



OccamVTS: Distilling Vision Models to 1% Parameters for Time Series Forecasting

Sisuo Lyu¹, Siru Zhong¹, Weilin Ruan¹, Qingxiang Liu¹, Qingsong Wen², Hui Xiong¹, Yuxuan Liang¹

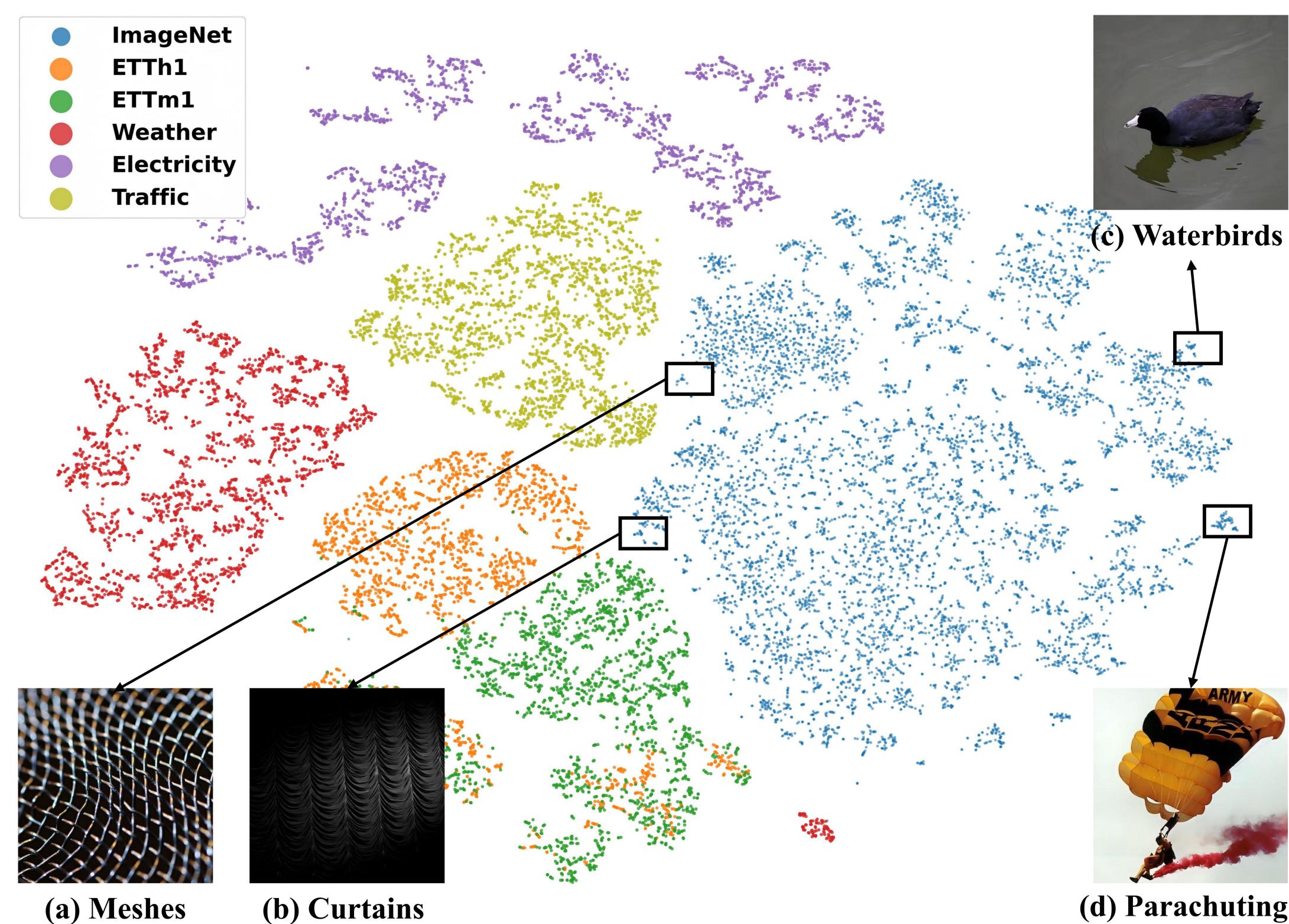
¹The Hong Kong University of Science and Technology (Guangzhou) ²Squirrel Ai Learning



