Dataset Details

The details of the datasets used in the experiments are as follows:

- **SleepEEG:** The SleepEEG dataset(Kemp et al. 2000) consists of 153 whole-night sleep electroencephalogram (EEG) recordings, monitored using sleep cassette tapes. The data comes from 82 healthy subjects, with EEG signals sampled at 100 Hz. Each sample is associated with one of five sleep patterns/stages: Wakefulness (W), Non-Rapid Eye Movement (N1, N2, N3), and Rapid Eye Movement (REM). This dataset includes a mix of high- and low-frequency patterns and has been used in several studies as a pre-training dataset. We used the preprocessed version of the dataset, which was released by the TF-C(Zhang et al. 2022) researchers.
- **Gesture:** The Gesture(Liu et al. 2009) dataset contains accelerometer measurements for 8 simple gestures, each varying based on the path of hand movement. The eight gestures include sliding the hand left, right, up, and down, waving in a clockwise or counterclockwise direction, waving in a square pattern, and waving in the shape of a right arrow. The classification labels correspond to these 8 different types of gestures. This dataset originates from the UCR dataset, and TF-C researchers have merged and cleaned it, making it widely used as a standalone small-sample dataset.
- **FD-B:** FD-B is a subset extracted from the FD(Lessmeier et al. 2016) dataset under condition B, which was collected from an electromechanical drive system that monitors the condition of rolling bearings and detects any damage. The data collected under different conditions is divided into 4 subsets, with parameters including rotational speed, load torque, and radial force. Each rolling bearing can be categorized into one of three classes: undamaged, internally damaged, and externally damaged. We used the preprocessed data provided by the TF-C researchers.
- **EMG:** Electromyography(EMG) measures the electrical activity of muscles in response to nerve stimulation and can be used to diagnose certain muscular dystrophies and neuropathies. The EMG(Goldberger et al. 2000) dataset consists of single-channel EMG recordings from the tibialis anterior muscle of three volunteers: one healthy, one with neuropathy, and one with myopathy. The data is sampled at a frequency of 4 kHz. Each patient, corresponding to their condition, represents a distinct classification category, with a total of three classes. We used the preprocessed data provided by the TF-C researchers.
- **EPI:** The Epilepsy(EPI)(Andrzejak et al. 2001) dataset contains single-channel EEG measurements from 500 subjects. For each subject, brain activity was recorded for 23.6 seconds. The dataset was then divided and shuffled into 11,500 samples, each 1 second long, sampled at a frequency of 178 Hz. There are a total of 11,500 EEG samples classified into 2 categories, corresponding to epilepsy patients and normal patients. We also used the preprocessed data provided by the TF-C researchers.
- **HAR:** The HAR(Anguita et al. 2013) dataset includes recordings of 30 healthy volunteers performing six daily activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down. Each sample is labeled with one of these six activities. The dataset was obtained by measuring tri-axial acceleration and tri-axial angular velocity using wearable sensors on a smartphone, sampled at a frequency of 50 Hz. We also used the preprocessed dataset provided by the TF-C researchers.
- **128 UCR:** The UCR dataset(Dau et al. 2018) includes a total of 128 time series classification datasets, making it one of the largest and most comprehensive benchmarks for time series classification. The dataset covers a wide range of fields, including healthcare, sports, and transportation, allowing for a thorough evaluation of algorithm classification performance. Although the dataset contains numerous subsets, the majority of these subsets have fewer than 500 training samples, making it suitable for evaluating transfer performance. In our experiments, we did not standardize the sample lengths; instead, we used the maximum sample length from each subset directly.
- **C-MAPSS:** The C-MAPSS(Saxena et al. 2008) dataset, generated by NASA using the Commercial Modular Aero-Propulsion System Simulation, is used to predict the remaining useful life (from 100% to 0%) of an engine. It records key monitoring parameters of major components such as the combustion chamber and compressor during each simulated flight. The dataset contains 4 subsets FD001 to FD004, covering different operating conditions and causes of degradation. FD001 and FD003 are single-condition datasets, while FD002 and FD004 are multi-condition composite datasets. In our experiments, we selected the Low Pressure Compressor Outlet Temperature as the input variable to predict the remaining life of the engine.
- **Bearing:** The Bearing dataset(Wang et al. 2018) originates from a real bearing machine setup, recording the vibration signals throughout the entire lifecycle of the bearings, from normal operation to eventual failure. It is also used to predict the remaining useful life. The dataset includes three different operational states with working speeds of 2100 rpm, 2250 rpm, and 2400 rpm. Vibration signals were collected using a DT9837 portable dynamic signal analyzer, with a sampling frequency of 25.6 kHz, a sampling interval of 1 minute, and each sampling duration of 1.28 seconds. We use a window size of 8192 to construct non-overlapping training and testing samples.

Detailed information about the datasets used in this study is shown in Table 1. The sample lengths for the Gesture, FD-B, EMG, EPI, and HAR datasets listed in the table correspond to the lengths of the original dataset. In our experiments, the sample length used for all these datasets is 178, following the protocols established by TF-C(Zhang et al. 2022) and SimMTM(Dong et al. 2023).

