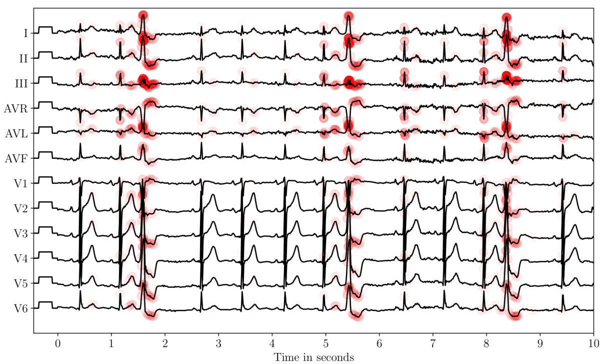
Semi-supervised Time Series Classification By Temporal Relation Prediction

Haoyi Fan [1], *Fengbin Zhang* [1], *Ruidong Wang* [1], *Xunhua Huang* [1], *Zuoyong Li* [2]

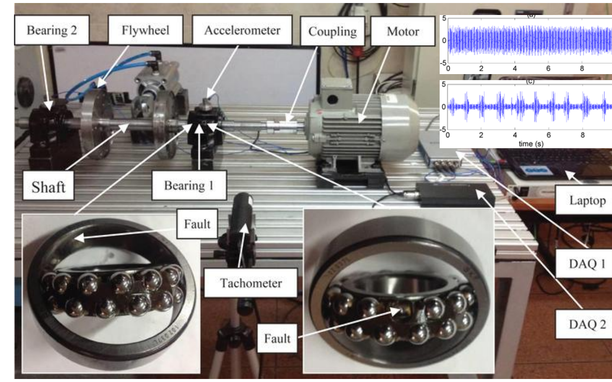
Harbin University of Science and Technology [1]
Minjiang University [2]
isfanhy@hrbust.edu.cn

