Paper Sharing

Self-Supervised Contrastive Pre-Training for Time Series via Time-Frequency Consistency

Lecturer：Yuxin Wu

2023.10.26

Introduction

- Time series plays important roles in many areas
  - Clinical diagnosis and Healthcare settings
  - Traffic analysis
  - Climate science
  - browsing behaviors

⭑ Learning broadly generalizable representations for time series remains a fundamentally challenging problem

Introduction

- Numerous immediate benefits → Pre-training capability
  - train a neural network model on a dataset
  - transferring it to a new target dataset for fine-tuning
  - expect the resulting performance to be good

☆ However…

Introduction

☆ However:

- A variety of realistic reasons:
  - Distribution shifts
  - Unknown properties of the target dataset during pre-training
- What makes situation worse? ➡ the complexity of time series
  - Large variations of temporal dynamics across datasets
  - Varying semantic meaning
  - Irregular sampling
  - System factors (e.g., different devices or subjects)

☆ What kind of inductive biases could facilitate generalizable representations of time series?

Introduction

☆ What kind of inductive biases could facilitate generalizable representations of time series?

☆ Require: the pre-training model captures a latent property hold true for previously unseen target datasets

A shared property:

- be shared between pre-training and target datasets
- enable knowledge transfer from pre-training to fine-tuning

- Maybe hard to understand? Let' s look at examples:

Introduction

☆ Require: the pre-training model captures a latent property

- be shared between pre-training and target datasets
- enable knowledge transfer from pre-training to fine-tuning

☆ In Computer Vision (CV) :

- initial neural layers capture universal visual elements (edges and shapes)
- relevant regardless of image style and tasks

☆ In Natural Language Processing (NLP) :

- linguistic principles of semantics
- grammar shared across different languages

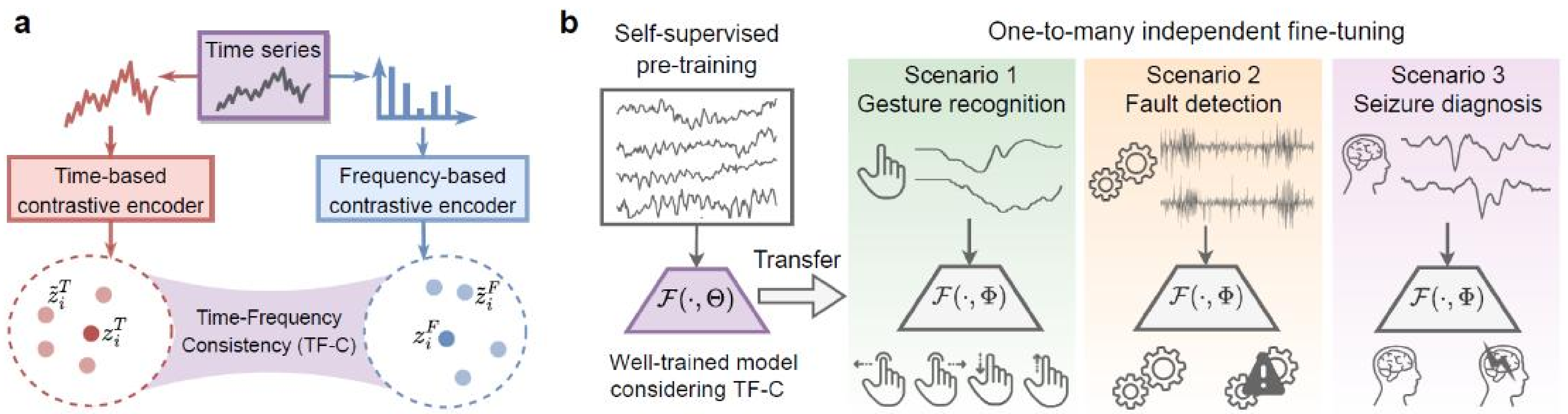
☆ However:

the complexity of time series ➡ such principle has not yet been established

Introduction

☆ How to solve the dilemma?

→ Time-Frequency Consistency(TF-C)



Rationale for Time-Frequency Consistency (TF-C)

- Central idea:

to identify a general property that is preserved across time series datasets and use it to induce transfer learning for effective pre-training

- time domain: how sensor readouts change with time
- how much of the signal lies within each frequency component over the entire spectrum

☆ The relationship between the two domains, grounded in signal processing theory, provides an invariance that is valid regardless of the time series distribution and thus can serve as an inductive bias for pretraining.

