

## A DATASET DESCRIPTION

### A.1 DATASET FOR SINGLE-TASK MODEL: iTRANSFORMER

For TS forecasting in a single-task setting, we evaluate the effectiveness of our proposed method using 13 datasets, with their statistics described in Table A.1. We adhere to the same data processing and train-validation-test split protocol as iTransformer (Liu et al., 2024a), ensuring that the training, validation, and test sets are separated in chronological order. The input length is consistently set to 96 across all datasets. Note that  $N$  and  $C$  denote the size of the dataset and number of channels in a dataset, respectively.

Dataset	$C$	Prediction Length	$(N_{\text{train}}, N_{\text{val}}, N_{\text{test}})$
ETTh1 (Zhou et al., 2021)	7	{96, 192, 336, 720}	(8545, 2881, 2881)
ETTh2 (Zhou et al., 2021)	7	{96, 192, 336, 720}	(8545, 2881, 2881)
ETTm1 (Zhou et al., 2021)	7	{96, 192, 336, 720}	(34465, 11521, 11521)
ETTm2 (Zhou et al., 2021)	7	{96, 192, 336, 720}	(34465, 11521, 11521)
Exchange (Wu et al., 2021)	8	{96, 192, 336, 720}	(5120, 665, 1422)
Weather (Wu et al., 2021)	21	{96, 192, 336, 720}	(36792, 5271, 10540)
ECL (Wu et al., 2021)	321	{96, 192, 336, 720}	(18317, 2633, 5261)
Traffic (Wu et al., 2021)	862	{96, 192, 336, 720}	(12185, 1757, 3509)
Solar-Energy (Lai et al., 2018)	137	{96, 192, 336, 720}	(36601, 5161, 10417)
PEMS03 (Liu et al., 2022)	358	{12, 24, 48, 96}	(15617, 5135, 5135)
PEMS04 (Liu et al., 2022)	307	{12, 24, 48, 96}	(10172, 3375, 3375)
PEMS07 (Liu et al., 2022)	883	{12, 24, 48, 96}	(16911, 5622, 5622)
PEMS08 (Liu et al., 2022)	170	{12, 24, 48, 96}	(10690, 3548, 3548)

Table A.1: Single-task forecasting datasets.

## A.2 DATASET FOR MULTI-TASK MODEL: UNITS

The datasets used in the experiment are aggregated from the Monash Forecasting Repository (Godahewa et al., 2021), the Time Series Classification Website (Middlehurst et al., 2024), and the Time Series Library (Wu et al., 2023). The combined training set includes more than 35 million time steps and over 6,000 variables (channels). Note that  $N$ ,  $L$ ,  $C$  denote the training size, input length, and number of channels in a dataset, respectively.

### A.2.1 MULTI-TASK LEARNING

For TS forecasting and classification in a multi-task setting, we evaluate the effectiveness of our proposed method using 20 datasets for forecasting and 18 datasets for classification. The statistics of these datasets are summarized in Table A.2 and A.3, respectively.

Category	Dataset	Prediction Length	$N$	$L$	$C$
Finance	NN5 (Taieb et al., 2012)	112	409	112	111
	Exchange (Wu et al., 2021)	192 336	5024 4880	96	8
Electricity	ECL (Wu et al., 2021)	96	18221	96	321
		192	18125		
		336	17981		
		720	17597		
	ETTh1 (Zhou et al., 2021)	96	8449	96	7
		192	8353		
		336	8209		
		720	7825		
Illness	ILI (Wu et al., 2021)	60	581	36	7
Traffic	Traffic (Wu et al., 2021)	96	12089	96	862
		192	11993		
		336	11849		
		720	11465		
Weather	Weather (Wu et al., 2021)	96	36696	96	21
		192	36600		
		336	36456		
		720	36072		

Table A.2: Multi-task forecasting datasets.

Category	Dataset	# classes	$N$	$L$	$C$
Finance	SharePriceIncrease (Dau et al., 2019)	2	965	60	1
Audio	JapaneseVowels (Bagnall et al., 2018)	9	270	29	12
	SpokenArabicDigits (Bagnall et al., 2018)	10	6599	93	13
	Heartbeat (Bagnall et al., 2018)	2	204	405	61
ECG	ECG5000 (Dau et al., 2019)	5	500	140	1
	NonInvasiveFetalECGThorax1 (Dau et al., 2019)	52	1800	750	1
EEG	Blink (Bagnall et al., 2018)	2	500	510	4
	FaceDetection (Bagnall et al., 2018)	2	5890	62	144
	SelfRegulationSCP2 (Bagnall et al., 2018)	2	200	1152	7
Sensors	ElectricDevices (Dau et al., 2019)	7	8926	96	1
	Trace (Dau et al., 2019)	4	100	275	1
	FordB (Dau et al., 2019)	2	3636	500	1
Human Activity	MotionSenseHAR (Bagnall et al., 2018)	6	966	200	12
	EMOPain (Bagnall et al., 2018)	3	968	180	30
	UWaveGestureLibrary (Bagnall et al., 2018)	8	120	315	3
Traffic	Chinatown (Dau et al., 2019)	2	20	24	1
	MelbournePedestrian (Dau et al., 2019)	10	1194	24	1
	PEMS-SF (Bagnall et al., 2018)	7	267	144	963

Table A.3: Multi-task classification datasets.

