Detailed Proofs

Properties of Data Augmentations that Preserve Pseudo Labels.

We assume that v_i and v_j are two augmented instances of inputs x_i and x_j, respectively. Preserving pseudo labels defined in one-hot encoding requires that the map between variable x and v is one-to-many. Formally, $x_i \neq x_j \rightarrow v_i \neq v_j$. This can be proved by contradiction. If we have a pair of (i, j) that $i \neq j$ and $v_i = v_j$, then we have $f_{\boldsymbol{\theta}}(v_i) = f_{\boldsymbol{\theta}}(v_j)$, showing that augmentations cannot preserve pseudo labels.

Property 1 *(Preserving Fidelity). If augmentation* v *preserves the one-hot encoding pseudo label, the mutual information between* v *and downstream task label* y *(although not visible to training) is equivalent to that between raw input* x *and* y*, i.e., MI*$(\mathsf{v};\mathsf{y}) =$ *MI*$(\mathsf{x};\mathsf{y})$.

Proof. From the definition of mutual information, we have

$$\begin{aligned}\text{MI}(\mathsf{v};\mathsf{y}) &= H(\mathsf{y}) - H(\mathsf{y}|\mathsf{v}) \\ &= H(\mathsf{y}) + \sum_{v,y} p(v,y)\log\frac{p(v,y)}{p(v)} \\ &= H(\mathsf{y}) + \sum_{x,y}\sum_{v\in\mathbb{V}(x)} p(v,y)\log\frac{p(v,y)}{p(v)},\end{aligned}$$

where $\mathbb{V}(x)$ is the set of augmented instances of a time series instance x. In the unsupervised setting where the ground-truth label y is unknown, we assume that the augmentation v is a (probabilistic) function of x only. The only qualifier means $p(v|x,y) = p(v|x)$. Since the mapping from x to v is one-to-many. For each $v \in \mathbb{V}(x)$ we have $p(v,y) = p(v,x,y)$ and $p(v) = p(v|x)p(x)$. Thus, we have

$$\begin{aligned}\frac{p(v,y)}{p(v)} &= \frac{p(v,x,y)}{p(v|x)p(x)} = \frac{p(v|x,y)p(x,y)}{p(v|x)p(x)} \\ &= \frac{p(v|x)p(x,y)}{p(v|x)p(x)} = \frac{p(x,y)}{p(x)} \\ \text{MI}(\mathsf{v};\mathsf{y}) &= H(\mathsf{y}) + \sum_{x,y}\sum_{v\in\mathbb{V}(x)} p(v,y)\log\frac{p(x,y)}{p(x)} \\ &= H(\mathsf{y}) + \sum_{x,y}[\sum_{v\in\mathbb{V}(x)} p(v,y)]\log\frac{p(x,y)}{p(x)} \\ &= H(\mathsf{y}) + \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)} \\ &= \text{MI}(\mathsf{x};\mathsf{y}).\end{aligned}$$

Property 2 *(Adding New Information). By preserving the one-hot encoding pseudo label, augmentation* v *contains new information comparing to the raw input* x*, i.e.,* $H(\mathsf{v}) \geq H(\mathsf{x})$.

In information theory, entropy describes the amount of information of a random variable. For simplicity, we assume a finite number of augmented instances for each input, and each augmented instance is generated independently. Then, we have $p(x) = \sum_{v\in\mathbb{V}(x)} p(v)$. Then we have that the entropy of variable x is no larger than the entropy of v.

$$\begin{aligned}H(\mathsf{x}) &= -\sum_x p(x)\log p(x) \\ &= -\sum_x[\sum_{v\in\mathbb{V}(x)} p(v)]\log[\sum_{v\in\mathbb{V}(x)} p(v)] \\ &= -\sum_x\sum_{v\in\mathbb{V}(x)} p(v)\log[\sum_{v\in\mathbb{V}(x)} p(v)] \\ &\leq -\sum_x\sum_{v\in\mathbb{V}(x)} p(v)\log p(v) \\ &= -\sum_v p(v)\log p(v) = H(\mathsf{v})\end{aligned}$$

Rationality of Approximation of Bernoulli Distribution with Binary Concrete Distribution in Eq. (10).

In the binary concrete distribution, parameter τ controls the temperature that achieves the trade-off between binary output and continuous optimization. When $\tau \rightarrow 0$, we have $\lim_{\tau\rightarrow 0} P(a_i = 1) = p_i$, which is equivalent to the Bernoulli distribution.

Proof.

$$\begin{aligned}&\lim_{\tau\rightarrow 0} P(a_i = 1) \\ &= \lim_{\tau\rightarrow 0} P(\sigma((\log\epsilon - \log(1-\epsilon) + \log\frac{p_i}{1-p_i})/\tau) = 1) \\ &= P(\log\epsilon - \log(1-\epsilon) + \log\frac{p_i}{1-p_i} > 0) \\ &= P(\log\frac{\epsilon}{1-\epsilon} - \log\frac{1-p_i}{p_i} > 0) \\ &= P(\frac{\epsilon}{1-\epsilon} > \frac{1-p_i}{p_i})\end{aligned}$$

Since ϵ, and p_i are both in $(0, 1)$, and function $\frac{x}{1-x}$ are monotonically increasing in this region. Thus, we have

$$\lim_{\tau\rightarrow 0} P(a_i = 1) = P(\epsilon > 1 - p_i) = p_i$$

Algorithms

Training Algorithm

The training algorithm of InfoTS under both supervised and unsupervised settings is described in Algorithm 1. We first randomly initiate parameters in the encoder, meta-learner network, and classifier (line 2). Given a batch of training instances $\mathbb{X}_B \subseteq \mathbb{X}$, for each candidate transformation t_i, we utilize binary concrete distribution to get parameters a_i (lines 6-7), which indicates whether the transformation should be applied (line 8). $\mathbb{V}_B^{(i)}$ denotes the batch of augmented instances generated from transformation function t_i. The final augmented instances are generated by adaptively considering all candidates transformations (line 10). Parameters $\boldsymbol{\theta}$ in the encoder are updated by minimizing the contrastive objective (lines 11-13). Meta-learner network is then optimized with information-aware criteria (lines 14-16). Then, classifier $h_{\boldsymbol{w}}$ is optimized with the classification objective (line 17).