Task	Dataset	Sub-datasets	Train	Val.	Test	Length	Freq.(Hz)	Classes
Pre-training	SleepEEG	-	371,055	107,730	-	200	100	5
Classification	Gesture	-	320	120	120	178	100	8
	FD-B	-	60	21	13,559	5,120	64k	3
	EMG	-	122	41	41	1,500	4k	3
	EPI	-	60	20	11,420	178	178	2
	HAR	-	10,299	1,471	2,947	206	50	6
	128 UCR	-	16∼8,926	-	20∼16,800	15∼2,844	-	2∼60
Regression	C-MAPSS	FD001	13,713	3,518	9,699	35	1	-
		FD002	35,831	9,088	25,244	35	1	-
		FD003	17,406	3,914	13,196	35	1	-
		FD004	41,831	10,952	32,896	35	1	-
	Bearing	OC-A	123	644	1,328	8,192	25.6k	-
		OC-B	491	644	3,656	8,192	25.6k	-
		OC-C	2,535	9,984	8,000	8,192	25.6k	-

Table 1: Detailed information about the datasets used in this study

Hyperparameter	Value	Description
$epoch$	100	The maximum pre-training epoch.
bs	512	The batch size during pre-training.
lr	0.0002	Initial learning rate.
α	0.995	The momentum factor.
β_1	0.0	Minimum length ratio of frequency masking.
β_2	0.7	Maximum length ratio of frequency masking.

Table 2: The hyperparameter configuration of FEI.

Pre-training Details of FEI

The proposed FEI has few tunable hyperparameters, with the key performance-related settings listed in Table 2. During pre-training, an exponential decay strategy with a decay coefficient of 0.9 is used. The validation set from the pre-training dataset SleepEEG is used to monitor test loss and prevent overfitting, employing an early stopping mechanism with a patience value of 5 steps. The optimizer used is AdamW, with β coefficients set to [0.9, 0.999].

Background and Baselines

Background

Currently, research on representation learning methods for time series is still in its early stages. Unlike other fields such as computer vision (CV) and natural language processing (NLP), it has yet to establish community standards. The benchmark experimental protocols for time series representation learning research are not fully established. By reviewing recent studies and their corresponding publicly available code, we found that existing experimental protocols mainly fall into two categories:

1. Unsupervised/self-supervised pre-training is conducted on the target dataset, followed by either end-to-end fine-tuning or the construction of non-end-to-end machine learning algorithms on the same dataset. This constitutes an integrated approach, as seen in methods like TS2Vec (Yue et al. 2022), InfoTS (Luo et al. 2023), TimesURL (Liu and Chen 2024), and TimeDRL (Chang et al. 2024).
2. Unsupervised/self-supervised pre-training is conducted on a pre-training dataset, followed by end-to-end fine-tuning on a target dataset that is different from the pre-training dataset. This is a two-stage approach, as seen in methods like TF-C(Zhang et al. 2022) and SimMTM(Dong et al. 2023).

Both of these approaches have certain limitations in validating the generalization performance of representation learning algorithms. First, Approach 1) cannot explicitly validate the generalization quality of the representations learned during the unsupervised representation learning stage because both pre-training and validation are confined to the same dataset. Approach 2) can only demonstrate that the representation learning algorithm helped the encoder find a good initial state during pre-training. However, it cannot validate whether the representations learned by the encoder meet the general requirements of the time series domain, as this approach lacks methods like linear evaluation that directly assess representation quality. Representation quality determines the potential of a method to train a general model with a larger encoder.

The evaluation of both representation quality and end-to-end fine-tuning performance is equally important in the field of time series representation learning, as it helps push the field towards generalization and standardization. The lack of consistency in experimental strategies has also led to significant fragmentation in the conclusions drawn by research in this field. One of the main objectives of this study is to develop a more comprehensive experimental strategy and benchmark to help unify research in this field. Therefore,

Method	Hyperparameter	Value
TS2Vec	$epoch$	30
	bs	32
	lr	0.001
TimeDRL	$epoch$	30
	bs	32
	lr	0.0001
	contrastive weight	0.1
	position embedding	learnable
TF-C	$epoch$	100
	bs	256
	lr	0.0003
	temperature	0.2
	scale ratio	1.5
	jitter ratio	2
	max. seg.	12
TimesURL	$epoch$	30
	bs	128
	lr	0.0001
	lmd	0.01
	temperature	1.0
	segment num.	3
	mask ratio	0.05

Table 3: The main hyperparameter settings of baselines (part 1).

in the implementation of our baseline, we consider the differences in implementing various approaches and conduct the experiments within a unified experimental framework as much as possible.

Baselines

Based on the above objectives, we conduct unified pre-training using 1D ResNet as the encoder for various baseline methods, including 1) TS2Vec, 2) TimeDRL, 3) TF-C, 4) TimesURL, 5) SimMTM, and 6) InfoTS, with the exception of TimesURL. Additionally, we design 2 types of experiments—linear evaluation and end-to-end fine-tuning on small samples—to comprehensively assess the performance differences among these methods. These methods are reproduced based on their publicly available code. The hyperparameter settings for all methods are kept consistent with those in their open-source code. Since TimeDRL has significantly different pre-training hyperparameters across different datasets, we select the hyperparameter settings that are similar across most datasets. The main hyperparameter settings for each baseline are shown in Table 3 and 4.