## A.2.2 FEW-SHOT LEARNING

For TS forecasting, classification, imputation, and anomaly detection in a few-shot setting, we evaluate the effectiveness of our proposed method using nine datasets for forecasting, six datasets for classification, four datasets for imputation, and five datasets for anomaly detection. The statistics of these datasets related to forecasting and classification are summarized in Table A.4, Table A.5, A.6, and A.7, respectively.

Category	Dataset	Prediction Length	$N$	$L$	$C$
Electricity	ETTh2 (Zhou et al., 2021)	96	8449	96	7
		192	8353		
		336	8209		
		720	7825		
	ETTm1 (Zhou et al., 2021)	96	34369	96	7
		192	34273		
		336	34129		
		720	33745		
Weather	SaugeenRiverFlow (McLeod & Gweon, 2013)	24	18921	48	1

Table A.4: Few-shot forecasting datasets.

Category	Dataset	# classes	$N$	$L$	$C$
ECG	ECG200 (Dau et al., 2019)	2	100	96	1
EEG	SelfRegulationSCP1 (Bagnall et al., 2018)	2	268	896	6
Human Activity	RacketSports (Bagnall et al., 2018)	4	151	30	6
	Handwriting (Bagnall et al., 2018)	26	150	152	3
	Epilepsy (Bagnall et al., 2018)	4	137	207	3
Sensor	StarLightCurves (Dau et al., 2019)	3	1000	1024	1

Table A.5: Few-shot classification datasets.

Category	Dataset	$L$	$C$
Electricity	ETTh1 (Zhou et al., 2021)	96	7
	ETTh1 (Zhou et al., 2021)	96	7
	ECL (Wu et al., 2021)	96	321
Weather	Weather (Wu et al., 2021)	96	21

Table A.6: Few-shot imputation datasets.

Category	Dataset	$L$	$C$
Machine	SMD (Su et al., 2019)	96	38
	PSM (Abdulaal et al., 2021)	96	25
Spacecraft	MSL (Hundman et al., 2018)	96	55
	SMAP (Hundman et al., 2018)	96	25
Infrastructure	SWaT (Mathur & Tippenhauer, 2016)	96	51

Table A.7: Few-shot anomaly detection datasets.

### A.2.3 ZERO-SHOT LEARNING

For TS forecasting in a zero-shot setting, we evaluate the effectiveness of our proposed method using six datasets. Three of these datasets are used for the zero-shot setting with unseen datasets, while the remaining four datasets are used for the zero-shot setting with new prediction lengths. The statistics for the three unseen datasets are summarized in Table A.8.

Category	Dataset	Prediction Length	$L$	$C$
Electricity	Solar (NREL, 2006)	64	128	137
Weather	SaugeenRiverFlow (McLeod & Gweon, 2013)	128	256	1
Healthcare	Hospital (Hyndman et al., 2008)	16	32	767

Table A.8: Zero-shot forecasting datasets.

## B IMPLEMENTATION DETAILS

It is important to note that we follow the experimental settings of iTransformer for single-task and UniTS for multi-task settings, respectively. The following sections outline the specific settings we adhered to.

### B.1 IMPLEMENTATION FOR SINGLE-TASK MODEL: iTRANSFORMER

Following iTransformer (Liu et al., 2024a), we use the Adam optimizer (Kinga et al., 2015) and L2 loss for model optimization. The batch size is consistently set to 32, and the number of training epochs is fixed at 10. Since our approach is plug-and-play, we do not adjust any hyperparameters for our method; instead, we use the same hyperparameters employed by iTransformer.

### B.2 IMPLEMENTATION FOR MULTI-TASK MODEL: UniTS

**Model architecture.** In a multi-task setting, the UniTS network consists of three UniTS blocks, along with one GEN tower and one CLS tower. For each data source, specific prompt and task tokens are assigned, with forecasting tasks on the same source but with varying forecast lengths using the same prompt and GEN token. To enable zero-shot learning on new datasets, a shared prompt and GEN token are applied across all data sources. The embedding dimensions are set to 64 for the supervised version, and 32 for the prompt-tuning version, and all blocks in UniTS retain the same feature shape.