Background

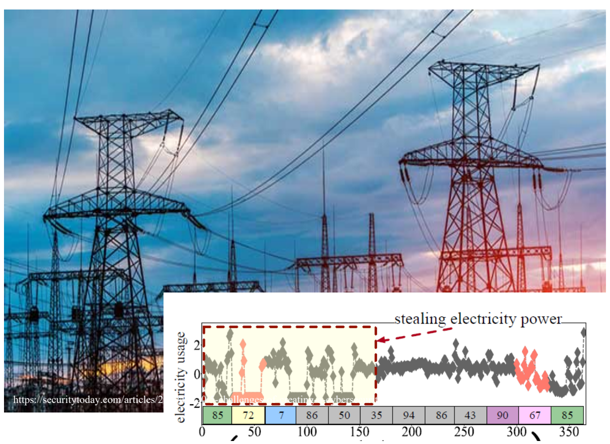


Healthcare Diagnosis

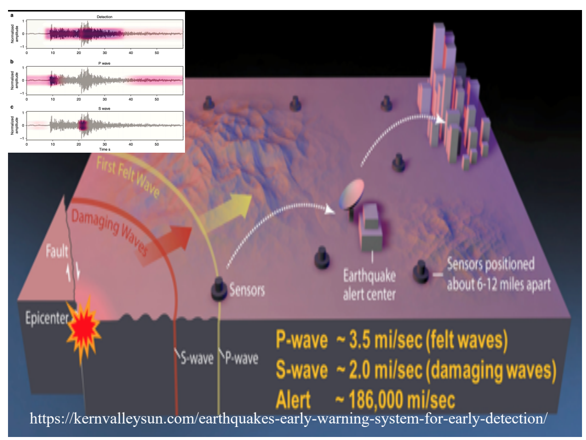
Finance Forecasting



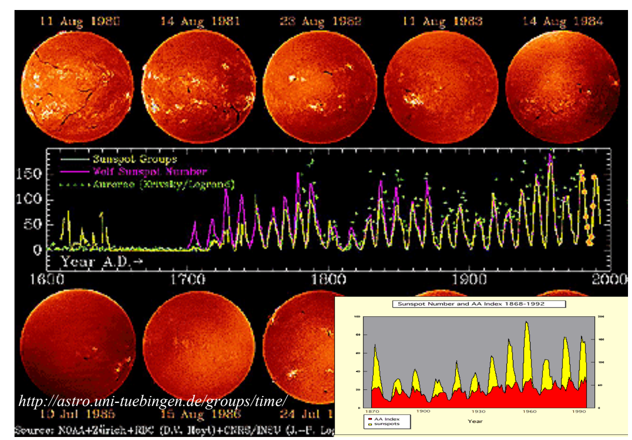
Industrial Fault Detection



Cybersecurity Monitoring



Disaster Warning

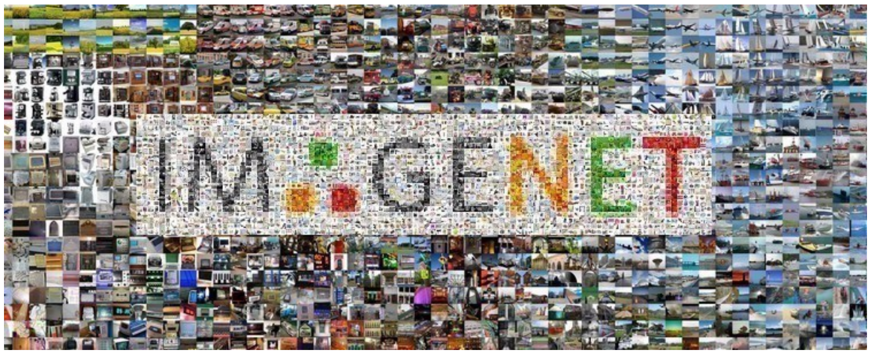


Astrophysical Analysis

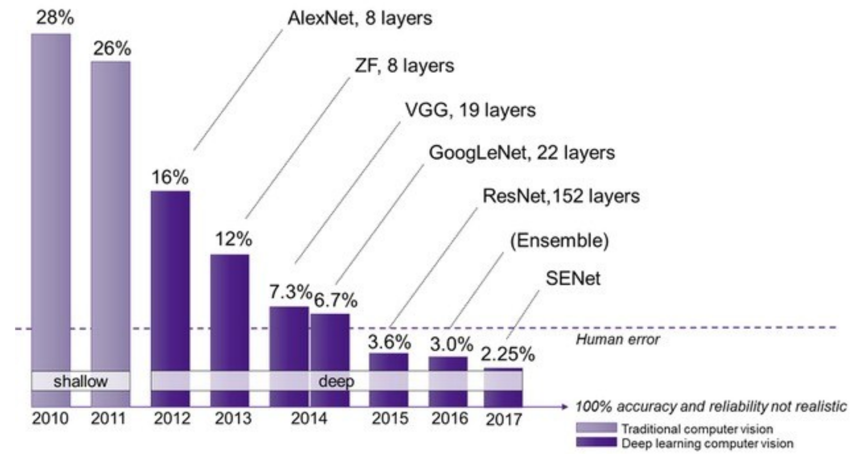
Strodthoff, Nils, et al. "Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL." arXiv preprint arXiv:2004.13701, (2020).
Li, Chuan, et al. “Fuzzy determination of informative frequency band for bearing fault detection.” Journal of Intelligent & Fuzzy Systems, (2016).
Cheng, Ziqiang, et al. “Time2Graph: Revisiting Time Series Modeling with Dynamic Shapelets.” AAAI, (2020).
Mousavi, S. Mostafa, et al. “Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking.” Nature communications, (2020).
Koenig, Michael, Rüdiger Staubert, and Jens Timmer. "Analyzing X-ray Variability by State Space Models." Astronomical Time Series, (1997).

Powerful Supervised Learning

Deep Learning + Supervised Learning



https://cv.gluon.ai/build/examples_datasets/imagenet.html



https://semiengineering.com/new-vision-technologies-for-real-world-applications/

1000 categories:
Training: 1 million, Test: 100 k

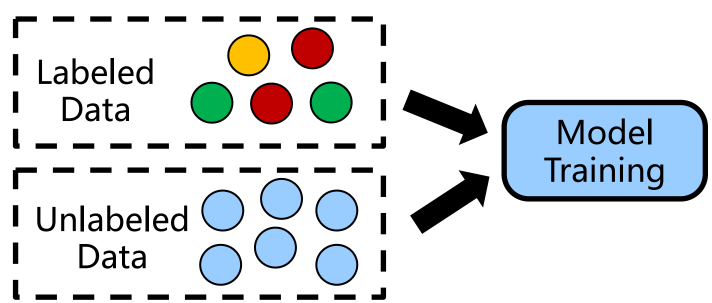
Deep Supervised Learning is powerful … when task and data permit it.

Challenges:

Labeled data can be hard to get
- Labels may require human efforts
- Labels may require special devices

Unlabeled data are usually abundant

Semi-supervised learning:



Modern Semi-supervised Learning Paradigm

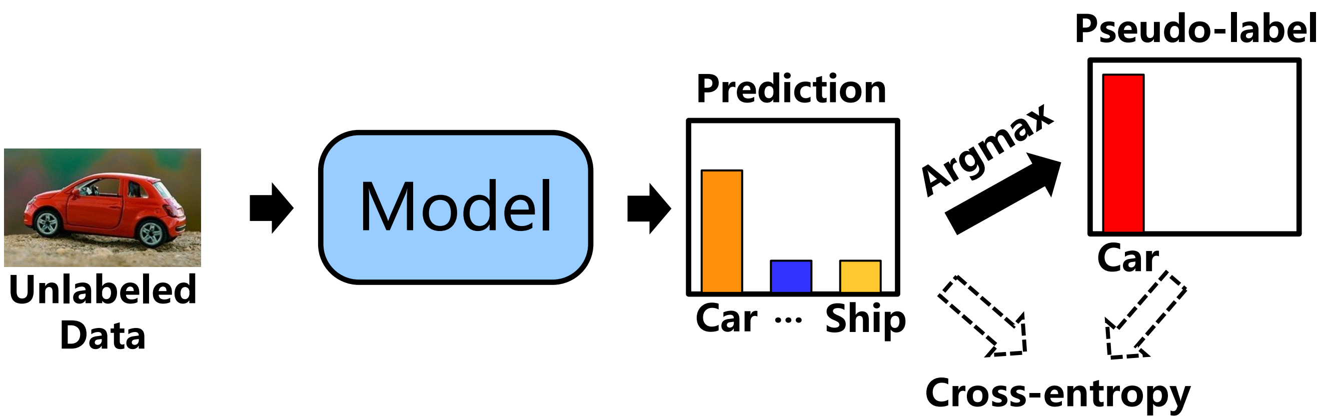
Semi-supervised Learning:

- Entropy minimization based methods
- Consistency regularization based methods
- Hybrid methods

Modern Semi-supervised Learning Paradigm

Entropy minimization based methods

EntMin [1]
Pseudo-label [2]
Noise Student [3]
Meta Pseudo Labels [4]

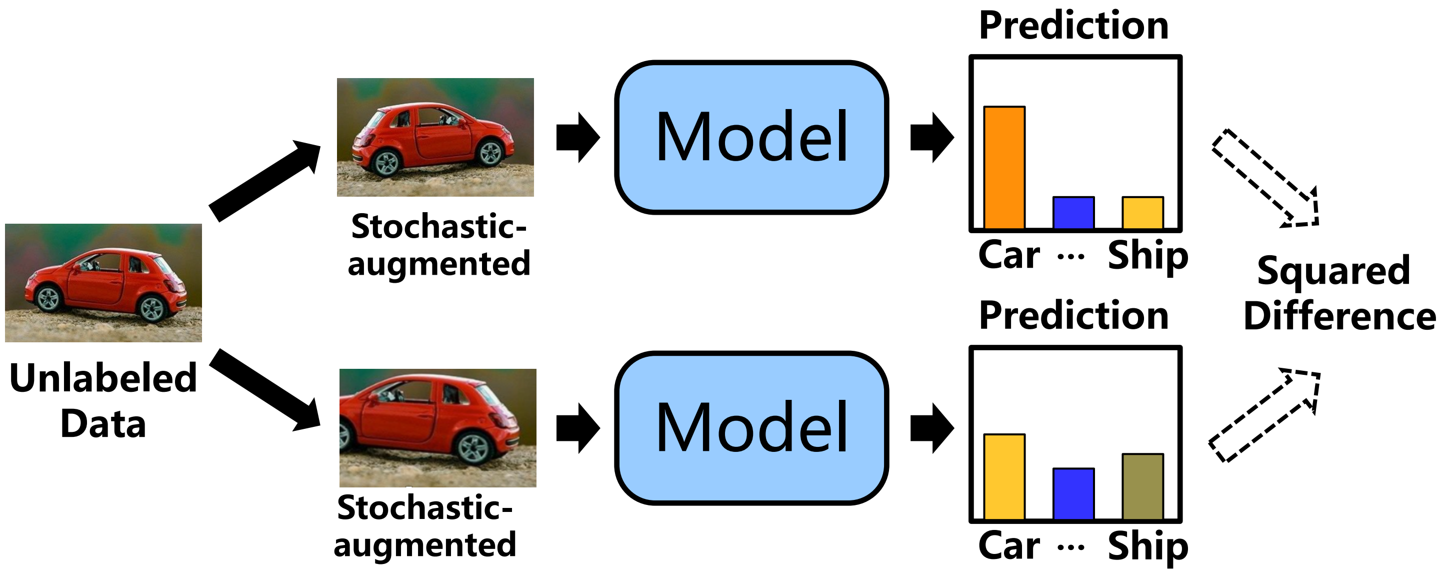


[1] Grandvalet, Yves, and Yoshua Bengio. "Semi-supervised learning by entropy minimization." *NIPS,* (2004).
[2] Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." *ICML, (*2013*).*
[3] Xie, Qizhe, et al. "Self-training with noisy student improves imagenet classification." *CVPR,* (2020).
[4] Pham, Hieu, et al. "Meta pseudo labels." *arXiv preprint arXiv:2003.1058, (2020).*

Modern Semi-supervised Learning Paradigm

Consistency regularization based methods

π-Model [5]
Temporal Ensembling [6]
Mean Teacher [7]
VAT [8]