Algorithm 1: Algorithm for InfoTS in both supervised and unsupervised settings

Input: time series dataset $\mathbb{X}$, the label set $\mathbb{Y}$ (supervised setting), a set of candidate transformations $\mathbb{T}$, hyper-parameters α and β,
Initialize the encoder $f_{\boldsymbol{\theta}}$, parameters in meta-learner network $\{q_i\}_{i=1}^{|\mathbb{T}|}$, and the classifier $h_{\boldsymbol{w}}$.
for each epoch **do**
  for each training batch $\mathbb{X}_B \subseteq \mathbb{X}$ **do**
    for each transformation $t_i \in \mathbb{T}$ **do**
      $p_i \leftarrow \sigma(q_i)$
      $a_i \leftarrow$ binaryConcrete(p_i)
      $\mathbb{V}_B^{(i)} \leftarrow$ transform each $x \in \mathbb{X}_B$ with Eq. (9)
    end for
    Get $\mathbb{V}_B$ by averaging $\{\mathbb{V}_B^{(i)}\}_{i=1}^{|\mathbb{T}|}$.
    Compute global contrastive loss $\mathcal{L}_g$ with Eq. (6)
    Compute local contrastive loss $\mathcal{L}_c$ with Eq. (7)
    Update parameters $\boldsymbol{\theta}$ in the encoder with Eq. (8)
    Compute fidelity loss with Eq. (2)
    Compute variety loss with Eq. (4)
    Update parameters $\{q_i\}_{i=1}^{|\mathbb{T}|}$ in the meta-learner network with Eq. (5)
    Update parameters $\boldsymbol{w}$ in the classifier $h_{\boldsymbol{w}}$ with the cross-entropy loss
  end for
end for

Implementation of Local-Wise Contrastive

Local-wise contrastive loss aims to capture the intra-temporal relations in each time series instance. For an augmented instance v, we first split it into multiple subsequences, as shown in Figure 4. Each subsequence has length L. For each subsequence s, the neighboring subsequences within window size 1 are considered as positive samples. If s locates at the end of v, then we choose the subsequence in front of s as the positive pair p. Otherwise, we choose the subsequence following s instead. Subsequences out of window size 1 are considered as negative samples.

Experimental Settings

Data Augmentations

We follow (Fan, Zhang, and Gao 2020) to set up candidate data augmentations, including jittering, scaling, cutout, time warping, window slicing, window warping and subsequence augmentation (Franceschi, Dieuleveut, and Jaggi 2019). Detailed descriptions are listed as follows.

- *Jittering* augmentation adds the random noise sampled from a Gaussian distribution $\mathcal{N}(0, 0.3)$ to the input time series.
- *Scaling* augmentation multiplies the input time series by a scaling factor sampled from a Gaussian distribution $\mathcal{N}(0, 0.5)$.
- *Cutout* operation replaces features of 10% randomly sampled time stamps of the input with zeros.
- *Time warping* random changes the speed of the timeline[1]. The number of speed changes is 100 and the maximal ratio of max/min speed is 10. If necessary, over-sampling or sampling methods are adopted to ensure the length of the augmented instance is the same as the original one.
- *Window slicing* randomly crops half the input time series and then linearly interpolates it back to the original length (Le Guennec, Malinowski, and Tavenard 2016).
- *Window warping* first randomly selects 30% of the input time series along the timeline and then warps the time dimension by 0.5 or 2. Finally, we adopt linear interpolation to transform it back to the original length (Le Guennec, Malinowski, and Tavenard 2016).
- *Subsequence* operation random selects a subsequence from the input time series (Yue et al. 2021).

With the first 100 time stamps in the univariate Electricity dataset as an example, we visualize the original time series and the augmented ones in Figure 5.

Hardware and Implementations

All experiments are conducted on a Linux machine with 4 NVIDIA GeForce RTX 2080 Ti GPUs, each with 11GB memory. CUDA version is 10.1 and Driver Version is 418.56. Our method InfoTS is implemented with Python 3.7.7 and Pytorch 1.7.1.

Hyperparameters

We train and evaluate our methods with the following hyperparameters and configurations.

- Optimizer: Adam optimizer (Kingma and Ba 2014) with learning rate and decay rates setting to 0.001 and (0.9,0.999), respectively.
- SVM: scikit-learn implementation (Pedregosa et al. 2011) with penalty $C \in \{10^i | i \in [-4, 4] \cup \infty\}$ (Franceschi, Dieuleveut, and Jaggi 2019).
- Encoder architecture: We follow (Yue et al. 2021) to design the encoder. Specifically, the output dimension of the linear projection layer is set to 64, the same for the number of channels in the following dilated CNN module. In the CNN module, GELU (Hendrycks and Gimpel 2016) is adopted as the activation function, and the kernel size is set to 3. The dilation is set to 2^i in the i-the block.
- Classifier architecture: a fully connected layer that maps the representations to the label is adopted.
- Trade-off hyperparameters: β in Eq. (5) and α in Eq. (8) are searched in $[0.1, 0.5, 1.0, 5, 10]$. Parameter sensitivity studies are shown in Section .
- Temperature in binary concrete distribution: we follow the practice in (Jang, Gu, and Poole 2016) to adopt the strategy by starting the training with a high temperature 2.0, and anneal to a small value 0.1, with a guided schedule.