**Model training.** In multi-task settings, models are trained jointly on multiple tasks following a unified protocol. To match the largest dataset, samples from each dataset are repeated within each epoch. Supervised training is conducted over 5 epochs with gradient accumulation, yielding an effective batch size of 1024. The initial learning rate is set at  $3.2e-2$  and is adjusted using a multi-step decay schedule. For self-supervised pretraining, the models training with an are trained for 10 epochs with effective batch size of 4096, starting with a learning rate of  $6.4e-3$ , which is adjusted using a cosine decay schedule.

## C APPLICATION TO ITRANSFORMER

To demonstrate the effectiveness of our method on a model with a single-task setting, we apply it to the TS forecasting task using iTransformer (Liu et al., 2024a) on 13 datasets, with the results shown in Table C.1.

Metric		iTransformer		+ CM	
		MSE	MAE	MSE	MAE
ETTh1	96	0.387	0.405	<b>0.385</b>	<b>0.404</b>
	192	0.441	0.436	<b>0.438</b>	<b>0.434</b>
	336	0.491	0.462	<b>0.475</b>	<b>0.454</b>
	720	0.509	0.494	<b>0.477</b>	<b>0.474</b>
	Avg.	0.457	0.449	<b>0.444</b>	<b>0.441</b>
ETTh2	96	0.301	0.350	<b>0.295</b>	<b>0.347</b>
	192	0.381	0.399	<b>0.380</b>	<b>0.397</b>
	336	<b>0.423</b>	<b>0.432</b>	0.427	0.434
	720	<b>0.430</b>	0.446	0.432	<b>0.445</b>
	Avg.	0.384	0.407	<b>0.383</b>	<b>0.406</b>
ETTm1	96	0.342	0.377	<b>0.331</b>	<b>0.369</b>
	192	0.383	0.396	<b>0.372</b>	<b>0.390</b>
	336	0.418	0.418	<b>0.412</b>	<b>0.414</b>
	720	0.487	0.456	<b>0.479</b>	<b>0.453</b>
	Avg.	0.408	0.412	<b>0.398</b>	<b>0.406</b>
ETTm2	96	0.186	<b>0.272</b>	<b>0.184</b>	<b>0.272</b>
	192	0.254	0.314	<b>0.251</b>	<b>0.311</b>
	336	0.317	0.353	<b>0.312</b>	<b>0.350</b>
	720	0.416	0.409	<b>0.412</b>	<b>0.408</b>
	Avg.	0.293	0.337	<b>0.289</b>	<b>0.335</b>
Exchange	96	0.086	0.206	<b>0.085</b>	<b>0.205</b>
	192	0.181	0.303	<b>0.180</b>	<b>0.302</b>
	336	0.338	0.422	<b>0.337</b>	<b>0.421</b>
	720	0.869	0.704	<b>0.850</b>	<b>0.696</b>
	Avg.	0.368	0.409	<b>0.363</b>	<b>0.406</b>
Weather	96	0.174	0.215	<b>0.165</b>	<b>0.209</b>
	192	0.224	0.258	<b>0.213</b>	<b>0.251</b>
	336	0.281	0.298	<b>0.274</b>	<b>0.296</b>
	720	0.359	0.351	<b>0.350</b>	<b>0.346</b>
	Avg.	0.260	0.281	<b>0.250</b>	<b>0.275</b>
Solar	96	0.201	0.234	<b>0.197</b>	<b>0.231</b>
	192	0.238	0.263	<b>0.232</b>	<b>0.260</b>
	336	0.248	0.273	<b>0.241</b>	<b>0.270</b>
	720	0.249	0.275	<b>0.241</b>	<b>0.273</b>
	Avg.	0.234	0.261	<b>0.228</b>	<b>0.258</b>

  

