

Table 4: Model Efficiency and Performance Comparison for Different Datasets with $T = 96$. Parameters (Params) are measured in millions (M), GPU memory (GPU) in MiB, computation time per epoch in seconds (s) on NVIDIA V100 GPU with batch size 32.

(a) Traffic Dataset

Model	Params (M)	GPU(MiB)	Time (s)	MSE
Autoformer	14.914	18.811	51.0	0.668
iTransformer	6.405	62.710	126.0	0.388
PatchTST	3.755	22.132	1042.0	0.413
MICN	236.151	32.310	84.0	0.511
TimesNet	30.170	111.998	6563.0	0.611
DLinear	0.009	12.861	7.7	0.485
Koopa	5.429	50.335	25.5	0.401
SKOLR	1.479	5.915	216.0	0.368

(b) Electricity Dataset

Model	Params (M)	GPU(MiB)	Time (s)	MSE
Autoformer	11.214	17.373	68.7	0.182
iTransformer	4.957	86.478	58.6	0.134
PatchTST	6.904	73.517	1231.0	0.143
MICN	6.635	32.668	18.0	0.165
TimesNet	15.037	33.435	11351.0	0.170
DLinear	0.019	76.016	6.8	0.153
Koopa	4.076	31.067	33.1	0.136
SKOLR	1.541	6.163	99.1	0.132

(c) ETTh1 Dataset

Model	Params (M)	GPU(MiB)	Time (s)	MSE
Autoformer	10.536	16.523	29.5	0.634
iTransformer	0.237	27.245	4.1	0.393
PatchTST	3.752	22.018	8.5	0.372
MICN	252.001	65.974	21.1	0.406
TimesNet	0.605	26.053	22.1	0.411
DLinear	0.140	26.440	0.6	0.379
Koopa	0.135	31.951	10.1	0.371
SKOLR	0.429	1.717	2.8	0.371

(d) ETTm2 Dataset

Model	Params (M)	GPU(MiB)	Time (s)	MSE
Autoformer	10.536	14.599	152.6	0.241
iTransformer	0.237	27.245	13.1	0.177
PatchTST	10.056	39.910	980.0	0.171
MICN	252.001	65.974	84.2	0.197
TimesNet	1.192	34.783	113.0	0.187
DLinear	18.291	9.312	1.9	0.172
Koopa	0.135	31.951	48.2	0.171
SKOLR	0.429	1.717	12.6	0.171