Representational Time-Frequency Consistency (TF-C). *Let $\boldsymbol{x}_i$ be a time series and $\mathcal{F}$ be a model satisfying TF-C. Then, time-based representation $\boldsymbol{z}_i^{\mathrm{T}}$ and frequency-based representation $\boldsymbol{z}_i^{\mathrm{F}}$ as well as representations of $\boldsymbol{x}_i$'s local augmentations are proximal in the latent time-frequency space.*

Rationale for Time-Frequency Consistency (TF-C)

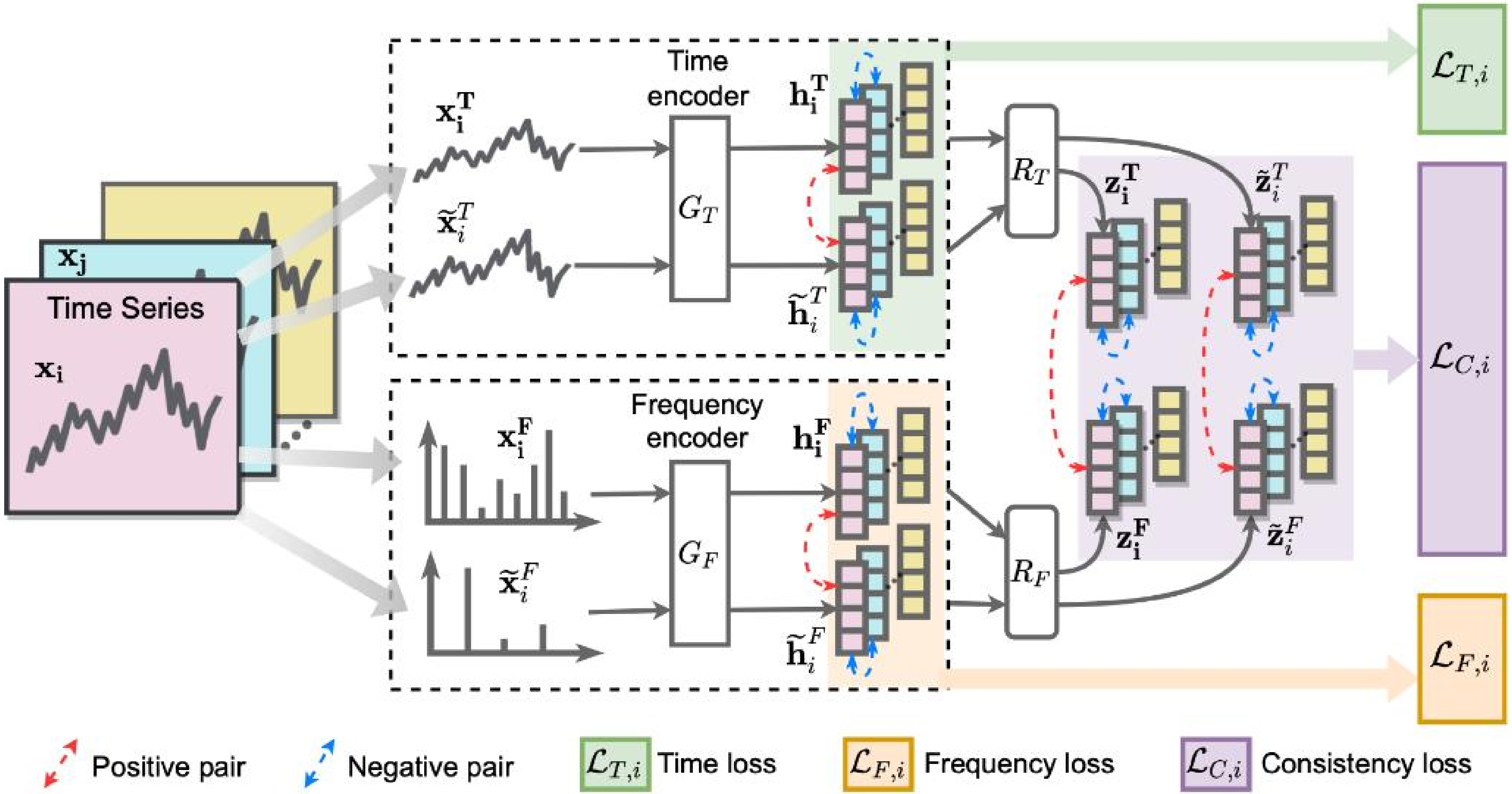
Representational Time-Frequency Consistency (TF-C). *Let $\boldsymbol{x}_i$ be a time series and $\mathcal{F}$ be a model satisfying TF-C. Then, time-based representation $\boldsymbol{z}_i^{\mathrm{T}}$ and frequency-based representation $\boldsymbol{z}_i^{\mathrm{F}}$ as well as representations of $\boldsymbol{x}_i$'s local augmentations are proximal in the latent time-frequency space.*