Full Results

Downstream Tasks

In this study, we validate performance on both classification and regression tasks. For classification tasks, the performance metrics include accuracy, precision, recall, and F1 score. For regression tasks, the performance metrics include MSE and MAE. The complete experimental results for both tasks are presented in Tables 5, 6, 7, and 8.All experiments were conducted on an Ubuntu 18.04 system with Python 3.8.15 and PyTorch 2.0.1, utilizing an Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz, 64GB of RAM, and an Nvidia RTX 3090 GPU. We fixed the random seed at '2024' to ensure the reproducibility of all results.

Method	Hyperparameter	Value
SimMTM	$epoch$	30
	bs	128
	lr	0.0001
	length of masking	$1/3$
	temperature	0.2
	positive num.	3
	mask ratio	0.5
InfoTS	$epoch$	30
	bs	512
	lr	0.001
	meta lr	0.01
	β	0.5
	meta β	0.1
	p_1	0.2
	p_2	0.0
	mask mode	binomial

Table 4: The main hyperparameter settings of baselines (part 2).

Ablation

We have reported the precision results of 6 ablation models on the EMG dataset in the main text. The complete ablation results are presented in Table 9.

The ablation results show that each module design of FEI has varying degrees of impact on FEI's generalization ability. Overall, the most significant influences are from target embedding inference, mask prompting, and momentum encoder. The target embedding inference is the direct means by which FEI controls the embedding space. The mask prompting is the basis for meaningful embedding inference in FEI, as it indicates the inference targets, i.e., the frequency masking positions. Due to the mask inference branch, removing the momentum encoder will not fully degrade FEI, but it remains crucial for target embedding inference. The other modules serve as enhancement modules for FEI, having a relatively smaller impact on performance but still being indispensable. Among these, the subspace projector acts as a relaxation factor, allowing FEI to perform inferences in a more achievable subspace rather than the original embedding space. It has a significant impact on the transfer results for some datasets (e.g., the EMG dataset show in Table 9). The mask inference and target embedding inference jointly reinforce the learning process of FEI.

Masking Strategy

We describe 3 masking strategies used in FEI to construct target series in the main text: Discrete Frequency

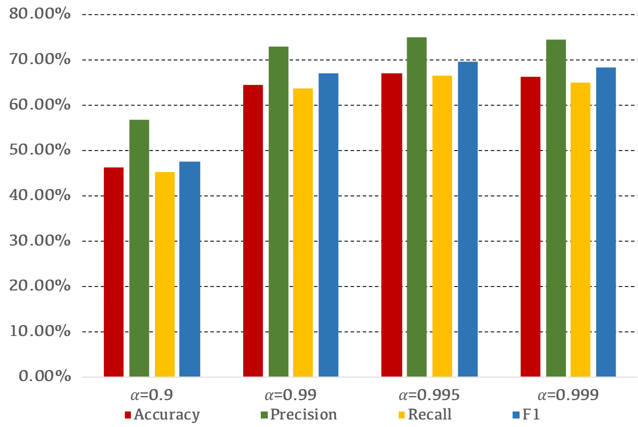

Figure 1: The linear evaluation results for varying momentum factors α on FD-B dataset.

Masking(DFM), Continuous Frequency Masking(CFM), and Time-domain Masking(TDM), with the full results listed in Table 10.

Sensitivity

We have reported FEI's sensitivity to masking ratio $\beta_1$and β_2. In addition, the momentum factor α determines the update speed of the momentum encoder, which is crucial for most methods based on momentum update strategies, such as BYOL(Grill et al. 2020) and I-JEPA(Assran et al. 2023). Therefore, we further analyze the impact of the momentum update factor α on FEI. The results are shown in Figure 1.The results indicate that the optimal range for α in FEI is between 0.99 and 0.999, within which sufficient representational generalization can be achieved. Ultimately, we use $\alpha = 0.995$.

Visualization

FEI demonstrates good sensitivity to changes in the frequencies of unseen samples. We have already shown the FEI's inference visualization on the unseen Gesture dataset in the main text. Here, we provide additional, more detailed visualizations on the **FD-B** and **EMG** datasets, which have significant frequency differences from the pre-training dataset SleepEEG, as shown in Figures 2 and 3.

In the figures, "Target Series #n" represents the target series constructed from the original series using the frequency mask shown below. In the "Inference Results" section, the inverted triangle represents the true embedding of the target series, the star represents the inferred embedding of the target series obtained by FEI using the mask prompt and original series, and the circle represents the embedding of the original series. All embeddings are displayed using t-SNE dimensionality reduction.

Both sets of visualizations include low-frequency and high-frequency masking to varying degrees, and FEI accurately infers embeddings, even though these samples were never used to train FEI. Thanks to the modeling approach of embedding inference, FEI can generate meaningful embedding for time series sample based on its frequency characteristics, which is the source of FEI's strong generalization ability.