More Experimental Results

Parameter Sensitivity Studies

In this part, we adopt the electricity dataset to analyze the effects of two important hyper-parameters in our method InfoTS. The hyperparameter α in Eq. (8) controls the trade-off

[1] https://tsaug.readthedocs.io/

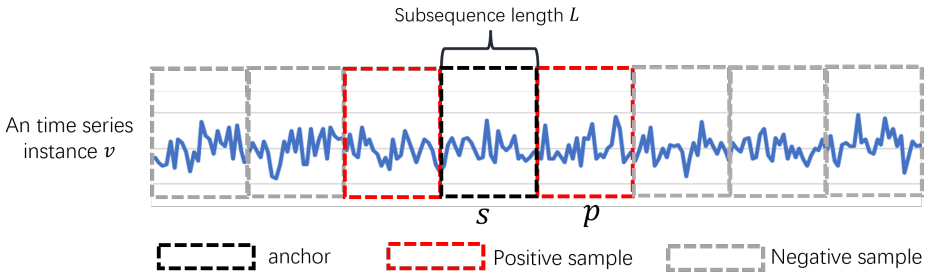
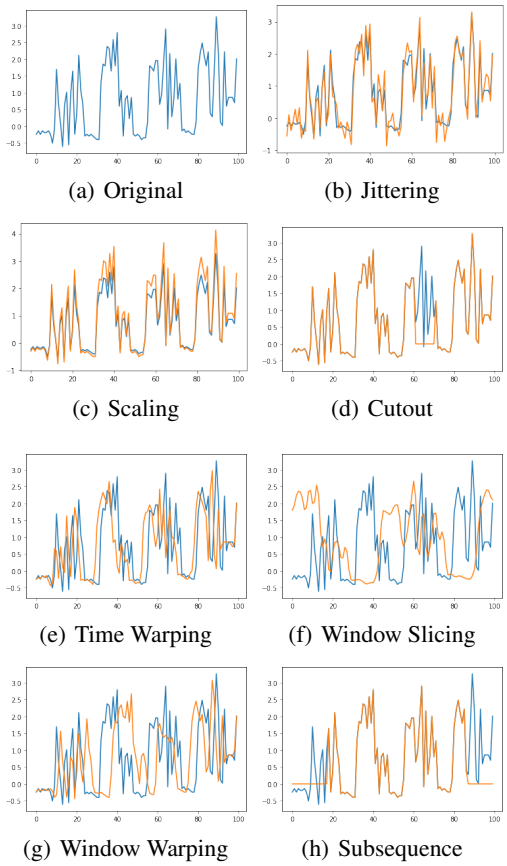


Figure 4: Positive and negative samples for a subsequence s.

Table 4: Multivariate time series forecasting results.

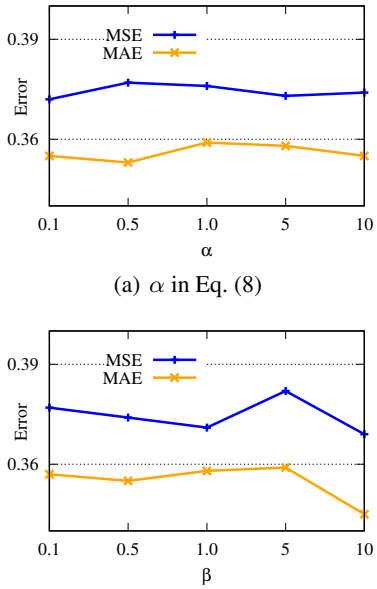
		InfoTS		TS2Vec		Informer		StemGNN		TCN		LogTrans		LSTnet	
Dataset	L_y	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh_1$	24	**0.564**	**0.520**	0.599	0.534	0.577	0.549	0.614	0.571	0.767	0.612	0.686	0.604	1.293	0.901
	48	**0.607**	**0.553**	0.629	0.555	0.685	0.625	0.748	0.618	0.713	0.617	0.766	0.757	1.456	0.960
	168	0.746	0.638	0.755	0.636	0.931	0.752	**0.663**	**0.608**	0.995	0.738	1.002	0.846	1.997	1.214
	336	**0.904**	0.722	0.907	**0.717**	1.128	0.873	0.927	0.730	1.175	0.800	1.362	0.952	2.655	1.369
	720	1.098	0.811	**1.048**	**0.790**	1.215	0.896	–	–	1.453	1.311	1.397	1.291	2.143	1.380
$ETTh_2$	24	**0.383**	0.462	0.398	**0.461**	0.720	0.665	1.292	0.883	1.365	0.888	0.828	0.750	2.742	1.457
	48	**0.567**	0.582	0.578	**0.573**	1.457	1.001	1.099	0.847	1.395	0.960	1.806	1.034	3.567	1.687
	168	**1.789**	**1.048**	1.901	1.065	3.489	1.515	2.282	1.228	3.166	1.407	4.070	1.681	3.242	2.513
	336	**2.120**	**1.161**	2.304	1.215	2.723	1.340	3.086	1.351	3.256	1.481	3.875	1.763	2.544	2.591
	720	**2.511**	**1.316**	2.650	1.373	3.467	1.473	–	–	3.690	1.588	3.913	1.552	4.625	3.709
$ETTm_1$	24	0.391	0.408	0.443	0.436	**0.323**	**0.369**	0.620	0.570	0.324	0.374	0.419	0.412	1.968	1.170
	48	0.503	0.475	0.582	0.515	0.494	0.503	0.744	0.628	**0.477**	**0.450**	0.507	0.583	1.999	1.215
	96	**0.537**	**0.503**	0.622	0.549	0.678	0.614	0.709	0.624	0.636	0.602	0.768	0.792	2.762	1.542
	288	**0.653**	**0.579**	0.709	0.609	1.056	0.786	0.843	0.683	1.270	1.351	1.462	1.320	1.257	2.076
	672	**0.757**	**0.642**	0.786	0.655	1.192	0.926	–	–	1.381	1.467	1.669	1.461	1.917	2.941
Electricity	24	**0.255**	**0.350**	0.287	0.374	0.312	0.387	0.439	0.388	0.305	0.384	0.297	0.374	0.356	0.419
	48	**0.279**	**0.368**	0.307	0.388	0.392	0.431	0.413	0.455	0.317	0.392	0.316	0.389	0.429	0.456
	168	**0.302**	**0.385**	0.332	0.407	0.515	0.509	0.506	0.518	0.358	0.423	0.426	0.466	0.372	0.425
	336	**0.320**	**0.399**	0.349	0.420	0.759	0.625	0.647	0.596	0.349	0.416	0.365	0.417	0.352	0.409
Avg.		**0.805**	**0.627**	0.852	0.645	1.164	0.781	0.977	0.706	1.243	0.854	1.402	1.032	1.836	1.374