An anology

Method

- TF-C has four neural network components:
  - Contrastive time encoder
  - Contrastive frequency encoder
  - two cross-space projectors that map time-based and frequency-based representations to the same time-frequency space.

Method



Time Series
$\mathbf{x_i}$
$\mathbf{x_j}$
$\mathbf{x_i^T}$
$\tilde{\mathbf{x}}_i^T$
Time encoder
G_T
$\mathbf{h_i^T}$
$\tilde{\mathbf{h}}_i^T$
$\mathbf{x_i^F}$
$\tilde{\mathbf{x}}_i^F$
Frequency encoder
G_F
$\mathbf{h_i^F}$
$\tilde{\mathbf{h}}_i^F$
R_T
R_F
$\mathbf{z_i^T}$
$\tilde{\mathbf{z}}_i^T$
$\mathbf{z_i^F}$
$\tilde{\mathbf{z}}_i^F$
$\mathcal{L}_{T,i}$
$\mathcal{L}_{C,i}$
$\mathcal{L}_{F,i}$
Positive pair
Negative pair
$\mathcal{L}_{T,i}$ Time loss
$\mathcal{L}_{F,i}$ Frequency loss
$\mathcal{L}_{C,i}$ Consistency loss

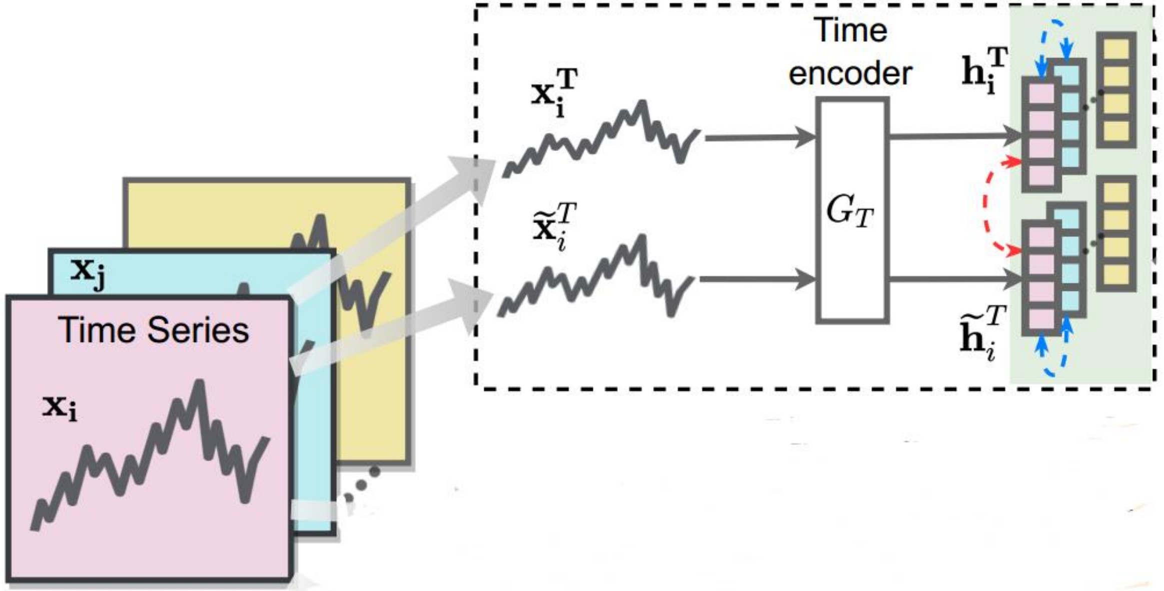
Method in detail

- Time-based Contrastive Encoder

$$\boldsymbol{x}_i^{\mathrm{T}} \rightarrow \mathcal{X}_i^{\mathrm{T}} \qquad \widetilde{\boldsymbol{x}}_i^{\mathrm{T}} \in \mathcal{X}_i^{\mathrm{T}}$$