Future Work

The proposed FEI introduces a new modeling approach for self-supervised representation learning of time series. Through extensive experiments, we have validated and analyzed the effectiveness and superiority of FEI. However, advancing this field remains crucial, as developing a fully generalized temporal representation model holds significant importance for the entire time series analysis domain. Based on the current limitations of FEI and recent research progress, we offer some recommendations for future research.

FEI successfully achieves sample-level general representation modeling and has shown significant improvements in sequence-level downstream tasks such as classification and regression. However, we have not yet explored how this architecture could be applied to finer-grained modeling at the time-step level. This could be valuable for certain downstream tasks like point-to-point anomaly detection or stepwise time series prediction. Our findings suggest that continuous semantic modeling at the step level is beneficial for obtaining generalizable time series representation models. **This leads to our first suggestion for future work: exploring continuous semantic modeling frameworks at the time-step level, which may not be limited to the FEI architecture.**

The inherent diversity of time series data, which can describe a wide range of objects, results in significant differences in key features such as trends, cycles, and noise levels between different series. Thus, learning universal representations for time series remains a highly challenging field. Constructing a comprehensive, large-scale time series corpus that covers all possible objects of description is extremely difficult. This challenge also differentiates the design of self-supervised learning algorithms in this field from those in CV and NLP, which benefit from vast amounts of data. The frequency inference approach of FEI offers a new perspective for training general time series representation models with limited samples. By using frequency-domain processing methods such as frequency masking, as demonstrated in this paper, it is possible to construct a large number of new samples that are temporally continuous but semantically distinct from the original samples. These differences and relationships among the constructed samples can guide self-supervised learning. Our proposed FEI utilizes embedding inference to model semantic relationships, representing a practical application of this idea. Moving forward, we believe it is possible to develop representation learning models in the time series domain that achieve sufficient representational power with fewer pretraining samples. **Therefore, a second potential research direction is to explore frameworks for mining semantic relationships within time series, which can train general representation models based on a large number of constructible samples. This presents an alternative research path to building vast time series pretraining corpora.**

Figure 2: Visualization of target series and inference results on the **FD-B** dataset.

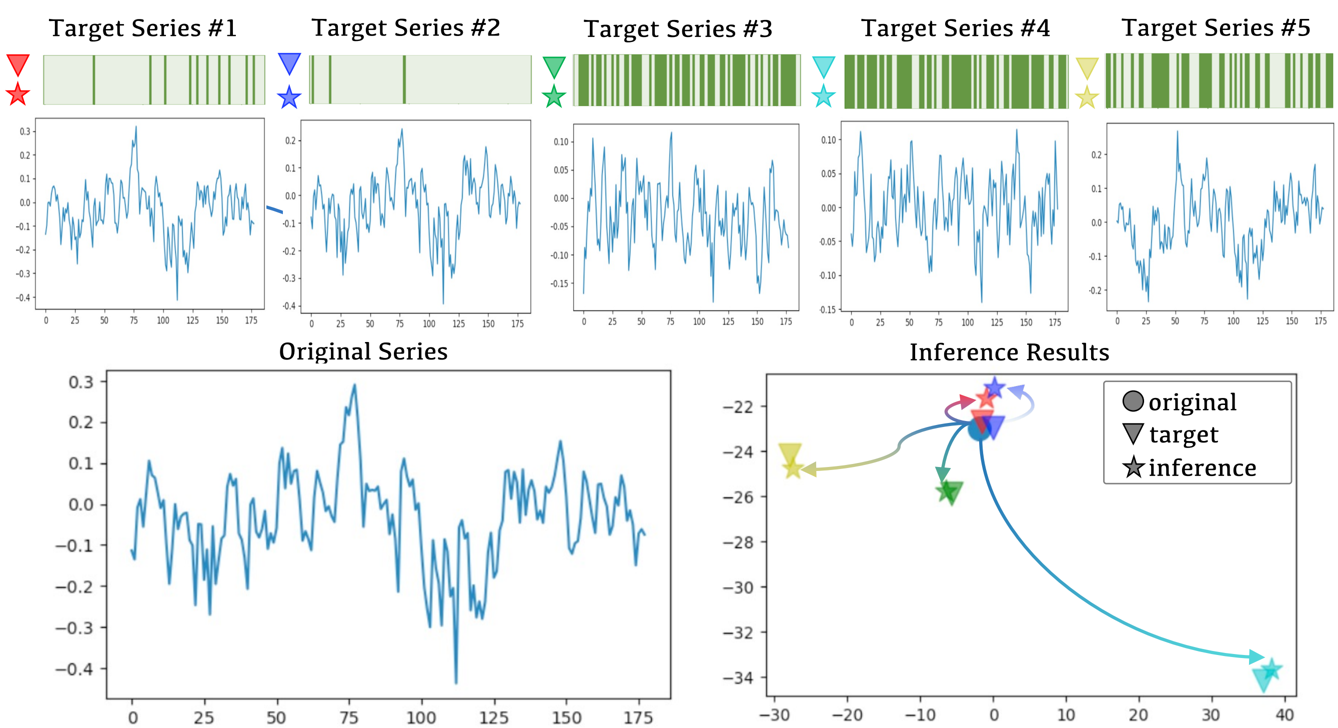

Figure 3: Visualization of target series and inference results on the **EMG** dataset.