(a) Original (b) Jittering (c) Scaling (d) Cutout (e) Time Warping (f) Window Slicing (g) Window Warping (h) Subsequence

Figure 5: Examples of candidate augmentations on Electricity univariate dataset. Blue lines are the original time series data and orange ones are augmented instances.

between local and global contrastive losses when training the encoder. β in Eq. (5) achieves the balance between high variety and high fidelity when training the meta-learner network. We tune these parameters in range $[0.1, 0.5, 1.0, 5, 10]$ and show the results in Figure 6. These figures show that our method achieves high performance with a wide range of selections, demonstrating the robustness of the proposed method. In general, setting trade-off parameters to 0.5 or 1 achieves good performance.



(a) α in Eq. (8)

(b) β in Eq. (5)

Figure 6: Parameter sensitivity studies.

Case Study on Signal Detection

In this part, we show the potential usage of InfoTS to detect the informative signals in the time series. We adopt the CricketX dataset as an example for the case study. Subsequence augmentations on 0-100, 100-200, and 200-300 periods are adopted as candidate transformations. We observe that the one operated on 100-200 period has high fidelity and variety, leading to better accuracy performance, which is consistent with the visualization results in Figure 7.

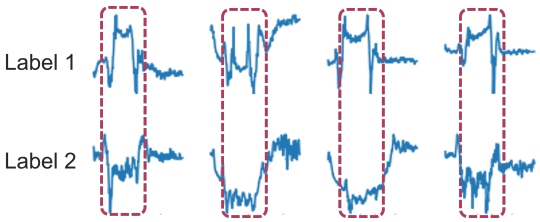


Figure 7: The informative signals locate in the middle periods of time series in the CricketX dataset.

Table 5: Effectiveness of each candidate data transformation on Electricity.

	InfoTS		Cutout		Jittering		Scaling		Time Warp		Window Slice		Window Warp		Subsequence	
L_y	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	**0.245**	**0.269**	0.254	0.277	0.251	0.275	0.252	0.273	0.251	0.274	0.258	0.280	0.253	0.277	0.248	0.273
48	**0.294**	**0.301**	0.304	0.309	0.297	0.302	0.302	0.307	0.305	0.309	0.310	0.314	0.307	0.310	0.295	**0.301**
168	**0.402**	**0.367**	0.412	0.381	0.403	0.373	0.407	0.377	0.415	0.382	0.415	0.381	0.416	0.382	0.405	0.372
336	**0.533**	**0.453**	0.555	0.465	0.545	0.458	0.552	0.461	0.555	0.469	0.551	0.470	0.554	0.466	0.546	0.456
Avg.	**0.369**	**0.348**	0.381	0.358	0.374	0.352	0.377	0.354	0.381	0.359	0.383	0.361	0.383	0.359	0.374	0.350

Table 6: Effectiveness of each candidate data transformation on ETTh1.

	InfoTS		Cutout		Jittering		Scaling		Time Warp		Window Slice		Window Warp		Subsequence	
L_y	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	**0.039**	0.149	0.045	0.158	0.045	0.160	**0.039**	**0.148**	0.043	0.155	0.041	0.151	0.043	0.156	0.045	0.161
48	**0.056**	**0.179**	0.061	0.185	0.062	0.188	0.060	0.185	0.061	0.186	0.064	0.190	0.063	0.191	0.063	0.188
168	**0.100**	**0.239**	0.110	0.251	0.115	0.261	0.111	0.255	0.111	0.253	0.118	0.265	0.118	0.265	0.125	0.271
336	**0.117**	**0.264**	0.136	0.287	0.127	0.278	0.130	0.281	0.148	0.302	0.133	0.288	0.146	0.303	0.139	0.291
720	**0.141**	**0.302**	0.167	0.330	0.143	0.304	0.155	0.318	0.168	0.331	0.151	0.315	0.147	0.308	0.153	0.314
Avg.	**0.091**	**0.227**	0.104	0.242	0.098	0.238	0.099	0.237	0.106	0.246	0.101	0.242	0.103	0.245	0.105	0.245