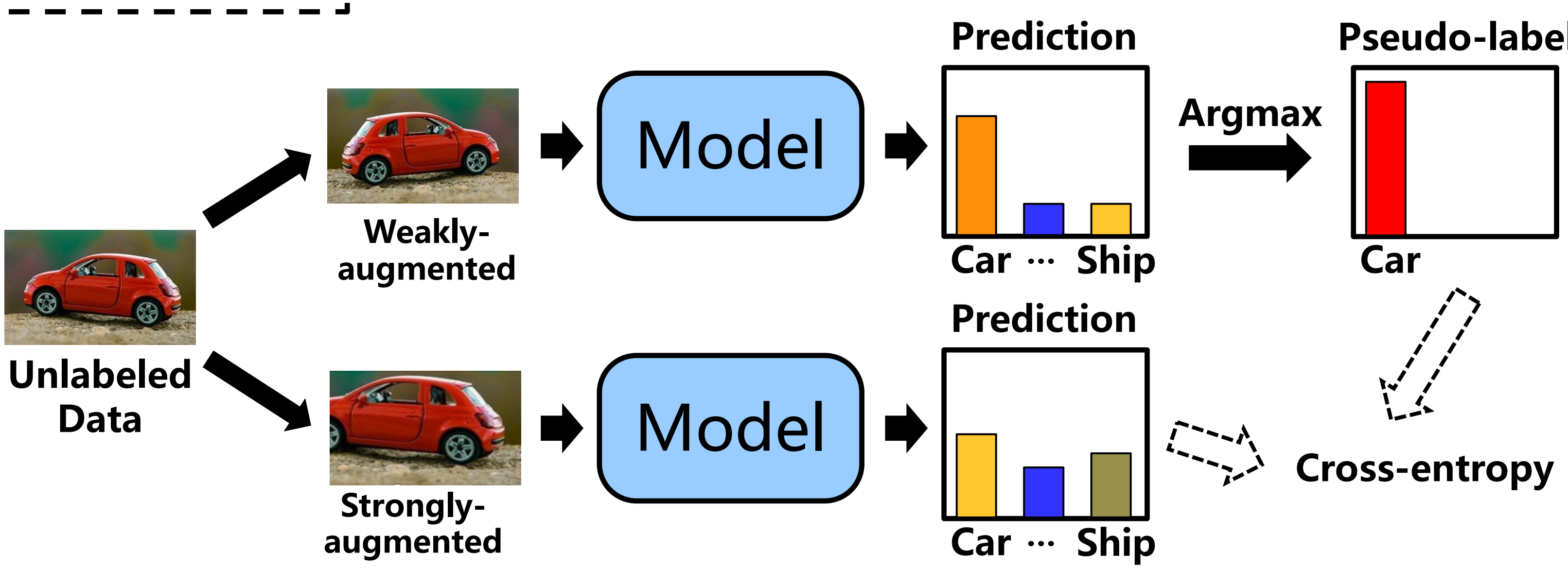


[5] Sajjadi, Mehdi, Mehran Javanmardi, and Tolga Tasdizen. "Regularization with stochastic transformations and perturbations for deep semi-supervised learning." *NIPS,* (2016).
[6] Laine, Samuli, and Timo Aila. "Temporal ensembling for semi-supervised learning." *ICLR* (2016).
[7] Tarvainen, Antti, and Harri Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results." *NIPS* (2017).
[8] Miyato, Takeru, et al. "Virtual adversarial training: a regularization method for supervised and semi-supervised learning." *TPAMI* (2018).

Modern Semi-supervised Learning Paradigm

Hybrid methods

Mixmatch [9]
Remixmatch [10]
Dividemix [11]
Fixmatch [12]



Berthelot, David, et al. "Mixmatch: A holistic approach to semi-supervised learning." *NeurIPS* (2019).
Berthelot, David, et al. "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring." *ICLR,* (2020).
Li, Junnan, Richard Socher, and Steven CH Hoi. "Dividemix: Learning with noisy labels as semi-supervised learning." *ICLR,* (2020).
Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *NeurIPS,* (2020).