$$\boldsymbol{h}_i^{\mathrm{T}} = G_{\mathrm{T}}(\boldsymbol{x}_i^{\mathrm{T}}) \text{ and } \widetilde{\boldsymbol{h}}_i^{\mathrm{T}} = G_{\mathrm{T}}(\widetilde{\boldsymbol{x}}_i^{\mathrm{T}}).$$

$$\boldsymbol{x}_j^{\mathrm{T}} \in \mathcal{D}^{\text{pret}} \left\{ \begin{array}{l} (\boldsymbol{x}_i^{\mathrm{T}}, \widetilde{\boldsymbol{x}}_i^{\mathrm{T}}) \\ \\ (\boldsymbol{x}_i^{\mathrm{T}}, \boldsymbol{x}_j^{\mathrm{T}}) \text{ and } (\boldsymbol{x}_i^{\mathrm{T}}, \widetilde{\boldsymbol{x}}_j^{\mathrm{T}}) \end{array} \right.$$

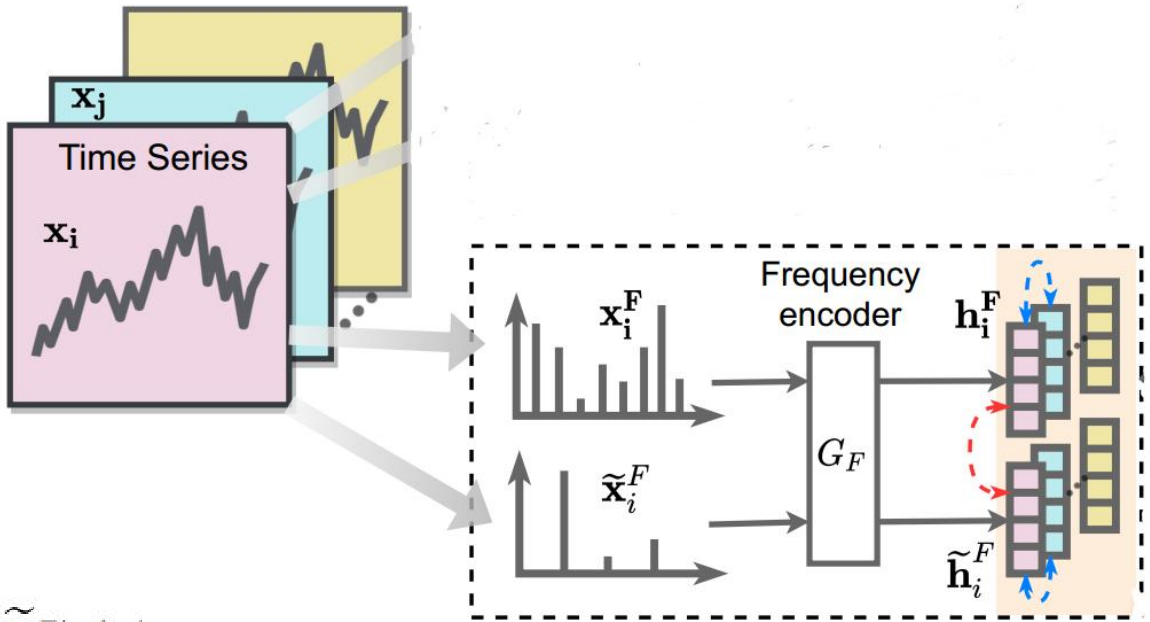


$$\mathcal{L}_{\mathrm{T},i} = d(\boldsymbol{h}_i^{\mathrm{T}}, \widetilde{\boldsymbol{h}}_i^{\mathrm{T}}, \mathcal{D}^{\text{pret}}) = -\log \frac{\exp(\operatorname{sim}(\boldsymbol{h}_i^{\mathrm{T}}, \widetilde{\boldsymbol{h}}_i^{\mathrm{T}})/\tau)}{\sum_{\boldsymbol{x}_j \in \mathcal{D}^{\text{pret}}} \mathbb{1}_{i \neq j} \exp(\operatorname{sim}(\boldsymbol{h}_i^{\mathrm{T}}, G_{\mathrm{T}}(\boldsymbol{x}_j))/\tau)}$$