Metric		iTransformer		+ CM	
		MSE	MAE	MSE	MAE
PEMS03	12	0.071	0.174	<b>0.063</b>	<b>0.168</b>
	24	0.097	0.208	<b>0.087</b>	<b>0.197</b>
	48	0.161	0.272	<b>0.133</b>	<b>0.250</b>
	96	0.240	0.338	<b>0.212</b>	<b>0.316</b>
	Avg.	0.142	0.248	<b>0.124</b>	<b>0.231</b>
PEMS04	12	0.081	0.188	<b>0.075</b>	<b>0.181</b>
	24	0.099	0.211	<b>0.086</b>	<b>0.196</b>
	48	0.133	0.246	<b>0.108</b>	<b>0.222</b>
	96	0.172	0.283	<b>0.125</b>	<b>0.242</b>
	Avg.	0.121	0.232	<b>0.098</b>	<b>0.210</b>
PEMS07	12	0.067	0.165	<b>0.061</b>	<b>0.157</b>
	24	0.088	0.190	<b>0.076</b>	<b>0.179</b>
	48	0.113	0.218	<b>0.086</b>	<b>0.188</b>
	96	0.140	0.246	<b>0.104</b>	<b>0.208</b>
	Avg.	0.102	0.205	<b>0.082</b>	<b>0.183</b>
PEMS08	12	0.088	0.193	<b>0.085</b>	<b>0.190</b>
	24	0.138	0.243	<b>0.126</b>	<b>0.234</b>
	48	0.334	0.353	<b>0.178</b>	<b>0.241</b>
	96	0.458	0.436	<b>0.221</b>	<b>0.260</b>
	Avg.	0.254	0.306	<b>0.152</b>	<b>0.231</b>
ECL	96	0.148	0.240	<b>0.140</b>	<b>0.235</b>
	192	0.167	0.258	<b>0.158</b>	<b>0.252</b>
	336	0.179	0.272	<b>0.172</b>	<b>0.267</b>
	720	0.220	0.310	<b>0.202</b>	<b>0.295</b>
	Avg.	0.179	0.270	<b>0.168</b>	<b>0.262</b>
Traffic	96	0.395	0.268	<b>0.391</b>	<b>0.266</b>
	192	0.417	0.277	<b>0.409</b>	<b>0.275</b>
	336	0.433	0.283	<b>0.426</b>	<b>0.282</b>
	720	0.467	0.300	<b>0.460</b>	<b>0.300</b>
	Avg.	0.428	0.282	<b>0.422</b>	<b>0.281</b>

Table C.1: TS forecasting results with 13 datasets.

## D APPLICATION TO UNITS

To demonstrate the effectiveness of our method on a TS foundation model, we apply it to four different TS tasks using UniTS (Gao et al., 2024) on datasets from various domains, under multiple settings, including multi-task, few-shot, and zero-shot settings. All experimental settings follow those outlined in UniTS (Gao et al., 2024). The sections and tables outlining the full experiment results are listed in Table D.1.

Settings	Section	TS downstream tasks			
		FCST	CLS	IMP	AD
Multi-task	D.1	Table 3	Table D.2	-	-
Few-shot	D.2	Table D.3,D.4,D.5	Table D.6,D.7,D.8	Table D.9	Table D.10
Zero-shot	4.2.3	Table 3	-	-	-

Table D.1: Summary of experiments.

### D.1 MULTI-TASK LEARNING

For experiments under multi-task settings, we perform 20 TS forecasting and 18 classification tasks, where the full results are shown in Table 3 and Table D.2, respectively.

18 Tasks	Shared (1 model)				Task-specific (18 models)					
	UniTS + CM		UniTS		iTransformer	TimesNet	PatchTST	Pyraformer	Autoformer	GPT4TS
	Sup.	PT	Sup.	PT	Sup.					FT
Heartbeat	67.3	70.2	59.0	69.3	66.8	<b>72.7</b>	65.9	<b>72.7</b>	<a href="#">71.7</a>	69.8
JapaneseVowels	94.1	93.2	93.5	90.8	<a href="#">95.9</a>	<b>97.6</b>	94.1	85.4	94.1	94.6
PEMS-SF	<a href="#">83.2</a>	82.1	<a href="#">83.2</a>	85.0	<a href="#">83.2</a>	77.5	<b>83.8</b>	<a href="#">83.2</a>	79.2	79.2
SelfRegulationSCP2	<b>58.3</b>	51.7	47.8	53.3	48.9	52.8	48.9	<a href="#">56.7</a>	45.0	45.6
SpokenArabicDigits	97.1	93.5	97.5	92.0	<a href="#">97.8</a>	<b>98.7</b>	97.5	92.1	97.3	97.5
UWaveGestureLibrary	<b>84.4</b>	<a href="#">83.8</a>	79.1	75.6	82.2	<b>84.4</b>	81.9	72.2	42.2	81.9
ECG5000	<a href="#">93.4</a>	<a href="#">93.4</a>	92.6	<a href="#">93.4</a>	<a href="#">93.3</a>	92.6	<b>94.3</b>	91.4	91.9	93.0
NonInvasiveFetalECGThorax1	<a href="#">89.5</a>	55.2	<b>90.5</b>	27.1	88.2	88.9	86.5	21.4	21.7	89.7
Blink	<b>99.1</b>	<a href="#">95.6</a>	<b>99.1</b>	91.1	<a href="#">93.3</a>	87.6	89.6	88.2	63.1	92.4
FaceDetection	64.7	54.6	64.1	57.6	66.0	<a href="#">66.2</a>	63.9	<b>67.3</b>	59.2	66.1
ElectricDevices	<a href="#">62.4</a>	60.5	60.3	55.4	57.3	49.5	59.5	<b>65.4</b>	56.1	62.9
Trace	<b>99.0</b>	<a href="#">93.0</a>	91.0	82.0	79.0	91.0	77.0	74.0	60.0	96.0
FordB	<b>76.2</b>	64.2	<a href="#">76.0</a>	62.8	72.7	68.9	61.4	55.3	66.4	77.7
MotionSenseHAR	92.8	<b>94.3</b>	92.8	93.2	<a href="#">93.6</a>	90.6	75.8	88.7	30.2	96.2
EMOPain	75.5	<a href="#">80.8</a>	78.0	80.3	79.4	78.0	79.2	<b>81.4</b>	69.9	79.4
Chinatown	<a href="#">97.7</a>	<b>98.0</b>	<a href="#">97.7</a>	<b>98.0</b>	97.4	<a href="#">97.7</a>	<a href="#">97.7</a>	27.4	96.8	96.5
MelbournePedestrian	<a href="#">89.3</a>	78.3	87.3	77.0	<a href="#">89.3</a>	<b>95.7</b>	80.4	52.3	75.0	94.0
SharePriceIncrease	62.9	66.6	61.9	<b>68.4</b>	61.9	65.0	<a href="#">68.0</a>	63.1	61.5	63.7
1st Count (/18)	5	2	2	2	0	5	2	4	0	-
2nd Count (/18)	6	5	3	1	5	2	2	2	1	-
Average Score	<b>82.0</b>	78.3	80.6	75.1	80.3	<a href="#">80.9</a>	78.1	68.8	65.6	82.0