Motivation

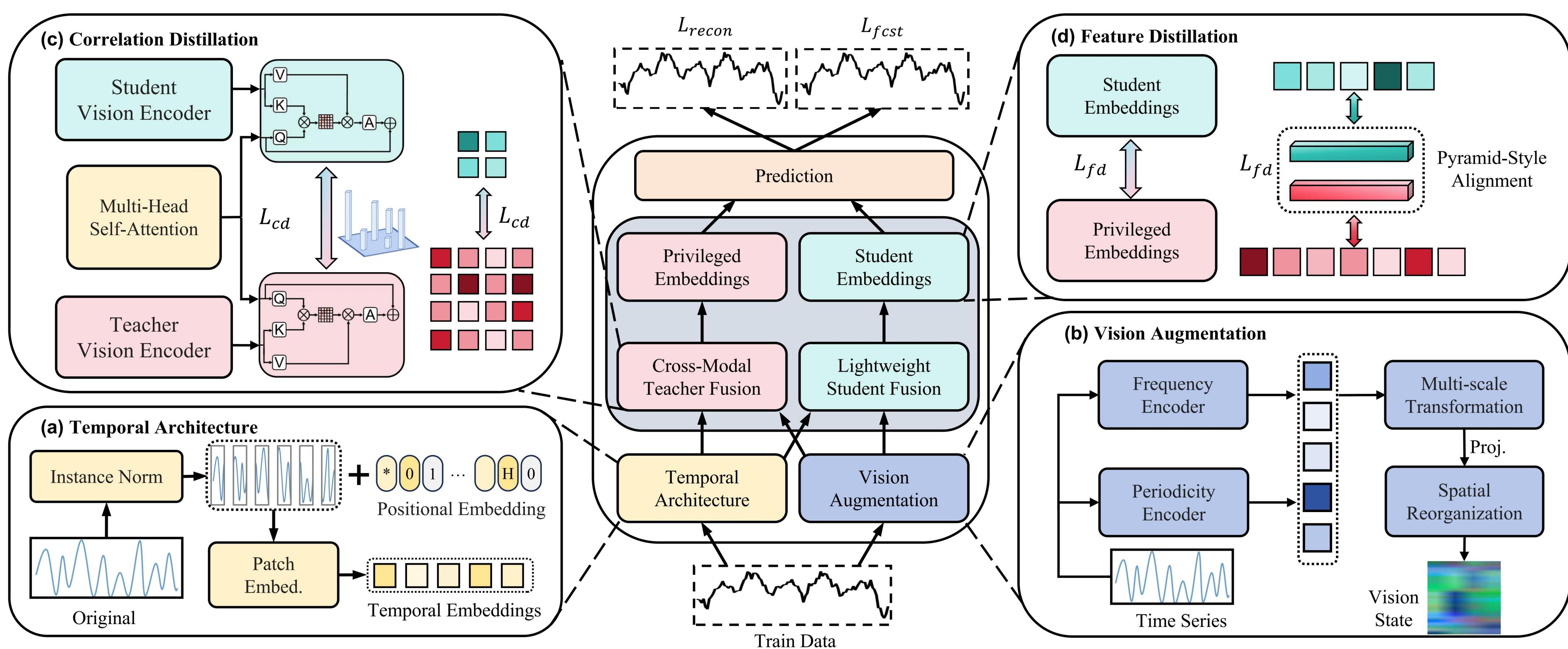
➤ Cross-Modal Analysis: Time-Series Dynamics Align with Low-Level

Visual Textures



Architecture

➤ The First Vision-Distillation Framework for Time Series Forecasting



Experiments

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	FEDformer	Autoformer
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETT1	0.403	0.421	0.416	0.433	0.434	0.444	0.405	0.420	0.419	0.432	0.447	0.443	0.454
ETT2	0.336	0.383	0.338	0.342	0.341	0.391	0.339	0.380	0.365	0.395	0.387	0.427	0.414
ETTm1	0.347	0.373	0.354	0.377	0.347	0.377	0.369	0.378	0.381	0.396	0.352	0.387	0.400
ETTm2	0.245	0.307	0.252	0.313	0.258	0.317	0.248	0.311	0.269	0.320	0.275	0.323	0.333
Weather	0.224	0.259	0.229	0.268	0.230	0.269	0.224	0.263	0.226	0.262	0.240	0.272	0.229
ECL	0.162	0.259	0.168	0.267	0.170	0.270	0.172	0.272	0.165	0.253	0.182	0.273	0.199
Traffic	0.407	0.279	0.415	0.292	0.419	0.297	0.419	0.298	0.416	0.264	0.485	0.298	0.550