Method in detail

- Frequency-based Contrastive Encoder

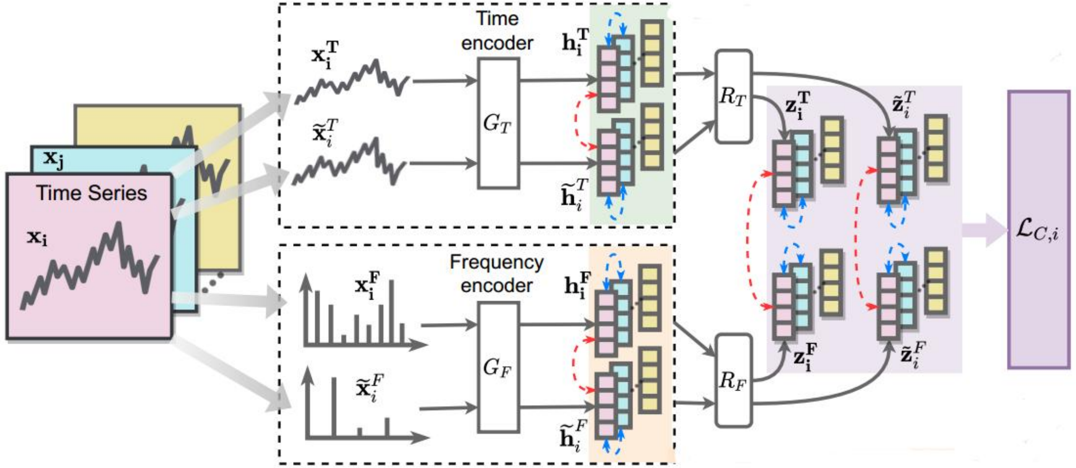
Quite the same!



$$\mathcal{L}_{\mathrm{F},i} = d(\boldsymbol{h}_i^{\mathrm{F}}, \widetilde{\boldsymbol{h}}_i^{\mathrm{F}}, \mathcal{D}^{\mathrm{pret}}) = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{h}_i^{\mathrm{F}}, \widetilde{\boldsymbol{h}}_i^{\mathrm{F}})/\tau)}{\sum_{\boldsymbol{x}_j \in \mathcal{D}^{\mathrm{pret}}} \mathbb{1}_{i \neq j} \exp(\mathrm{sim}(\boldsymbol{h}_i^{\mathrm{F}}, G_{\mathrm{F}}(\boldsymbol{x}_j))/\tau)}$$