Table D.2: Results of multi-task classification.

## D.2 FEW-SHOT LEARNING

For the few-shot tasks, we conduct four distinct tasks: forecasting (FCST), classification (CLS), imputation (IMP), and anomaly detection (AD), which are discussed in Sections D.2.1, D.2.2, D.2.3, and D.2.4, respectively.

### D.2.1 FEW-SHOT FORECASTING

The results of few-shot forecasting with data ratios of 5%, 15%, and 20% are shown in Tables D.3, D.4, and D.5, respectively.

5%		iTransformer		UniTS				UniTS + CM			
		FT		PT		FT		PT		FT	
Data	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	0.554	0.500	<b>0.405</b>	<b>0.417</b>	<u>0.418</u>	<u>0.424</u>	0.421	0.427	0.421	0.425
	192	0.440	0.438	0.400	0.406	0.377	0.397	<u>0.386</u>	<u>0.402</u>	<b>0.370</b>	<b>0.389</b>
	336	0.478	0.467	0.425	0.433	<u>0.420</u>	0.433	0.423	<u>0.431</u>	<b>0.416</b>	<b>0.425</b>
	720	0.483	0.480	0.446	0.457	0.439	0.452	<b>0.424</b>	<u>0.444</u>	<u>0.428</u>	<b>0.443</b>
RiverFlow	24	1.141	0.514	1.115	0.504	<u>1.112</u>	0.504	<b>1.097</b>	<u>0.503</u>	<b>1.097</b>	<b>0.500</b>
ETTm1	96	0.504	0.462	0.436	0.434	<u>0.384</u>	<u>0.404</u>	0.428	0.436	<b>0.354</b>	<b>0.384</b>
	192	0.555	0.485	0.462	0.448	<u>0.414</u>	<u>0.418</u>	0.475	0.458	<b>0.393</b>	<b>0.405</b>
	336	0.567	0.496	0.560	0.494	<u>0.453</u>	<u>0.442</u>	0.550	0.493	<b>0.420</b>	<b>0.423</b>
	720	0.659	0.539	0.703	0.558	<u>0.526</u>	<u>0.483</u>	0.689	0.554	<b>0.483</b>	<b>0.455</b>
Average		0.598	0.487	0.549	0.461	<u>0.505</u>	<u>0.440</u>	0.546	0.462	<b>0.489</b>	<b>0.429</b>

Table D.3: Results of few-shot forecasting (5%).