Drawbacks of Existing Works

1. Rely on the model confidence of pseudo label for unlabeled data.

2. Sensitive to the choice of augmentation.

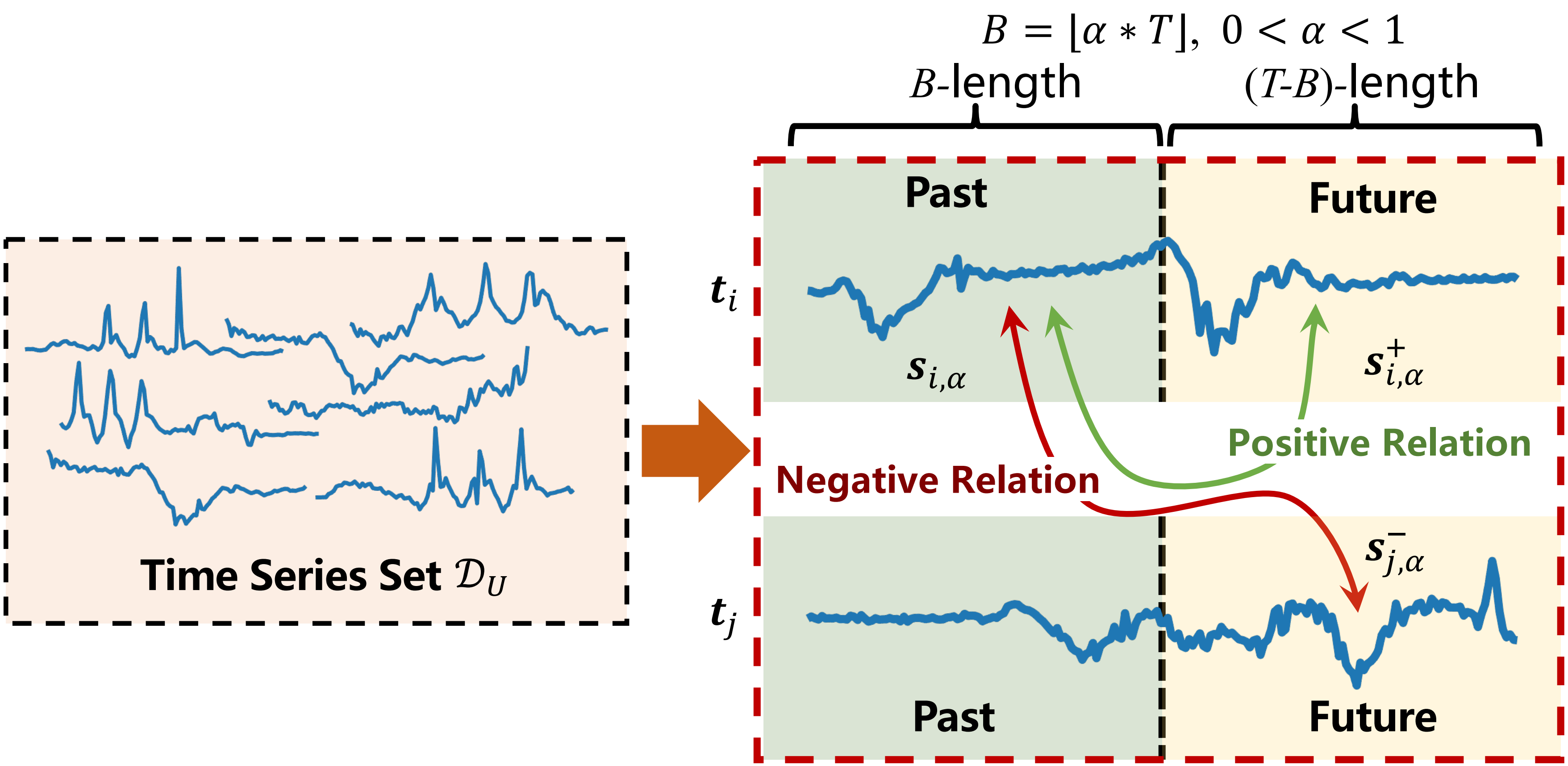
3. Neglect the underlying temporal structure of time series.