Method in detail

- Time-Frequency Consistency

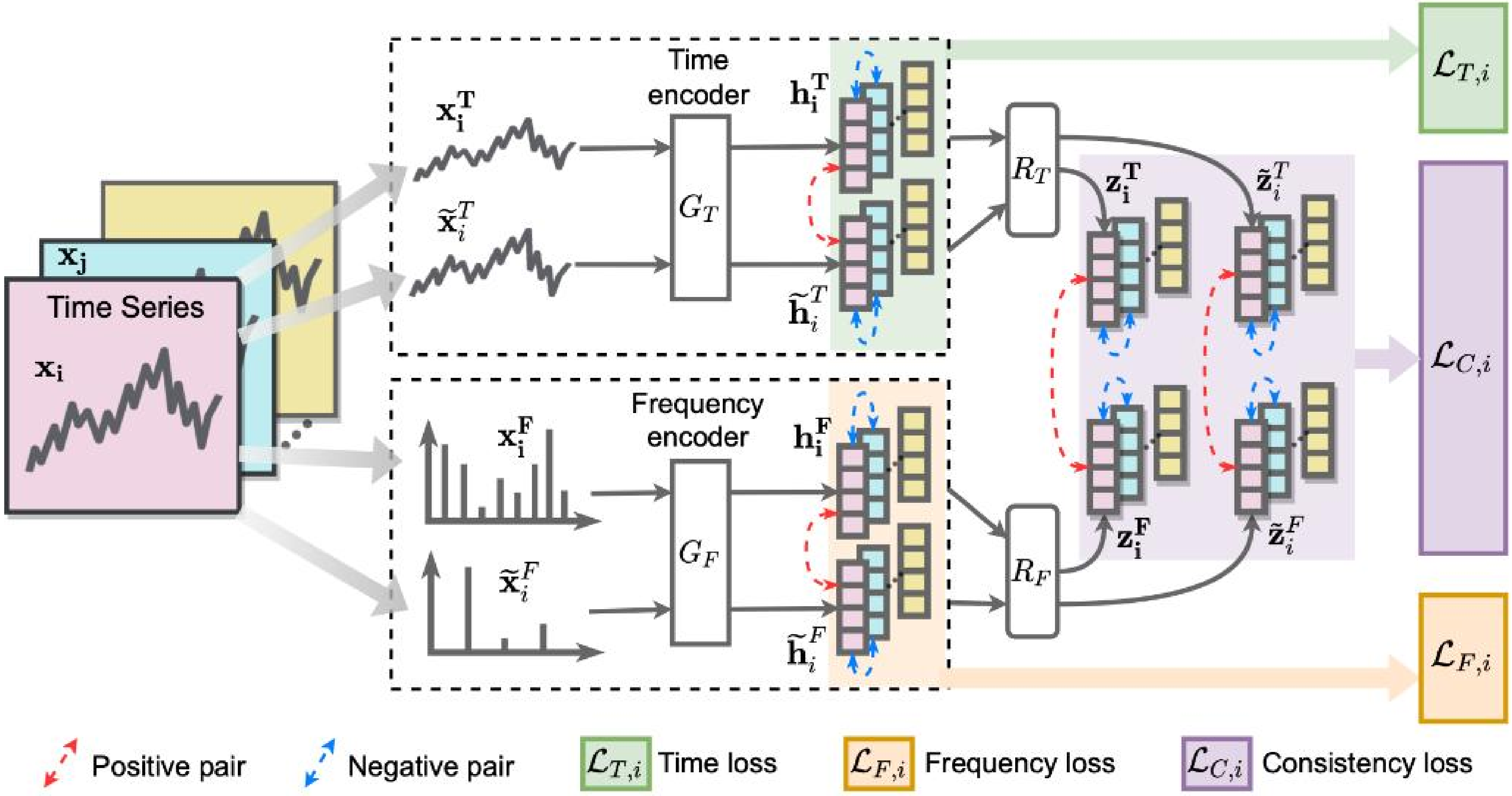


$$z_i^{\mathrm{T}} = R_{\mathrm{T}}(h_i^{\mathrm{T}}),\ \widetilde{z}_i^{\mathrm{T}} = R_{\mathrm{T}}(\widetilde{h}_i^{\mathrm{T}}),\ z_i^{\mathrm{F}} = R_{\mathrm{F}}(h_i^{\mathrm{F}}),\ \text{and}\ \widetilde{z}_i^{\mathrm{F}} = R_{\mathrm{F}}(\widetilde{h}_i^{\mathrm{F}})$$

use $S_i^{\mathrm{TF}} = d(z_i^{\mathrm{T}}, z_i^{\mathrm{F}}, \mathcal{D}^{\mathrm{pret}})$ to denote the distance between z_i^{T} and z_i^{F} $S_i^{\mathrm{T\widetilde{F}}}$ $S_i^{\mathrm{\widetilde{T}F}}$ $S_i^{\mathrm{\widetilde{T}\widetilde{F}}}$

$$\mathcal{L}_{\mathrm{C},i} = \sum_{S^{\mathrm{pair}}} (S_i^{\mathrm{TF}} - S_i^{\mathrm{pair}} + \delta), \quad S^{\mathrm{pair}} \in \{S_i^{\mathrm{T\widetilde{F}}}, S_i^{\mathrm{\widetilde{T}F}}, S_i^{\mathrm{\widetilde{T}\widetilde{F}}}\}$$

Implementation



$$\mathcal{L}_{\text{TF-C},i} = \lambda(\mathcal{L}_{\text{T},i} + \mathcal{L}_{\text{F},i}) + (1 - \lambda)\mathcal{L}_{\text{C},i}$$

Experiments: One-to-One Pre-Training Evaluation

Table 1: One-to-one pre-training evaluation (Scenario 3). Pre-training is performed on HAR, followed by fine-tuning on GESTURE. Results for other three scenarios are shown in Tables 4-6.