15%		iTransformer		UniTS				UniTS + CM			
		FT		PT		FT		PT		FT	
Data	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	0.441	0.440	0.403	0.412	<b>0.399</b>	<b>0.409</b>	0.416	0.423	<u>0.403</u>	<u>0.411</u>
	192	0.398	0.410	0.396	0.404	0.394	<u>0.399</u>	<u>0.388</u>	0.403	<b>0.387</b>	<b>0.399</b>
	336	0.436	0.441	0.432	0.435	0.441	0.435	<b>0.419</b>	<u>0.435</u>	<u>0.430</u>	<b>0.431</b>
	720	0.438	0.453	0.448	0.457	0.449	0.453	<b>0.415</b>	<b>0.442</b>	<u>0.433</u>	<u>0.446</u>
RiverFlow	24	<b>1.067</b>	<b>0.467</b>	1.077	0.492	<u>1.069</u>	0.489	1.073	0.492	1.072	<u>0.487</u>
ETTm1	96	0.423	0.419	0.407	0.420	<u>0.353</u>	<u>0.386</u>	0.408	0.426	<b>0.342</b>	<b>0.380</b>
	192	0.464	0.439	0.434	0.432	<u>0.384</u>	<u>0.400</u>	0.449	0.447	<b>0.377</b>	<b>0.399</b>
	336	0.492	0.457	0.490	0.464	<u>0.416</u>	<u>0.420</u>	0.502	0.475	<b>0.406</b>	<b>0.148</b>
	720	0.558	0.493	0.641	0.537	<u>0.480</u>	<u>0.455</u>	0.621	0.530	<b>0.470</b>	<b>0.451</b>
Average		0.524	0.450	0.525	0.450	<u>0.487</u>	<u>0.428</u>	0.522	0.452	<b>0.481</b>	<b>0.425</b>

Table D.4: Results of few-shot forecasting (15%).

20%		iTransformer		UniTS				UniTS + CM			
		FT		PT		FT		PT		FT	
Data	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	0.418	0.426	0.411	0.414	<b>0.391</b>	<b>0.405</b>	0.411	0.422	<u>0.395</u>	<u>0.409</u>
	192	0.395	0.407	0.383	<b>0.398</b>	0.395	0.403	<b>0.381</b>	<u>0.400</u>	<u>0.390</u>	<u>0.400</u>
	336	0.431	0.438	<b>0.419</b>	<u>0.431</u>	0.430	<b>0.430</b>	<u>0.423</u>	<b>0.430</b>	0.438	0.433
	720	<u>0.431</u>	<u>0.449</u>	0.440	0.453	0.444	0.449	<b>0.418</b>	<b>0.422</b>	0.456	0.456
RiverFlow	24	<b>1.056</b>	<b>0.462</b>	1.069	<u>0.487</u>	1.069	0.489	1.071	0.487	<u>1.067</u>	0.489
ETTm1	96	0.408	0.410	0.409	0.421	<u>0.344</u>	<u>0.379</u>	0.403	0.425	<b>0.339</b>	<b>0.376</b>
	192	0.444	0.428	0.443	0.439	<u>0.377</u>	<u>0.397</u>	0.450	0.450	<b>0.375</b>	<b>0.396</b>
	336	0.471	0.445	0.505	0.472	<u>0.408</u>	<u>0.418</u>	0.507	0.481	<b>0.403</b>	<b>0.415</b>
	720	0.536	0.482	0.648	0.536	<u>0.472</u>	<u>0.453</u>	0.621	0.531	<b>0.466</b>	<b>0.448</b>
Average		0.510	0.438	0.525	0.450	<u>0.486</u>	<u>0.425</u>	0.521	0.453	<b>0.482</b>	<b>0.425</b>

Table D.5: Results of few-shot forecasting (20%).

## D.2.2 FEW-SHOT CLASSIFICATION

The results of few-shot classification with data ratios of 5%, 15%, and 20% are shown in Tables D.6, D.7, and D.8, respectively.

5%	iTransformer	UniTS		UniTS + CM	
	FT	PT	FT	PT	FT
ECG200	<u>78.0</u>	67.0	77.0	<b>80.0</b>	77.0
Handwriting	<u>5.4</u>	4.6	4.7	4.8	<b>5.5</b>
SelfRegulationSCP1	62.8	66.2	<u>74.7</u>	<b>77.8</b>	73.7
RacketSports	37.5	31.6	35.5	<u>39.5</u>	<b>47.4</b>
Epilepsy	39.9	44.9	<u>47.1</u>	44.9	<b>57.2</b>
StarLightCurves	85.1	82.3	83.8	<b>86.3</b>	<u>85.4</u>
Average	51.4	49.4	53.8	<b>54.9</b>	<u>54.8</u>

Table D.6: Results of few-shot classification (5%).

15%	iTransformer	UniTS		UniTS + CM	
	FT	PT	FT	PT	FT
ECG200	<u>81.0</u>	74.0	78.0	73.2	<b>82.0</b>
Handwriting	<b>9.8</b>	7.3	8.1	<u>9.2</u>	8.5
SelfRegulationSCP1	67.9	59.0	<b>76.5</b>	<u>69.3</u>	68.6
RacketSports	<b>54.6</b>	40.1	50.7	44.7	<u>51.3</u>
Epilepsy	41.3	52.9	58.0	<u>61.6</u>	<b>68.1</b>
StarLightCurves	84.2	85.8	<b>87.1</b>	<u>85.9</u>	85.5
Average	56.5	53.2	<u>59.7</u>	55.4	<b>60.4</b>

Table D.7: Results of few-shot classification (15%).