4. Unable to discover more general patterns.

Method — SemiTime

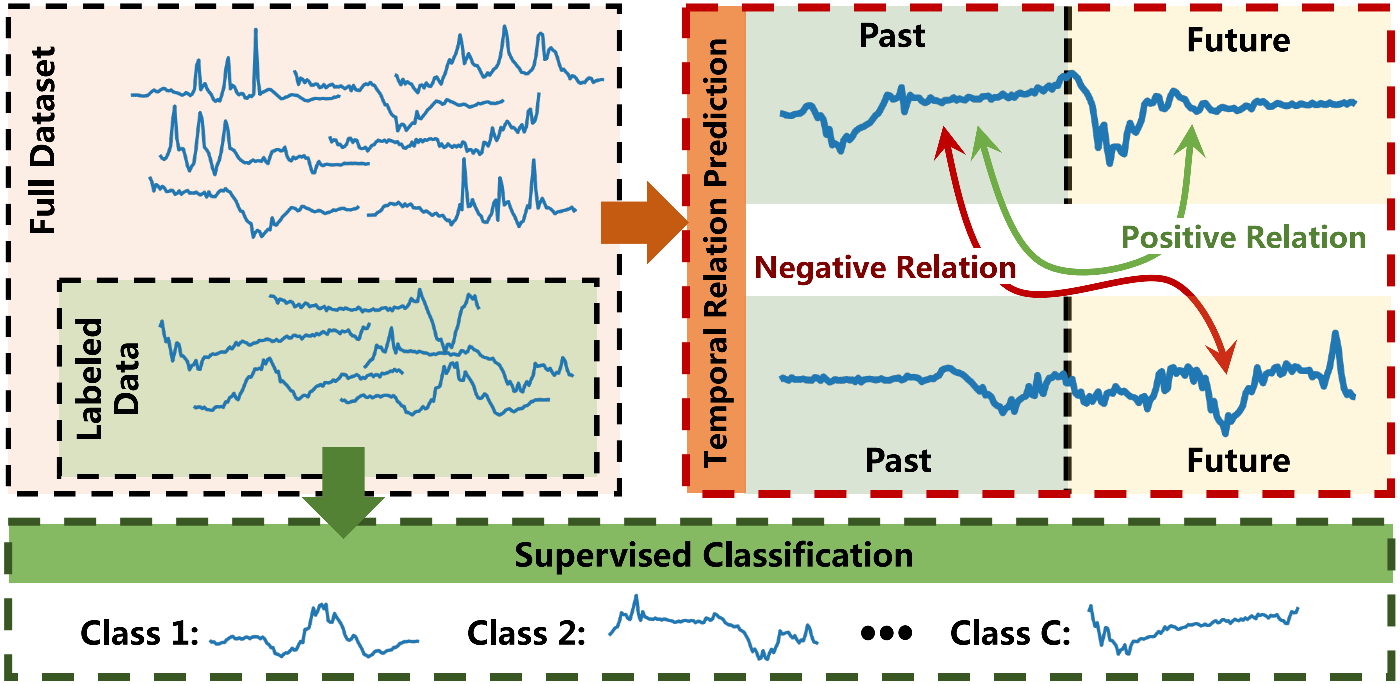
Definition

— Past-Future Temporal Relation



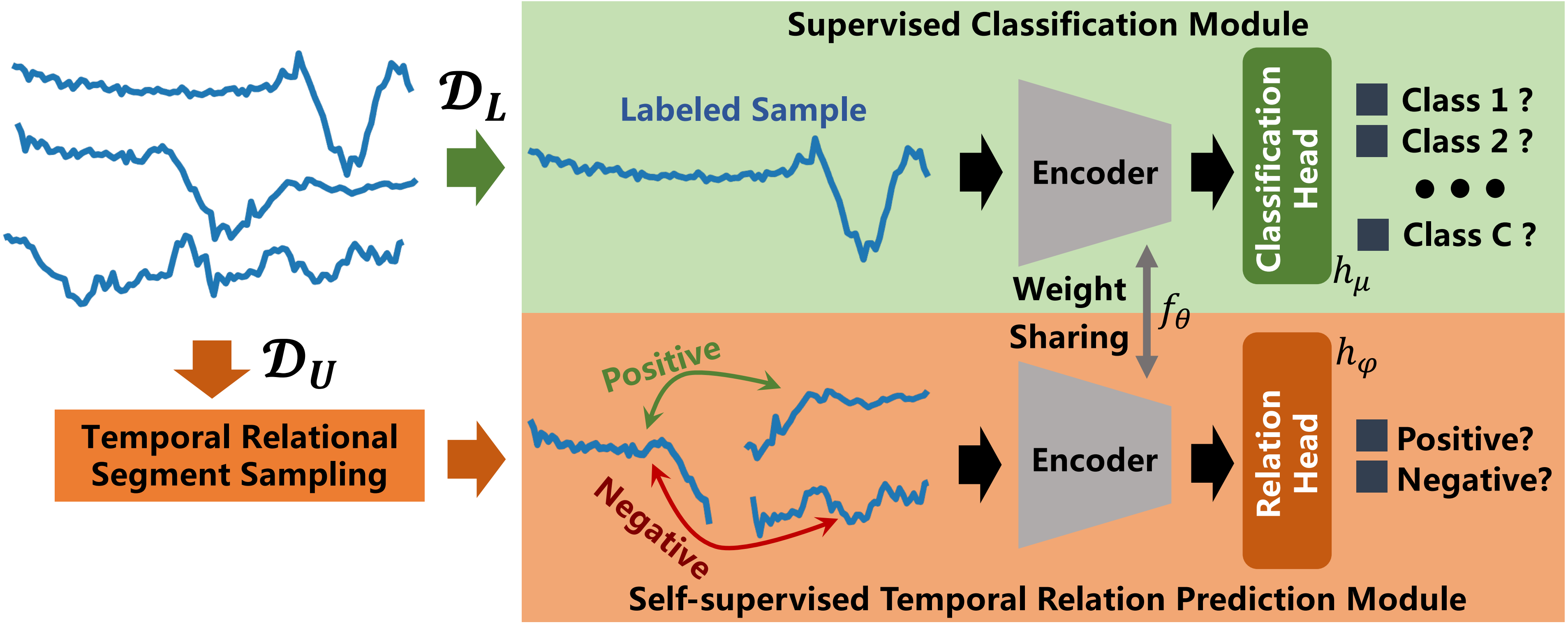
Method — SemiTime

Overview of the proposed method



Method — SemiTime

Architecture detail



Loss :

Supervised Loss :

$$\mathcal{L}_{cls} = -\frac{1}{|\mathcal{D}_L|}\sum_{i=1}^{|\mathcal{D}_L|} y_i \cdot \log(p_i)$$

Self-supervised Loss :

$$\mathcal{L}_{rel} = -\frac{1}{2|\mathcal{D}_U|}\sum_{i=1}^{2|\mathcal{D}_U|} \tilde{y}_i \cdot \log(p_i) + (1-\tilde{y}_i)\cdot(1-\log(p_i))$$

Experiment

Datasets

Table 1. Statistics of Datasets.