Models	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC
Non-DL (KNN)	0.6766±0.0000	0.6500±0.0000	0.6821±0.0000	0.6442±0.0000	0.8190±0.0000	0.5231±0.0000
Random Init.	0.4219±0.0865	0.4751±0.0925	0.4963±0.1026	0.4886±0.0967	0.7129±0.1206	0.3358±0.1194
TS-SD	0.6937±0.0533	0.6806±0.0496	0.6883±0.0525	0.6785±0.0495	0.8708±0.0305	0.6261±0.0790
TS2vec	0.6453±0.0260	0.6287±0.0339	0.6451±0.0218	0.6261±0.0294	0.8890±0.0054	0.6670±0.0118
CLOCS	0.4731±0.0229	0.4639±0.0432	0.4766±0.0266	0.4392±0.0198	0.8161±0.0068	0.4916±0.0103
Mixing-up	0.7183±0.0123	0.7001±0.0166	0.7183±0.0123	0.6991±0.0145	**0.9127±0.0018**	0.7654±0.0071
TS-TCC	0.7593±0.0242	0.7668±0.0257	0.7566±0.0231	0.7457±0.0210	0.8866±0.0040	0.7217±0.0121
SimCLR	0.4383±0.0652	0.4255±0.1072	0.4383±0.0652	0.3713±0.0919	0.7721±0.0559	0.4116±0.0971
TF-C (Ours)	**0.7824±0.0237**	**0.7982±0.0496**	**0.8011±0.0322**	**0.7991±0.0296**	0.9052±0.0136	**0.7861±0.0149**

Experiments: One-to-Many Pre-Training Evaluation

Table 2: One-to-many pre-training evaluation. Pre-training is performed on SLEEPEEG, followed by an independent fine-tuning on EPILEPSY, FD-B, GESTURE, and EMG.