20%	iTransformer	UniTS		UniTS + CM	
	FT	PT	FT	PT	FT
ECG200	81.0	76.0	77.0	<b>85.0</b>	<u>82.0</u>
Handwriting	<b>11.8</b>	8.0	8.5	7.6	<u>9.8</u>
SelfRegulationSCP1	<u>77.1</u>	68.6	70.6	<b>77.8</b>	74.4
RacketSports	<u>54.6</u>	51.3	<b>57.9</b>	38.8	50.7
Epilepsy	62.3	<u>81.9</u>	72.5	<b>84.1</b>	61.6
StarLightCurves	84.8	87.3	86.0	<b>90.0</b>	<u>87.8</u>
Average	59.9	58.9	<u>63.6</u>	60.0	<b>64.8</b>

Table D.8: Results of few-shot classification (20%).



### D.2.3 FEW-SHOT IMPUTATION

The results of few-shot imputation with data ratios of 25% and 50% are shown in Table D.9

Ratio			ECL	ETTh1	ETTh2	ETTh1	ETTh2	Weather	Avg.
25%	TimesNet	FT	0.245	0.369	0.193	0.442	0.119	0.106	0.246
	PatchTST		0.195	0.315	0.147	0.309	<u>0.092</u>	0.089	0.191
	iTransformer		0.174	0.301	0.185	0.254	0.113	0.087	0.186
	UniTS	PT	<u>0.139</u>	0.311	0.178	0.268	0.102	0.078	0.179
		FT	0.160	<u>0.284</u>	<u>0.150</u>	<u>0.241</u>	<b>0.090</b>	<u>0.077</u>	<u>0.167</u>
	UniTS + CM	PT	<u>0.139</u>	0.310	0.176	0.262	0.100	0.078	0.179
		FT	<b>0.129</b>	<b>0.275</b>	<b>0.149</b>	<b>0.231</b>	<b>0.090</b>	<b>0.073</b>	<b>0.158</b>
	50%	TimesNet	0.258	0.412	0.211	0.607	0.140	0.125	0.292
		PatchTST	0.230	0.353	0.175	0.442	<b>0.111</b>	0.105	0.236
		iTransformer	0.203	0.332	0.205	0.372	0.136	0.106	0.226
		UniTS	0.172	0.352	0.251	0.380	0.134	0.103	0.232
			0.191	<u>0.322</u>	<u>0.198</u>	<u>0.352</u>	0.118	<u>0.095</u>	<u>0.213</u>
		UniTS + CM	<u>0.162</u>	0.353	0.240	0.370	0.128	0.097	0.225
			<b>0.151</b>	<b>0.307</b>	<b>0.197</b>	<b>0.345</b>	<u>0.116</u>	<b>0.093</b>	<b>0.201</b>

Table D.9: Results of few-shot imputation.

### D.2.4 FEW-SHOT ANOMALY DETECTION

The results of few-shot anomaly detection with data ratio of 5% are shown in Table D.10.

		MSL	PSM	SMAP	SMD	SWAT	Avg.
Anomaly Trans.	-	78.0	90.2	68.3	77.8	81.5	79.2
TimesNet	FT	33.9	91.0	68.5	84.0	<b>93.4</b>	74.2
iTransformer	FT	<u>80.4</u>	96.5	67.2	82.4	89.0	83.1
PatchTST	FT	79.9	<u>96.6</u>	68.7	83.8	92.6	84.3
UniTS	PT	73.2	95.5	65.9	81.2	<u>92.9</u>	81.7
	FT	<b>81.3</b>	<b>97.3</b>	<u>71.6</u>	<u>85.5</u>	92.5	<u>85.6</u>
UniTS + CM	PT	73.7	95.5	66.0	82.0	<u>92.9</u>	82.0
	FT	<b>81.3</b>	<b>97.3</b>	<b>75.9</b>	<b>86.2</b>	92.6	<b>86.6</b>

Table D.10: Results of few-shot anomaly detection.

## E APPLICATION TO TIMESIAM

To demonstrate the effectiveness of our proposed model on TimeSiam (Dong et al., 2024), which uses a self-supervised pretraining framework for TS with Siamese networks, we conduct experiments with two datasets that vary in channel size: Exchange, with a small number of channels (8), and ECL, with a large number of channels (321). Specifically, we apply variants of our method by using the domain parameter only during the fine-tuning stage and during both pretraining and fine-tuning stages. The results, shown in Table E.1, validate both components of our method, with the best performance achieved when using domain parameters at both pretraining and fine-tuning stages.