Datasets	Methods	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)	Mean(%)
Gesture	Rand. Init.	12.50	12.50	1.56	2.78	7.34
	TS2Vec	63.33	63.33	60.40	61.08	62.04
	TimeDRL	50.00	50.00	43.47	43.47	46.74
	TF-C	57.50	57.50	54.75	54.30	46.01
	TimesURL	69.72	69.72	64.66	65.60	67.43
	SimMTM	74.17	74.17	71.68	70.45	72.62
	InfoTS	64.17	64.17	58.78	60.32	61.86
	FEI	**75.00**	**75.00**	**75.00**	**72.54**	**74.39**
FD-B	Rand. Init.	11.39	35.10	30.68	9.00	21.54
	TS2Vec	43.59	31.91	28.79	28.83	33.28
	TimeDRL	40.63	42.09	37.96	37.03	39.43
	TF-C	45.53	33.36	48.51	20.91	37.08
	TimesURL	54.44	63.99	52.88	55.40	56.68
	SimMTM	60.74	69.98	57.96	60.94	62.41
	InfoTS	60.71	70.33	63.41	64.10	64.64
	FEI	**67.25**	**75.09**	**66.58**	**69.68**	**69.68**
EMG	Rand. Init.	46.34	33.33	15.45	21.11	16.52
	TS2Vec	**92.68**	84.71	**95.45**	**88.22**	**90.27**
	TimeDRL	63.41	48.30	43.77	45.65	50.28
	TF-C	78.05	68.44	74.49	70.18	72.79
	TimesURL	**92.68**	84.71	**95.45**	**88.22**	**90.27**
	SimMTM	85.37	74.12	90.83	77.67	82.00
	InfoTS	87.80	66.67	59.02	62.54	78.69
	FEI	87.80	**90.20**	88.65	88.05	88.05
EPI	Rand. Init.	19.79	50.00	9.89	16.52	24.05
	TS2Vec	96.41	94.66	94.10	94.38	94.89
	TimeDRL	77.85	71.53	67.54	68.89	71.45
	TF-C	85.75	80.66	88.74	89.09	86.06
	TimesURL	95.42	93.62	92.23	92.90	93.54
	SimMTM	96.42	**95.43**	93.59	94.48	94.98
	InfoTS	96.27	94.82	93.62	94.21	94.73
	FEI	**96.84**	95.05	**95.00**	**95.02**	**95.02**
HAR	Rand. Init.	36.51	33.59	37.66	23.38	32.79
	TS2Vec	78.91	78.40	81.44	77.76	79.13
	TimeDRL	70.31	69.03	73.21	64.88	69.36
	TF-C	67.56	66.14	80.49	64.82	69.75
	TimesURL	79.10	78.51	81.40	78.08	79.27
	SimMTM	77.13	76.33	80.05	75.57	77.27
	InfoTS	78.35	77.55	**81.61**	76.37	78.47
	FEI	**79.54**	**79.06**	**81.61**	**78.30**	**79.63**
128 UCR	Rand. Init.	39.03	35.07	26.93	26.86	31.97
	TS2Vec	72.50	69.31	71.15	67.61	70.15
	TimeDRL	61.61	59.76	60.21	57.49	59.77
	TF-C	61.88	58.68	63.56	57.09	60.31
	TimesURL	69.53	66.24	69.02	64.30	67.27
	SimMTM	75.34	73.37	74.96	72.75	74.11
	InfoTS	73.13	70.74	72.34	70.18	71.60
	FEI	**78.17**	**76.37**	**77.94**	**76.04**	**77.13**

Table 5: The full results of **linear evaluation** on 6 classification datasets.

Method	C-MAPSS								Bearing					
	FD001		FD002		FD003		FD004		OC-A		OC-B		OC-C	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Rand. Init.	0.065	0.233	0.094	0.282	0.048	0.196	0.079	0.265	0.649	0.739	0.575	0.689	0.358	0.584
TS2Vec	0.037	0.153	**0.081**	**0.250**	0.037	0.148	0.233	0.405	0.103	**0.203**	0.194	0.253	0.084	0.136
TimeDRL	0.046	0.179	0.096	0.284	0.041	0.157	0.075	0.254	0.107	0.265	**0.074**	**0.177**	0.054	0.102
TF-C	0.353	0.495	0.268	0.499	0.573	0.610	0.606	0.642	0.523	0.533	0.472	0.543	0.812	0.858
TimesURL	0.046	0.174	0.096	0.274	0.045	0.160	0.173	0.336	**0.101**	0.205	0.140	0.248	0.034	0.110
SimMTM	0.035	0.148	0.094	0.278	0.035	0.142	0.077	0.254	0.105	0.265	0.146	0.248	0.037	0.083
InfoTS	0.036	0.149	0.092	0.275	**0.033**	0.160	0.088	0.274	0.105	0.263	0.137	0.241	0.026	0.068
FEI	**0.034**	**0.145**	0.099	0.284	**0.033**	**0.132**	**0.068**	**0.236**	0.104	0.261	0.108	0.216	**0.011**	**0.047**

Table 6: The full results of **linear evaluation** on 2 regression datasets.

