C1_p=20.0	150	0.0005	0.3	Yes (1.0)	**12.0735**	Plateaued
200	alternative	13	1.0, 50.0, 40.0	150	0.0005	0.3	Yes (1.0)	12.1566	201.67

500	baseline	18	C1_p=20.0	250	0.0003	0.3	Yes (1.0)	**12.5700**	~320
500	alternative	14	1.0, 50.0, 45.0	250	0.0003	0.3	Yes (1.0)	18.6	350.22