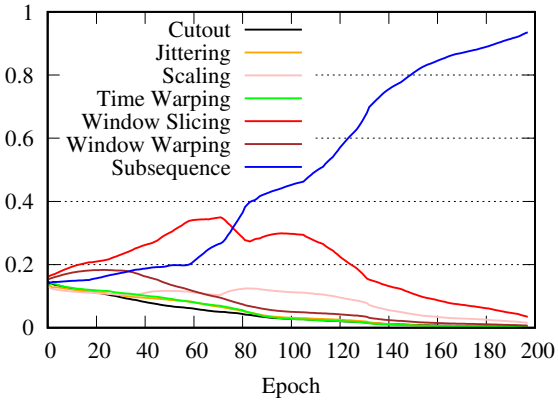
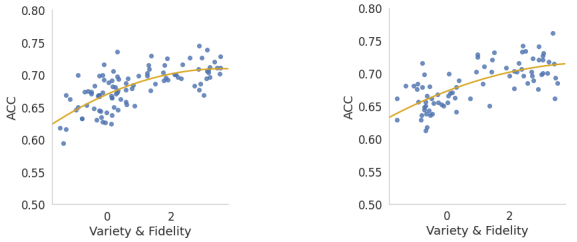


Figure 8: Weight updating process of meta-learner network in InfoTS.



(a) Supervised Setting (b) Unsupervised Setting

Figure 9: Evaluation of the criteria.

Updating Process of InfoTS

To show that our InfoTS can adaptively detect the most effective augmentation based on the data distribution, we conduct more ablation studies to investigate comprehensively into the proposed model. We compare performances of variants that each applies a single transformation to generate augmented instances in Table 5. From the table, we know that augmentation with subsequence benefits the most for the Electricity dataset. We visualize the weight updating process of InfoTS in Figure 8, with each line representing the normalized importance score of the corresponding transformation. The weight for subsequence increase with the epoch, showing that InfoTS tends to adopt subsequence as the optimal transformation. Consistency between accuracy performance and weight updating process demonstrates the effectiveness of InfoTS to adaptively select feasible transformations. Besides, as shown in Table 5, InfoTS outperforms the variant that uses subsequence only.

Note that although subsequence transformation works well for the Electricity dataset, it may generate uninformative augmented instances for other datasets. Table 6 shows the performances of each single transformation in the ETTh1 dataset. The subsequence transformation is no more effective than other candidate transformations. Besides, guided by the information-aware criteria, our method InfoTS can still outperform other variants. This comparison shows that the meta-learner network learns to consider the combinations, which is better than any (single) candidate augmentation.

Evaluation of The Criteria with Time series Classification

To empirically verify the effectiveness of the proposed criteria with Time series Classification in both supervised and unsupervised setting, we adopt the dataset CricketY from the UCR archive (Dau et al. 2019; Fan, Zhang, and Gao 2020), and conduct augmentations with different configurations. For each configuration, we calculate the criteria score and the corresponding classification accuracy within the setting in Section . As shown in Figure 9, in general, accuracy performance is positively related to the proposed criteria in both supervised and unsupervised settings, the results are consistent with the conclusion drawn from forecasting performances.

More Ablation Studies

To further check the effectiveness of the meta-learner network on automatically selecting suitable augmentations, we adopt another state-of-the-art baseline, TS2Vec as the backbone (Yue et al. 2021). We denote this variant as TS2Vec+Infoadpative, where the contrastive loss in TS2Vec is adopted to train the encoder and the proposed information-aware criteria are used to train the meta-learner network. The performances of the original TS2Vec and TS2Vec+Infoadpative on the Electricity dataset are shown in Table 7. The comparison shows that with information-aware adaptive augmentation, we can also consistently and significantly improve the performances of TS2vec.

Table 7: Effectiveness of meta-learner network with TS2Vec as the backbone.

	TS2vec		TS2Vec+Infoadpative	
L_y	MSE	MAE	MSE	MAE
24	0.260	0.288	0.250	0.273
48	0.319	0.324	0.298	0.302
168	0.427	0.394	0.411	0.372
336	0.565	0.474	0.561	0.463
Avg.	0.393	0.370	0.380	0.352

Full Results of Time Series Forecasting and Classification

The full results of multivariate time series forecasting are shown in Tabel 4. Results of StemGNN with $L_y = 720$ are not available due to the out-of-memory error (Yue et al. 2021). Full results of univariate time series classification on 128 UCR datasets are shown in Tabel 8. Results of T-Loss, TS-TCC, and TNC are not reported on several datasets because they are not able to deal with missing observations in time series data. These unavailable accuracy scores are dismissed when computing average accuracy and considered as 0 when calculating the average rank. Results of multivariate classification on 30 UEA datasets are listed in Tabel 9. Computations of average accuracy scores and ranks follow the ones in 128 UCR datasets.