Datasets	Methods	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)	Mean(%)
Gesture	Rand. Init.	68.33	68.33	63.87	64.34	66.22
	TS2Vec	72.50	72.50	71.59	71.72	72.08
	TimeDRL	66.67	66.67	61.33	61.63	64.08
	TF-C	70.00	70.00	65.32	66.73	68.01
	TimesURL	73.33	73.33	72.13	72.45	67.43
	SimMTM	76.67	76.67	75.28	74.03	75.66
	InfoTS	71.67	71.67	67.30	67.75	69.60
	FEI	**77.50**	**77.50**	**75.35**	**75.70**	**76.51**
FD-B	Rand. Init.	69.61	77.53	70.23	72.27	72.41
	TS2Vec	48.31	53.96	45.20	44.64	48.03
	TimeDRL	47.97	50.22	45.56	47.10	47.71
	TF-C	65.48	73.93	63.98	67.30	67.67
	TimesURL	54.41	63.90	53.33	56.45	57.02
	SimMTM	63.49	72.96	63.46	66.76	66.67
	InfoTS	62.99	72.57	65.12	66.87	66.89
	FEI	**70.99**	**78.52**	**71.46**	**74.29**	**73.82**
EMG	Rand. Init.	95.12	96.83	96.83	93.62	95.60
	TS2Vec	78.05	69.27	69.74	68.17	71.31
	TimeDRL	78.05	64.15	85.35	65.70	73.31
	TF-C	92.68	94.53	90.63	92.31	92.54
	TimesURL	73.17	70.05	69.09	67.03	69.84
	SimMTM	87.80	71.37	92.24	72.64	81.01
	InfoTS	**97.56**	**98.04**	**98.33**	**98.14**	**98.02**
	FEI	**97.56**	**98.04**	98.04	**98.14**	97.95
EPI	Rand. Init.	80.21	50.00	40.11	44.51	53.71
	TS2Vec	95.60	94.09	92.38	93.20	93.82
	TimeDRL	94.05	86.79	93.98	89.82	91.16
	TF-C	95.28	93.32	92.07	92.68	93.34
	TimesURL	96.67	96.01	93.86	94.89	95.36
	SimMTM	96.22	94.84	93.47	94.14	94.67
	InfoTS	97.07	95.32	95.43	**95.37**	95.80
	FEI	**97.24**	**96.43**	95.05	**95.72**	**96.11**
128 UCR	Rand. Init.	75.44	73.57	75.46	73.07	74.38
	TS2Vec	67.44	65.49	63.95	63.02	64.97
	TimeDRL	63.36	61.64	61.56	60.04	61.65
	TF-C	78.50	76.86	78.38	78.38	76.47
	TimesURL	79.41	77.79	78.84	77.21	78.31
	SimMTM	80.42	78.83	79.91	78.46	79.41
	InfoTS	81.78	80.28	81.31	79.93	80.82
	FEI	**82.65**	**81.25**	**82.19**	**80.95**	**81.76**

Table 7: The full results of **end-to-end fine-tuning** on 5 small-sample classification datasets.

Method	Bearing					
	OC-A		OC-B		OC-C	
	MSE	MAE	MSE	MAE	MSE	MAE
Rand. Init.	0.1618	0.3727	0.0449	0.1292	0.0116	0.0513
TS2Vec	0.1560	0.3378	0.3561	0.3629	0.0394	0.0939
TimeDRL	0.2312	0.3907	0.1626	0.3055	0.0463	0.1694
TF-C	0.5286	0.5327	0.4717	0.5429	0.8119	0.8583
TimesURL	0.1615	0.3443	0.2646	0.3051	0.0663	0.1190
SimMTM	0.1174	0.2352	0.0265	0.1317	0.0178	0.0522
InfoTS	0.1740	0.3563	0.0294	0.1105	0.0161	0.0662
FEI	**0.0631**	**0.1848**	**0.0256**	**0.1099**	**0.0107**	**0.0503**

Table 8: The full results of **end-to-end fine-tuning** on Bearing datasets.