Scenarios	Models	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC
SLEEPEEG ↓ EPILEPSY	**Non-DL (KNN)**	0.8525±0.0000	0.8639±0.0000	0.6431±0.0000	0.6791±0.0000	0.6434±0.0000	0.6279±0.0000
	Random Init.	0.8983±0.0656	0.9213±0.1369	0.7447±0.1135	0.7959±0.1208	0.8578±0.2153	0.6489±0.1926
	TS-SD	0.8952±0.0522	0.8018±0.2244	0.7647±0.1485	0.7767±0.1855	0.7677±0.2452	0.7940±0.1825
	TS2vec	0.9395±0.0044	0.9059±0.0116	0.9039±0.0118	0.9045±0.0067	0.9587±0.0086	0.9430±0.0103
	CLOCS	**0.9507±0.0027**	0.9301±0.0067	**0.9127±0.0165**	**0.9206±0.0066**	0.9803±0.0023	0.9609±0.0116
	Mixing-up	0.8021±0.0000	0.4011±0.0000	0.5000±0.0000	0.4451±0.0000	0.9743±0.0081	0.9618±0.0104
	TS-TCC	0.9253±0.0098	0.9451±0.0049	0.8181±0.0257	0.8633±0.0215	0.9842±0.0034	0.9744±0.0043
	SimCLR	0.9071±0.0344	0.9221±0.0166	0.7864±0.1071	0.8178±0.0998	0.9045±0.0539	0.9128±0.0205
	TF-C (Ours)	0.9495±0.0249	**0.9456±0.0108**	0.8908±0.0216	0.9149±0.0534	**0.9811±0.0237**	**0.9703±0.0199**
SLEEPEEG ↓ FD-B	**Non-DL (KNN)**	0.4473±0.0000	0.2847±0.0000	0.3275±0.0000	0.2284±0.0000	0.4946±0.0000	0.3308±0.0000
	Random Init.	0.4736±0.0623	0.4829±0.0529	0.5235±0.1023	0.4911±0.0590	0.7864±0.0349	0.7528±0.0254
	TS-SD	0.5566±0.0210	0.5710±0.0535	0.6054±0.0272	0.5703±0.0328	0.7196±0.0113	0.5693±0.0532
	TS2vec	0.4790±0.0113	0.4339±0.0092	0.4842±0.0197	0.4389±0.0107	0.6463±0.0130	0.4442±0.0162
	CLOCS	0.4927±0.0310	0.4824±0.0316	0.5873±0.0387	0.4746±0.0485	0.6992±0.0099	0.5501±0.0365
	Mixing-up	0.6789±0.0246	0.7146±0.0343	**0.7613±0.0198**	0.7273±0.0228	0.8209±0.0035	0.7707±0.0042
	TS-TCC	0.5499±0.0220	0.5279±0.0293	0.6396±0.0178	0.5418±0.0338	0.7329±0.0203	0.5824±0.0468
	SimCLR	0.4917±0.0437	0.5446±0.1024	0.4760±0.0885	0.4224±0.1138	0.6619±0.0219	0.5009±0.0477
	TF-C (Ours)	**0.6938±0.0231**	**0.7559±0.0349**	0.7202±0.0257	**0.7487±0.0268**	**0.8965±0.0135**	**0.7871±0.0267**
SLEEPEEG ↓ GESTURE	**Non-DL (KNN)**	0.6833±0.0000	0.6501±0.0000	0.6833±0.0000	0.6443±0.0000	0.8190±0.0000	0.5232±0.0000
	Random Init.	0.4219±0.0629	0.4751±0.0175	0.4963±0.0679	0.4886±0.0459	0.7129±0.0166	0.3358±0.1439
	TS-SD	0.6922±0.0444	0.6698±0.0472	0.6867±0.0488	0.6656±0.0443	0.8725±0.0324	0.6185±0.0966
	TS2vec	0.6917±0.0333	0.6545±0.0358	0.6854±0.0349	0.6570±0.0392	0.8968±0.0123	0.6989±0.0346
	CLOCS	0.4433±0.0518	0.4237±0.0794	0.4433±0.0518	0.4014±0.0602	0.8073±0.0109	0.4460±0.0384
	Mixing-up	0.6933±0.0231	0.6719±0.0232	0.6933±0.0231	0.6497±0.0306	0.8915±0.0261	0.7279±0.0558
	TS-TCC	0.7188±0.0349	0.7135±0.0352	0.7167±0.0373	0.6984±0.0360	0.9099±0.0085	0.7675±0.0201
	SimCLR	0.4804±0.0594	0.5946±0.1623	0.5411±0.1946	0.4955±0.1870	0.8131±0.0521	0.5076±0.1588
	TF-C (Ours)	**0.7642±0.0196**	**0.7731±0.0355**	**0.7429±0.0268**	**0.7572±0.0311**	**0.9238±0.0159**	**0.7961±0.0109**
SLEEPEEG ↓ EMG	**Non-DL (KNN)**	0.4390±0.0000	0.3772±0.0000	0.5143±0.0000	0.3979±0.0000	0.6025±0.0000	0.4084±0.0000
	Random Init.	0.7780±0.0729	0.5909±0.0625	0.6667±0.0135	0.6238±0.0267	0.9109±0.1239	0.7771±0.1427
	TS-SD	0.4606±0.0000	0.1545±0.0000	0.3333±0.0000	0.2111±0.0000	0.5005±0.0126	0.3775±0.0110
	TS2vec	0.7854±0.0318	**0.8040±0.0750**	0.6785±0.0396	0.6766±0.0501	**0.9331±0.0164**	**0.8436±0.0372**
	CLOCS	0.6985±0.0323	0.5306±0.0750	0.5354±0.0291	0.5139±0.0409	0.7923±0.0573	0.6484±0.0680
	Mixing-up	0.3024±0.0534	0.1099±0.0126	0.2583±0.0456	0.1541±0.0204	0.4506±0.1718	0.3660±0.1635
	TS-TCC	0.7889±0.0192	0.5851±0.0974	0.6310±0.0991	0.5904±0.0952	0.8851±0.0113	0.7939±0.0386
	SimCLR	0.6146±0.0582	0.5361±0.1724	0.4990±0.1214	0.4708±0.1486	0.7799±0.1344	0.6392±0.1596
	TF-C (Ours)	**0.8171±0.0287**	0.7265±0.0353	**0.8159±0.0289**	**0.7683±0.0311**	0.9152±0.0211	0.8329±0.0137

Conclusion

- Develop a pre-training approach that introduces time-frequency consistency (TF-C) as a mechanism to support knowledge transfer between time-series datasets.

- Use self-supervised contrastive estimation and injects TF-C into pre-training, bringing time-based and frequency-based representations and their local neighborhoods close together in the latent space

Thank you for listening

主講人：吴雨欣

2023.10.26