Dataset	Sample	Length	Class
CricketX	780	300	12
XJTU	1920	1024	15
InsectWingbeatSound	2200	256	11
MFPT	2574	1024	15
UWaveGestureLibraryAll	4478	945	8
EpilepticSeizure	11500	178	5

Baselines

Fully-supervised baseline
Pseudo-Label Lee et al. 2013
π-Model Laine et al. 2017
MTL Jawed et al. 2020

Evaluation Metric

Accuracy

Experiment

Results

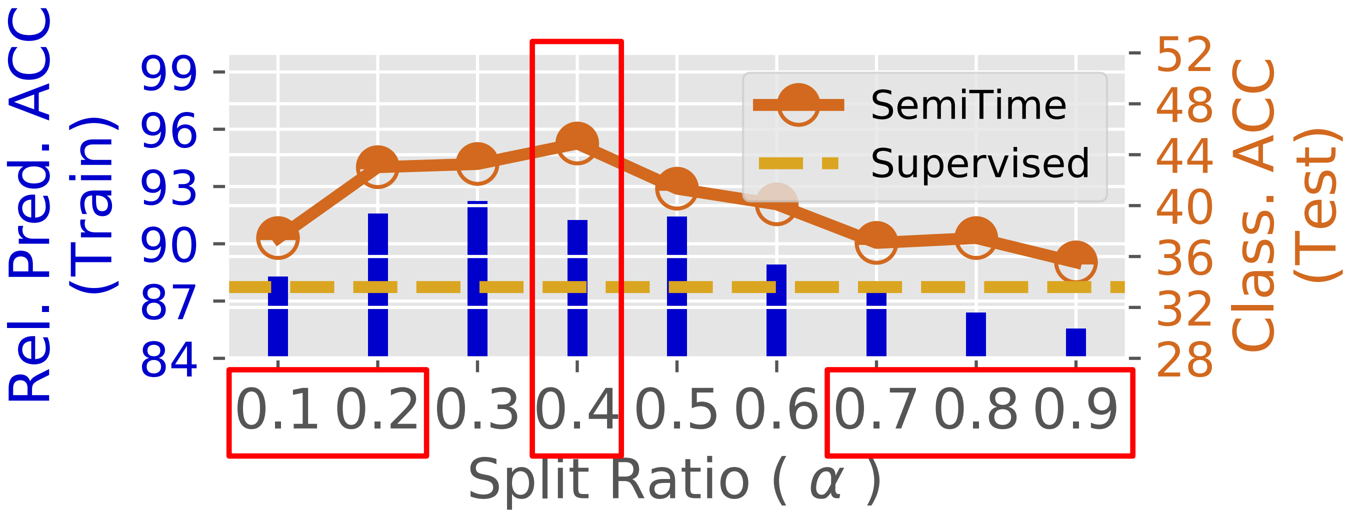
Table 2. Test classification accuracy (%, averages of 10 runs) for supervised baseline and semi-supervised learning on different datasets.All methods use the same 4-layer convolutional backbone. Best results are marked in **red** and the second-best in **blue**.

Label Ratio	10%	20%	40%	100%	10%	20%	40%	100%	10%	20%	40%	100%
Dataset	CricketX				XJTU				InsectWingbeatSound			
Supervised	33.62±0.95	38.79±2.08	52.64±2.53	62.98±2.01	69.71±1.96	83.32±1.59	94.03±1.56	97.92±0.61	50.96±1.58	55.95±0.76	61.41±0.96	**66.27±1.30**
Pseudo-Label [12]	38.87±2.26	44.44±2.91	53.39±2.18	-	74. 11.38%	85.19±1.82	93.97±2.79	-	43. 6.78%	48.35±1.81	55.32±2.04	-
Π-Model [13]	38. 9.62%	48.18±2.07	54.73±1.04	-	**75.96±0.52**	85.93±0.91	**95.03±1.34**	-	**51.47±0.36**	56.14±1.32	**62.20±0.53**	-
MTL [11]	**40.94±1.97**	**50.12±1.22**	**55.10±1.12**	**63.58±1.72**	73.22±1.86	**86.64±1.78**	94.02±1.65	**98.15±1.04**	50.45±1.01	**56.43±0.88**	60.90±0.87	64.14±1.08
Ours	**44.88±3.13**	**51.61±0.66**	**58.71±2.78**	**65.66±1.58**	**84.61±1.39**	**93.93±0.49**	**97.79±0.33**	**98.46±0.25**	**54.96±1.61**	**59.01±1.56**	**62.38±0.76**	**66.57±0.67**
Dataset	MFPT				UWaveGestureLibraryAll				EpilepticSeizure			
Supervised	50.88±0.32	57.14±0.54	69.76±0.48	81.63±0.15	75.81±0.84	81.53±0.54	85.81±0.66	89.5±0.68	68.40±0.43	70.77±0.70	73.49±0.60	77.77±1.13
Pseudo-Label [12]	**63.90±2.62**	65.30±1.70	69.60±2.27	-	75.72±1.85	81. 2.05%	**86.45±1.20**	-	68.57±0.50	72.92±0.48	74.60±0.65	-
Π-Model [13]	55.41±0.65	59. 5.49%	70.15±0.88	-	**77.26±0.31**	**82.87±0.64**	86.17±0.91	-	**69.60±0.34**	71. 3.23%	74.54±0.55	-
MTL [11]	56.11±1.25	**66.20±1.18**	**74.25±1.01**	**82.81±1.06**	76.35±0.56	81.77±0.94	86.01±0.68	**89.76±0.96**	68.71±0.94	**73.17±0.81**	**74.77±0.75**	**78.53±0.62**
Ours	**64.16±0.85**	**69.84±0.94**	**76.49±0.54**	**84.33±0.50**	**81.46±0.60**	**84.57±0.49**	**86.91±0.47**	**90.29±0.32**	**74.86±0.42**	**75.54±0.63**	**77.01±0.79**	**79.26±1.20**

At least 2.05% ~ 11.38% Accuracy improvement!