Datasets	Structure	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)
Gesture	**FEI**	75.00	75.00	75.00	72.54
	w/o emb. infer.	63.33 $\downarrow$ 11.67	63.33 $\downarrow$ 11.67	60.54 $\downarrow$ 14.46	60.84 $\downarrow$ 11.70
	w/o mask prompt	38.33 $\downarrow$ 36.67	38.33 $\downarrow$ 36.67	37.65 $\downarrow$ 37.35	37.30 $\downarrow$ 35.24
	w/o momentum	45.00 $\downarrow$ 30.00	45.00 $\downarrow$ 30.00	45.00 $\downarrow$ 30.00	42.71 $\downarrow$ 29.80
	w/o subspace	71.67 $\downarrow$ 3.33	71.67 $\downarrow$ 3.33	68.50 $\downarrow$ 6.50	68.25 $\downarrow$ 4.29
	w/o mask infer.	75.00 –	75.00 –	75.00 –	71.53 $\downarrow$ 1.01
	w/o detach	71.67 $\downarrow$ 3.33	71.67 $\downarrow$ 3.33	67.14 $\downarrow$ 7.86	68.48 $\downarrow$ 4.06
FD-B	**FEI**	67.25	75.09	66.58	69.68
	w/o emb. infer.	34.49 $\downarrow$ 32.76	36.93 $\downarrow$ 38.16	34.17 $\downarrow$ 32.41	31.17 $\downarrow$ 38.51
	w/o mask prompt	37.46 $\downarrow$ 29.79	38.22 $\downarrow$ 36.87	37.94 $\downarrow$ 28.64	30.84 $\downarrow$ 38.84
	w/o momentum	54.28 $\downarrow$ 13.00	57.21 $\downarrow$ 17.90	57.21 $\downarrow$ 9.4	51.14 $\downarrow$ 18.50
	w/o subspace	59.00 $\downarrow$ 8.25	66.69 $\downarrow$ 8.40	56.60 $\downarrow$ 9.98	59.53 $\downarrow$ 10.15
	w/o mask infer.	65.15 $\downarrow$ 2.10	73.15 $\downarrow$ 1.94	65.17 $\downarrow$ 1.41	68.23 $\downarrow$ 1.45
	w/o detach	63.04 $\downarrow$ 4.21	72.06 $\downarrow$ 3.03	64.13 $\downarrow$ 2.45	66.22 $\downarrow$ 3.46
EMG	**FEI**	87.80	90.20	88.65	88.05
	w/o emb. infer.	34.49 $\downarrow$ 32.76	36.93 $\downarrow$ 38.16	34.17 $\downarrow$ 32.41	31.17 $\downarrow$ 38.51
	w/o mask prompt	82.93 $\downarrow$ 4.87	62.75 $\downarrow$ 27.45	55.31 $\downarrow$ 33.34	58.73 $\downarrow$ 29.32
	w/o momentum	82.93 $\downarrow$ 4.90	62.75 $\downarrow$ 27.50	55.80 $\downarrow$ 32.90	58.87 $\downarrow$ 29.20
	w/o subspace	68.29 $\downarrow$ 19.51	51.60 $\downarrow$ 38.60	45.41 $\downarrow$ 43.24	48.25 $\downarrow$ 39.80
	w/o mask infer.	85.37 $\downarrow$ 2.43	79.24 $\downarrow$ 10.96	89.44 $\uparrow$ 0.79	82.63 $\downarrow$ 5.42
	w/o detach	87.80 –	85.70 $\downarrow$ 4.50	91.07 $\uparrow$ 2.42	87.91 $\downarrow$ 0.14
EPI	**FEI**	96.84	95.05	95.00	95.02
	w/o emb. infer.	94.44 $\downarrow$ 2.40	87.86 $\downarrow$ 7.38	94.41 $\downarrow$ 0.59	90.55 $\downarrow$ 4.47
	w/o mask prompt	95.64 $\downarrow$ 1.20	92.80 $\downarrow$ 2.25	93.38 $\downarrow$ 1.62	93.09 $\downarrow$ 1.93
	w/o momentum	95.20 $\downarrow$ 1.60	92.69 $\downarrow$ 2.40	92.69 $\downarrow$ 2.30	92.69 $\downarrow$ 2.30
	w/o subspace	95.15 $\downarrow$ 1.69	92.74 $\downarrow$ 2.31	92.10 $\downarrow$ 2.90	92.42 $\downarrow$ 2.60
	w/o mask infer.	96.04 $\downarrow$ 0.80	93.47 $\downarrow$ 1.58	94.00 $\downarrow$ 1.00	93.73 $\downarrow$ 1.29
	w/o detach	95.12 $\downarrow$ 1.72	94.04 $\downarrow$ 1.01	91.28 $\downarrow$ 3.72	92.57 $\downarrow$ 2.45
HAR	**FEI**	79.54	79.06	81.61	78.30
	w/o emb. infer.	75.06 $\downarrow$ 4.48	74.36 $\downarrow$ 4.70	76.86 $\downarrow$ 4.75	73.90 $\downarrow$ 4.40
	w/o mask prompt	68.44 $\downarrow$ 11.10	67.27 $\downarrow$ 11.79	70.53 $\downarrow$ 11.08	66.54 $\downarrow$ 11.76
	w/o momentum	74.92 $\downarrow$ 4.60	74.08 $\downarrow$ 5.00	76.83 $\downarrow$ 4.80	73.57 $\downarrow$ 4.70
	w/o subspace	78.25 $\downarrow$ 1.29	77.45 $\downarrow$ 1.61	81.26 $\downarrow$ 0.35	76.44 $\downarrow$ 1.86
	w/o mask infer.	78.01 $\downarrow$ 1.53	77.28 $\downarrow$ 1.78	79.83 $\downarrow$ 1.78	76.60 $\downarrow$ 1.70
	w/o detach	77.37 $\downarrow$ 2.17	76.51 $\downarrow$ 2.55	79.37 $\downarrow$ 2.24	75.91 $\downarrow$ 2.39

Table 9: The full results of ablation study.