		TimeSiam		+ CM					
Correlation matrix		-		✓		✓		✓	
Domain parameters	Pretrain	-		-		-		✓	
	Fine-tune	-		-		✓		✓	
Dataset	$H$	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange ( $C = 8$ )	96	0.092	0.215	<u>0.089</u>	<b>0.207</b>	<b>0.088</b>	<b>0.207</b>	<b>0.088</b>	<u>0.209</u>
	192	<b>0.182</b>	0.306	<b>0.182</b>	<u>0.304</u>	<b>0.182</b>	<b>0.303</b>	<b>0.182</b>	0.305
	336	0.341	0.426	0.336	<u>0.422</u>	<u>0.332</u>	<b>0.417</b>	<b>0.329</b>	<b>0.417</b>
	720	0.806	0.679	0.792	0.670	<u>0.788</u>	<u>0.668</u>	<b>0.783</b>	<b>0.666</b>
	Avg.	0.356	0.407	0.350	0.401	<u>0.349</u>	<u>0.399</u>	<b>0.346</b>	<b>0.398</b>
ECL ( $C = 321$ )	96	0.147	0.239	<b>0.140</b>	<b>0.236</b>	<b>0.140</b>	<b>0.236</b>	<u>0.141</u>	<u>0.237</u>
	192	0.162	0.253	<b>0.157</b>	<u>0.251</u>	<b>0.157</b>	<u>0.251</u>	<b>0.157</b>	<b>0.250</b>
	336	0.175	0.269	<u>0.173</u>	<u>0.268</u>	<u>0.173</u>	<u>0.268</u>	<b>0.172</b>	<b>0.267</b>
	720	0.215	0.304	<b>0.203</b>	<u>0.297</u>	<b>0.203</b>	<u>0.297</u>	<b>0.203</b>	<b>0.296</b>
	Avg.	0.175	0.266	<b>0.168</b>	<u>0.263</u>	<b>0.168</b>	<u>0.263</u>	<b>0.168</b>	<b>0.262</b>

Table E.1: Results of TS forecasting with TimeSiam.

## F MASKED CHANNEL PREDICTION

Tables F.1 and F.2 show the results of masked channel prediction for five datasets (Wu et al., 2021; Liu et al., 2022), indicating significant improvement when the CM is applied to iTransformer compared to when it is not used.

Horizon	Exchange			ECL		
	Avg. MSE(C1~C8)			Avg. MSE(C1~C321)		
	iTrans.	+ CM	Gain(%)	iTrans.	+ CM	Gain(%)
96	0.139	<b>0.138</b>	<b>1.2%</b>	0.846	<b>0.526</b>	<b>37.8%</b>
192	0.236	<b>0.232</b>	<b>1.5%</b>	0.849	<b>0.563</b>	<b>33.7%</b>
336	0.383	<b>0.374</b>	<b>2.4%</b>	0.861	<b>0.594</b>	<b>31.0%</b>
720	0.934	<b>0.917</b>	<b>1.8%</b>	0.891	<b>0.741</b>	<b>16.8%</b>
Avg.	0.423	<b>0.415</b>	<b>1.8%</b>	0.862	<b>0.606</b>	<b>29.7%</b>

Table F.1: Results of masked channel prediction (Exchange, ECL).

Horizon	PEMS04			PEMS07			PEMS08		
	Avg. MSE(C1~C307)			Avg. MSE(C1~C883)			Avg. MSE(C1~C170)		
	iTrans.	+ CM	Gain(%)	iTrans.	+ CM	Gain(%)	iTrans.	+ CM	Gain(%)
12	0.549	<b>0.300</b>	<b>45.4%</b>	0.835	<b>0.343</b>	<b>58.9%</b>	0.628	<b>0.200</b>	<b>68.1%</b>
24	0.718	<b>0.351</b>	<b>51.1%</b>	0.865	<b>0.448</b>	<b>48.1%</b>	0.678	<b>0.241</b>	<b>64.5%</b>
48	0.750	<b>0.409</b>	<b>45.5%</b>	1.038	<b>0.511</b>	<b>50.8%</b>	1.197	<b>1.059</b>	<b>11.5%</b>
96	0.758	<b>0.513</b>	<b>32.3%</b>	1.040	<b>0.640</b>	<b>38.5%</b>	1.375	<b>1.217</b>	<b>11.5%</b>
Avg.	0.694	<b>0.393</b>	<b>43.3%</b>	0.945	<b>0.486</b>	<b>48.6%</b>	0.970	<b>0.679</b>	<b>29.9%</b>

Table F.2: Results of masked channel prediction (PEMS datasets).