Table 1: Long-term forecasting results. Results are averaged over forecasting horizons $H \in \{96, 192, 336, 720\}$. Lower values indicate better performance. Red: best, Blue: second best. Full results see Appendix D.

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	FEDformer	Autoformer
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETT1	0.422	0.439	0.443	0.456	0.446	0.431	0.442	0.517	0.512	0.613	0.452	0.471	0.468
ETT2	0.344	0.390	0.356	0.402	0.357	0.402	0.379	0.391	0.402	0.433	0.452	0.460	0.479
ETTm1	0.356	0.379	0.364	0.387	0.365	0.387	0.398	0.431	0.487	0.461	0.371	0.393	0.401
ETTm2	0.253	0.313	0.261	0.321	0.262	0.322	0.323	0.291	0.351	0.311	0.367	0.336	0.373
Weather	0.227	0.262	0.230	0.268	0.231	0.269	0.233	0.274	0.241	0.271	0.242	0.281	0.276
ECL	0.181	0.283	0.206	0.310	0.209	0.312	0.188	0.291	0.168	0.271	0.187	0.277	0.172
Traffic	0.460	0.332	0.531	0.385	0.536	0.390	0.484	0.357	0.483	0.315	0.536	0.349	0.621

Table 2: Few-shot learning on 10% training data. We use the same protocol in Table 1. Full results see Appendix E.

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	FEDformer	Autoformer
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETT1 → ETT2	0.342	0.385	0.351	0.396	0.350	0.395	0.338	0.385	0.367	0.391	0.427	0.424	0.458
ETT1 → ETTm2	0.295	0.350	0.300	0.355	0.303	0.356	0.293	0.350	0.301	0.357	0.369	0.400	0.327
ETT2 → ETT1	0.429	0.446	0.452	0.466	0.532	0.508	0.496	0.480	0.511	0.498	0.679	0.577	0.865
ETT2 → ETTm2	0.285	0.343	0.295	0.352	0.297	0.353	0.329	0.370	0.342	0.378	0.444	0.432	0.474
ETTm1 → ETT2	0.357	0.398	0.359	0.400	0.359	0.399	0.354	0.397	0.417	0.422	0.452	0.441	0.454
ETTm1 → ETTm2	0.259	0.315	0.262	0.319	0.263	0.319	0.291	0.331	0.329	0.357	0.354	0.367	0.484
ETTm2 → ETT2	0.357	0.394	0.366	0.402	0.364	0.403	0.359	0.399	0.432	0.443	0.413	0.427	0.494
ETTm2 → ETTm1	0.403	0.410	0.451	0.442	0.437	0.434	0.432	0.426	0.427	0.448	0.554	0.478	0.588

Table 3: Zero-shot learning results. We use the same protocol in Table 1. Full results see Appendix F.

Methods	Ours	Only Teacher	Only Student	TimeVLM	Timemixer++	Timemixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	ETSformer	LightTS	FEDformer	Stationary	Autoformer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETT1 → ETT2	0.342	0.385	0.351	0.396	0.350	0.395	0.338	0.385	0.367	0.391	0.427	0.424	0.458	0.452	0.421	0.431	
ETT1 → ETTm2	0.295	0.350	0.300	0.355	0.303	0.356	0.293	0.350	0.301	0.357	0.369	0.400	0.327	0.361	0.337	0.374	
ETT2 → ETT1	0.429	0.446	0.452	0.466	0.532	0.508	0.496	0.480	0.511	0.498	0.679	0.577	0.865	0.621	0.657	0.563	0.703
ETT2 → ETTm2	0.285	0.343	0.295	0.352	0.297	0.353	0.329	0.370									