Datasets	Structure	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)
Gesture	DFM	75.00	75.00	75.00	72.54
	CFM	71.67 $\downarrow 3.33$	71.67 $\downarrow 3.33$	67.68 $\downarrow 7.32$	68.32 $\downarrow 4.22$
	TDM	79.17 $\uparrow$ 4.17	79.17 $\uparrow$ 4.17	77.18 $\uparrow$ 2.18	77.10 $\uparrow$ 4.56
FD-B	DFM	67.25	75.09	66.58	69.68
	CFM	60.65 $\downarrow 6.60$	70.94 $\downarrow 4.15$	59.12 $\downarrow 7.46$	62.65 $\downarrow 7.12$
	TDM	54.02 $\downarrow 13.23$	63.57 $\downarrow 11.52$	50.87 $\downarrow 15.71$	51.51 $\downarrow 18.17$
EMG	DFM	87.80	90.20	88.65	88.05
	CFM	87.80 $-$	80.78 $\downarrow 36.67$	93.06 $\uparrow$ 4.41	84.56 $\downarrow 3.49$
	TDM	73.17 $\downarrow 14.63$	69.85 $\downarrow 20.35$	81.06 $\downarrow 7.59$	73.63 $\downarrow 14.42$
EPI	DFM	96.84	95.05	95.00	95.02
	CFM	96.39 $\downarrow 0.45$	95.23 $\uparrow$ 0.18	93.66 $\downarrow 1.34$	94.42 $\downarrow 0.60$
	TDM	94.26 $\downarrow 2.58$	90.99 $\downarrow 4.06$	90.93 $\downarrow 4.07$	90.06 $\downarrow 4.06$
HAR	DFM	79.54	79.06	81.61	78.30
	CFM	78.59 $\downarrow 0.95$	77.85 $\downarrow 1.21$	80.50 $\downarrow 1.11$	77.38 $\downarrow 0.92$
	TDM	78.62 $\downarrow 0.92$	78.00 $\downarrow 1.06$	79.73 $\downarrow 1.88$	77.31 $\downarrow 0.99$

Table 10: The full results of different masking strategies.

References

Andrzejak, R. G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; and Elger, C. E. 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6): 061907.

Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J. L.; et al. 2013. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, 3.

Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.; LeCun, Y.; and Ballas, N. 2023. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15619–15629.

Chang, C.; Chan, C.-T.; Wang, W.-Y.; Peng, W.-C.; and Chen, T.-F. 2024. TimeDRL: Disentangled Representation Learning for Multivariate Time-Series. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 625–638. IEEE.

Dau, H. A.; Keogh, E.; Kamgar, K.; Yeh, C.-C. M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C. A.; Yanping; Hu, B.; Begum, N.; Bagnall, A.; Mueen, A.; Batista, G.; and Hexagon-ML. 2018. The UCR Time Series Classification Archive. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

Dong, J.; Wu, H.; Zhang, H.; Zhang, L.; Wang, J.; and Long, M. 2023. SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 29996–30025. Curran Associates, Inc.

Goldberger, A. L.; Amaral, L. A.; Glass, L.; Hausdorff, J. M.; Ivanov, P. C.; Mark, R. G.; Mietus, J. E.; Moody, G. B.; Peng, C.-K.; and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23): e215–e220.

Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. 2020. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33: 21271–21284.

Kemp, B.; Zwinderman, A. H.; Tuk, B.; Kamphuisen, H. A.; and Oberye, J. J. 2000. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, 47(9): 1185–1194.

Lessmeier, C.; Kimotho, J. K.; Zimmer, D.; and Sextro, W. 2016. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM Society European Conference*, volume 3.

Liu, J.; and Chen, S. 2024. TimesURL: Self-Supervised Contrastive Learning for Universal Time Series Representation Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(12): 13918–13926.

Liu, J.; Zhong, L.; Wickramasuriya, J.; and Vasudevan, V. 2009. uWave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6): 657–675.

Luo, D.; Cheng, W.; Wang, Y.; Xu, D.; Ni, J.; Yu, W.; Zhang, X.; Liu, Y.; Chen, Y.; Chen, H.; and Zhang, X. 2023. Time Series Contrastive Learning with Information-Aware Augmentations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4): 4534–4542.

Saxena, A.; Goebel, K.; Simon, D.; and Eklund, N. 2008. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management*, 1–9. IEEE.

Wang, B.; Lei, Y.; Li, N.; and Li, N. 2018. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 69(1): 401–412.

Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; and Xu, B. 2022. TS2Vec: Towards Universal Representation of Time Series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(8): 8980–8987.

Zhang, X.; Zhao, Z.; Tsiligkaridis, T.; and Zitnik, M. 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems*, 35: 3988–4003.