	InfoTS$_s$	InfoTS	TS2Vec	T-Loss	TNC	TS-TCC	TST	DTW
Adiac	**0.795**	0.788	0.775	0.675	0.726	0.767	0.550	0.604
ArrowHead	**0.874**	**0.874**	0.857	0.766	0.703	0.737	0.771	0.703
Beef	**0.900**	0.833	0.767	0.667	0.733	0.600	0.500	0.633
BeetleFly	0.950	0.950	0.900	0.800	0.850	0.800	**1.000**	0.700
BirdChicken	0.850	**0.900**	0.800	0.850	0.750	0.650	0.650	0.750
Car	**0.900**	0.883	0.883	0.833	0.683	0.583	0.550	0.733
CBF	**1.000**	0.999	**1.000**	0.983	0.983	0.998	0.898	0.997
ChlorineConcentration	0.825	0.822	**0.832**	0.749	0.760	0.753	0.562	0.648
CinCECGTorso	0.896	**0.928**	0.827	0.713	0.669	0.671	0.508	0.651
Coffee	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.821	**1.000**
Computers	0.720	**0.748**	0.660	0.664	0.684	0.704	0.696	0.700
CricketX	0.780	0.774	**0.805**	0.713	0.623	0.731	0.385	0.754
CricketY	**0.774**	**0.774**	0.769	0.728	0.597	0.718	0.467	0.744
CricketZ	**0.792**	0.787	**0.792**	0.708	0.682	0.713	0.403	0.754
DiatomSizeReduction	**0.997**	**0.997**	0.987	0.984	0.993	0.977	0.961	0.967
DistalPhalanxOutlineCorrect	**0.808**	0.801	0.775	0.775	0.754	0.754	0.728	0.717
DistalPhalanxOutlineAgeGroup	0.763	0.763	0.727	0.727	0.741	0.755	0.741	**0.770**
DistalPhalanxTW	0.720	**0.727**	0.698	0.676	0.669	0.676	0.568	0.590
Earthquakes	**0.821**	**0.821**	0.748	0.748	0.748	0.748	0.748	0.719
ECG200	**0.950**	0.930	0.920	0.940	0.830	0.880	0.830	0.770
ECG5000	**0.945**	**0.945**	0.935	0.933	0.937	0.941	0.928	0.924
ECGFiveDays	**1.000**	**1.000**	**1.000**	**1.000**	0.999	0.878	0.763	0.768
ElectricDevices	0.691	0.702	**0.721**	0.707	0.700	0.686	0.676	0.602
FaceAll	**0.929**	**0.929**	0.805	0.786	0.766	0.813	0.504	0.808
FaceFour	0.864	0.818	**0.932**	0.920	0.659	0.773	0.511	0.830
FacesUCR	0.917	0.913	**0.930**	0.884	0.789	0.863	0.543	0.905
FiftyWords	**0.809**	0.793	0.774	0.732	0.653	0.653	0.525	0.690
Fish	**0.949**	0.937	0.937	0.891	0.817	0.817	0.720	0.920
FordA	0.925	0.915	**0.948**	0.928	0.902	0.930	0.568	0.555
FordB	0.795	0.785	0.807	0.793	0.733	**0.815**	0.507	0.620
GunPoint	**1.000**	**1.000**	0.987	0.980	0.967	0.993	0.827	0.907
Ham	**0.848**	0.838	0.724	0.724	0.752	0.743	0.524	0.467
HandOutlines	**0.946**	**0.946**	0.930	0.922	0.930	0.724	0.735	0.881
Haptics	0.545	**0.546**	0.536	0.490	0.474	0.396	0.357	0.377
Herring	**0.703**	0.656	0.641	0.594	0.594	0.594	0.594	0.531
InlineSkate	0.420	**0.424**	0.415	0.371	0.378	0.347	0.287	0.384
InsectWingbeatSound	**0.664**	0.639	0.630	0.597	0.549	0.415	0.266	0.355
ItalyPowerDemand	**0.971**	0.966	0.961	0.954	0.928	0.955	0.845	0.950
LargeKitchenAppliances	0.851	0.853	**0.875**	0.789	0.776	0.848	0.595	0.795
Lightning2	**0.934**	**0.934**	0.869	0.869	0.869	0.836	0.705	0.869
Lightning7	0.863	**0.877**	0.863	0.795	0.767	0.685	0.411	0.726
Mallat	0.967	**0.974**	0.915	0.951	0.871	0.922	0.713	0.934
Meat	**0.967**	**0.967**	**0.967**	0.950	0.917	0.883	0.900	0.933
MedicalImages	**0.920**	0.820	0.793	0.750	0.754	0.747	0.632	0.737
MiddlePhalanxOutlineCorrect	**0.859**	**0.859**	0.838	0.825	0.818	0.818	0.753	0.698
MiddlePhalanxOutlineAgeGroup	**0.662**	**0.662**	0.636	0.656	0.643	0.630	0.617	0.500
MiddlePhalanxTW	**0.636**	0.617	0.591	0.591	0.571	0.610	0.506	0.506
MoteStrain	**0.873**	**0.873**	0.863	0.851	0.825	0.843	0.768	0.835
NonInvasiveFetalECGThorax1	**0.941**	**0.941**	0.930	0.878	0.898	0.898	0.471	0.790
NonInvasiveFetalECGThorax2	0.943	**0.944**	0.940	0.919	0.912	0.913	0.832	0.865
OliveOil	**0.933**	**0.933**	0.900	0.867	0.833	0.800	0.800	0.833
OSULeaf	0.760	0.760	**0.876**	0.760	0.723	0.723	0.545	0.591
PhalangesOutlinesCorrect	**0.826**	**0.826**	0.823	0.784	0.787	0.804	0.773	0.728
Phoneme	0.272	0.281	**0.312**	0.276	0.180	0.242	0.139	0.228
Plane	**1.000**	**1.000**	**1.000**	0.990	**1.000**	**1.000**	0.933	**1.000**
ProximalPhalanxOutlineCorrect	0.924	**0.927**	0.900	0.859	0.866	0.873	0.770	0.784
ProximalPhalanxOutlineAgeGroup	**0.883**	**0.883**	0.844	0.844	0.854	0.839	0.854	0.805
ProximalPhalanxTW	**0.849**	0.844	0.824	0.771	0.810	0.800	0.780	0.761
RefrigerationDevices	**0.624**	**0.624**	0.589	0.515	0.565	0.563	0.483	0.464
ScreenType	**0.510**	0.493	0.411	0.416	0.509	0.419	0.419	0.397
ShapeletSim	0.856	0.856	**1.000**	0.672	0.589	0.683	0.489	0.650
ShapesAll	0.855	0.852	**0.905**	0.848	0.788	0.773	0.733	0.768
SmallKitchenAppliances	**0.773**	**0.773**	0.733	0.677	0.725	0.691	0.592	0.643