Experiment

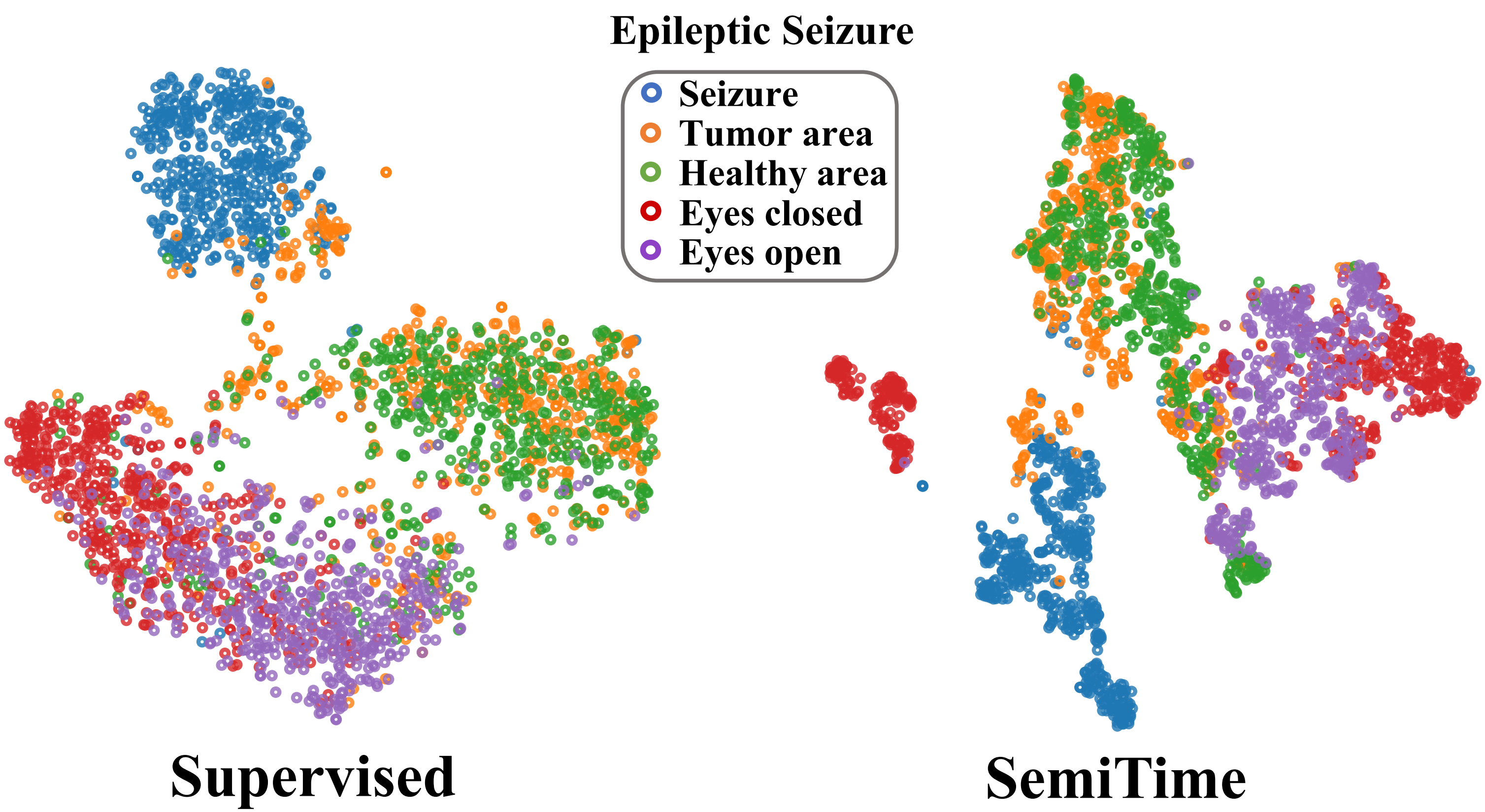
Parameter Sensitivity



SemiTime consistently outperforms the supervised baseline!

Experiment

Visualization



Semantic consistency is captured in the learned representations

Conclusion

- **Traditional semi-supervised model cannot effectively capture the underlying temporal structure of time series.**
- **We propose a general semi-supervised time series classification framework, named SemiTime, by exploring the semantic feature from unlabeled data in a self-supervised manner.**
- **We design a simple but effective temporal relational segments sampling strategy, and based on the sampled relational segments, the useful semantic feature can be extracted from the unlabeled time series data**

Thanks

Q&A

Contact : isfanhy@hrbust.edu.cn

Code and data : https://haoyfan.github.io/