	InfoTS$_s$	InfoTS	TS2Vec	T-Loss	TNC	TS-TCC	TST	DTW
SonyAIBORobotSurface1	0.921	**0.927**	0.903	0.902	0.804	0.899	0.724	0.725
SonyAIBORobotSurface2	**0.953**	**0.953**	0.890	0.889	0.834	0.907	0.745	0.831
StarLightCurves	**0.973**	**0.973**	0.971	0.964	0.968	0.967	0.949	0.907
Strawberry	**0.978**	**0.978**	0.965	0.954	0.951	0.965	0.916	0.941
SwedishLeaf	**0.954**	0.950	0.942	0.914	0.880	0.923	0.738	0.792
Symbols	**0.979**	**0.979**	0.976	0.963	0.885	0.916	0.786	0.950
SyntheticControl	**1.000**	**1.000**	0.997	0.987	**1.000**	0.990	0.490	0.993
ToeSegmentation1	0.930	0.934	**0.947**	0.939	0.864	0.930	0.807	0.772
ToeSegmentation2	**0.923**	0.915	0.915	0.900	0.831	0.877	0.615	0.838
Trace	**1.000**	**1.000**	**1.000**	0.990	**1.000**	**1.000**	**1.000**	**1.000**
TwoLeadECG	**0.999**	0.998	0.987	**0.999**	0.993	0.976	0.871	0.905
TwoPatterns	**1.000**	**1.000**	**1.000**	0.999	**1.000**	0.999	0.466	**1.000**
UWaveGestureLibraryX	**0.820**	0.819	0.810	0.785	0.781	0.733	0.569	0.728
UWaveGestureLibraryY	**0.745**	0.736	0.729	0.710	0.697	0.641	0.348	0.634
UWaveGestureLibraryZ	0.768	0.768	**0.770**	0.757	0.721	0.690	0.655	0.658
UWaveGestureLibraryAll	0.966	**0.967**	0.934	0.896	0.903	0.692	0.475	0.892
Wafer	**0.999**	0.998	0.998	0.992	0.994	0.994	0.991	0.980
Wine	**0.963**	**0.963**	0.889	0.815	0.759	0.778	0.500	0.574
WordSynonyms	**0.715**	0.704	0.704	0.691	0.630	0.531	0.422	0.649
Worms	**0.766**	0.753	0.701	0.727	0.623	0.753	0.455	0.584
WormsTwoClass	0.818	**0.857**	0.805	0.792	0.727	0.753	0.584	0.623
Yoga	**0.937**	0.869	0.887	0.837	0.812	0.791	0.830	0.837
ACSF1	0.850	0.850	**0.910**	0.900	0.730	0.730	0.760	0.640
AllGestureWiimoteX	0.560	0.630	**0.777**	0.763	0.703	0.697	0.259	0.716
AllGestureWiimoteY	0.623	0.686	**0.793**	0.726	0.699	0.741	0.423	0.729
AllGestureWiimoteZ	0.633	0.629	**0.770**	0.723	0.646	0.689	0.447	0.643
BME	**1.000**	**1.000**	0.993	0.993	0.973	0.933	0.760	0.900
Chinatown	0.985	**0.988**	0.968	0.951	0.977	0.983	0.936	0.957
Crop	**0.766**	**0.766**	0.756	0.722	0.738	0.742	0.710	0.665
EOGHorizontalSignal	0.577	0.572	0.544	**0.605**	0.442	0.401	0.373	0.503
EOGVerticalSignal	0.459	0.459	**0.503**	0.434	0.392	0.376	0.298	0.448
EthanolLevel	0.710	**0.712**	0.484	0.382	0.424	0.486	0.260	0.276
FreezerRegularTrain	**0.998**	0.996	0.986	0.956	0.991	0.989	0.922	0.899
FreezerSmallTrain	**0.991**	0.988	0.894	0.933	0.982	0.979	0.920	0.753
Fungi	0.866	0.946	0.962	**1.000**	0.527	0.753	0.366	0.839
GestureMidAirD1	0.592	0.592	**0.631**	0.608	0.431	0.369	0.208	0.569
GestureMidAirD2	0.459	0.492	0.515	0.546	0.362	0.254	0.138	**0.608**
GestureMidAirD3	0.323	0.315	**0.346**	0.285	0.292	0.177	0.154	0.323
GesturePebbleZ1	0.895	0.802	**0.930**	0.919	0.378	0.395	0.500	0.791
GesturePebbleZ2	**0.905**	0.842	0.873	0.899	0.316	0.430	0.380	0.671
GunPointAgeSpan	0.997	**1.000**	0.994	0.994	0.984	0.994	0.991	0.918
GunPointMaleVersusFemale	**1.000**	**1.000**	**1.000**	0.997	0.994	0.997	**1.000**	0.997
GunPointOldVersusYoung	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.838
HouseTwenty	**0.941**	0.924	**0.941**	0.933	0.782	0.790	0.815	0.924
InsectEPGRegularTrain	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.872
InsectEPGSmallTrain	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.735
MelbournePedestrian	**0.964**	0.962	0.959	0.944	0.942	0.949	0.741	0.791
MixedShapesRegularTrain	**0.940**	0.935	0.922	0.905	0.911	0.855	0.879	0.842
MixedShapesSmallTrain	**0.892**	0.887	0.881	0.860	0.813	0.735	0.828	0.780
PickupGestureWiimoteZ	**0.820**	**0.820**	**0.820**	0.740	0.620	0.600	0.240	0.660
PigAirwayPressure	0.433	0.432	**0.683**	0.510	0.413	0.380	0.120	0.106
PigArtPressure	0.820	0.830	**0.966**	0.928	0.808	0.524	0.774	0.245
PigCVP	0.654	0.653	**0.870**	0.788	0.649	0.615	0.596	0.154
PLAID	0.356	0.355	0.561	0.555	0.495	0.445	0.419	**0.840**
PowerCons	0.995	**1.000**	0.972	0.900	0.933	0.961	0.911	0.878
Rock	**0.760**	**0.760**	0.700	0.580	0.580	0.600	0.680	0.600
SemgHandGenderCh2	0.939	0.944	**0.963**	0.890	0.882	0.837	0.725	0.802
SemgHandMovementCh2	0.833	0.836	**0.893**	0.789	0.593	0.613	0.420	0.584
SemgHandSubjectCh2	0.945	0.924	**0.951**	0.853	0.771	0.753	0.484	0.727
ShakeGestureWiimoteZ	0.920	0.920	**0.940**	0.920	0.820	0.860	0.760	0.860
SmoothSubspace	**1.000**	**1.000**	0.993	0.960	0.913	0.953	0.827	0.827
UMD	**1.000**	**1.000**	**1.000**	0.993	0.993	0.986	0.910	0.993
DodgerLoopDay	**0.675**	**0.675**	0.562	–	–	–	0.200	0.500

	InfoTS$_s$	InfoTS	TS2Vec	T-Loss	TNC	TS-TCC	TST	DTW
DodgerLoopGame	**0.971**	0.942	0.841	–	–	–	0.696	0.877
DodgerLoopWeekend	**0.986**	**0.986**	0.964	–	–	–	0.732	0.949
AVG	**0.841**	0.838	0.836	0.806	0.761	0.757	0.639	0.729
Rank	**1.757**	1.969	2.328	3.640	4.508	4.383	6.117	5.125

Table 8: Full results of univariate time series classification on 128 UCR datasets.

Dataset	InfoTS$_s$	InfoTS	TS2Vec	T-Loss	TNC	TS-TCC	TST	DTW
ArticularyWordRecognition	**0.993**	0.987	0.987	0.943	0.973	0.953	0.977	0.987
AtrialFibrillation	**0.267**	0.200	0.200	0.133	0.133	**0.267**	0.067	0.200
BasicMotions	**1.000**	0.975	0.975	**1.000**	0.975	**1.000**	0.975	0.975
CharacterTrajectories	0.987	0.974	**0.995**	0.993	0.967	0.985	0.975	0.989
Cricket	**1.000**	0.986	0.972	0.972	0.958	0.917	**1.000**	**1.000**
DuckDuckGeese	0.600	0.540	**0.680**	0.650	0.460	0.380	0.620	0.600
EigenWorms	0.748	0.733	**0.847**	0.840	0.840	0.779	0.748	0.618
Epilepsy	**0.993**	0.971	0.964	0.971	0.957	0.957	0.949	0.964
ERing	**0.953**	0.949	0.874	0.133	0.852	0.904	0.874	0.133
EthanolConcentration	**0.323**	0.281	0.308	0.205	0.297	0.285	0.262	**0.323**
FaceDetection	0.525	0.534	0.501	0.513	0.536	**0.544**	0.534	0.529
FingerMovements	0.620	**0.630**	0.480	0.580	0.470	0.460	0.560	0.530
HandMovementDirection	**0.514**	0.392	0.338	0.351	0.324	0.243	0.243	0.231
Handwriting	**0.554**	0.452	0.515	0.451	0.249	0.498	0.225	0.286
Heartbeat	**0.771**	0.722	0.683	0.741	0.746	0.751	0.746	0.717
JapaneseVowels	0.986	0.984	0.984	**0.989**	0.978	0.930	0.978	0.949
Libras	**0.889**	0.883	0.867	0.883	0.817	0.822	0.656	0.870
LSST	0.593	0.591	0.537	0.509	**0.595**	0.474	0.408	0.551
MotorImagery	0.610	**0.630**	0.510	0.580	0.500	0.610	0.500	0.500
NATOPS	**0.939**	0.933	0.928	0.917	0.911	0.822	0.850	0.883
PEMS-SF	**0.757**	0.751	0.682	0.676	0.699	0.734	0.740	0.711
PenDigits	0.989	**0.990**	0.989	0.981	0.979	0.974	0.560	0.977
PhonemeSpectra	0.233	0.249	0.233	0.222	0.207	**0.252**	0.085	0.151
RacketSports	0.829	**0.855**	**0.855**	**0.855**	0.776	0.816	0.809	0.803
SelfRegulationSCP1	**0.887**	0.874	0.812	0.843	0.799	0.823	0.754	0.775
SelfRegulationSCP2	0.572	**0.578**	**0.578**	0.539	0.550	0.533	0.550	0.539
SpokenArabicDigits	0.932	0.947	**0.988**	0.905	0.934	0.970	0.923	0.963
StandWalkJump	**0.467**	**0.467**	**0.467**	0.333	0.400	0.333	0.267	0.200
UWaveGestureLibrary	0.884	0.884	**0.906**	0.875	0.759	0.753	0.575	0.903
InsectWingbeat	**0.472**	0.470	0.466	0.156	0.469	0.264	0.105	–
Avg. ACC	**0.730**	0.714	0.704	0.658	0.670	0.668	0.617	0.629
Avg. Rank	**1.967**	2.633	3.067	3.833	4.367	4.167	5.0	4.366

Table 9: Full results of multivariate time series classification on 30 UEA datasets.