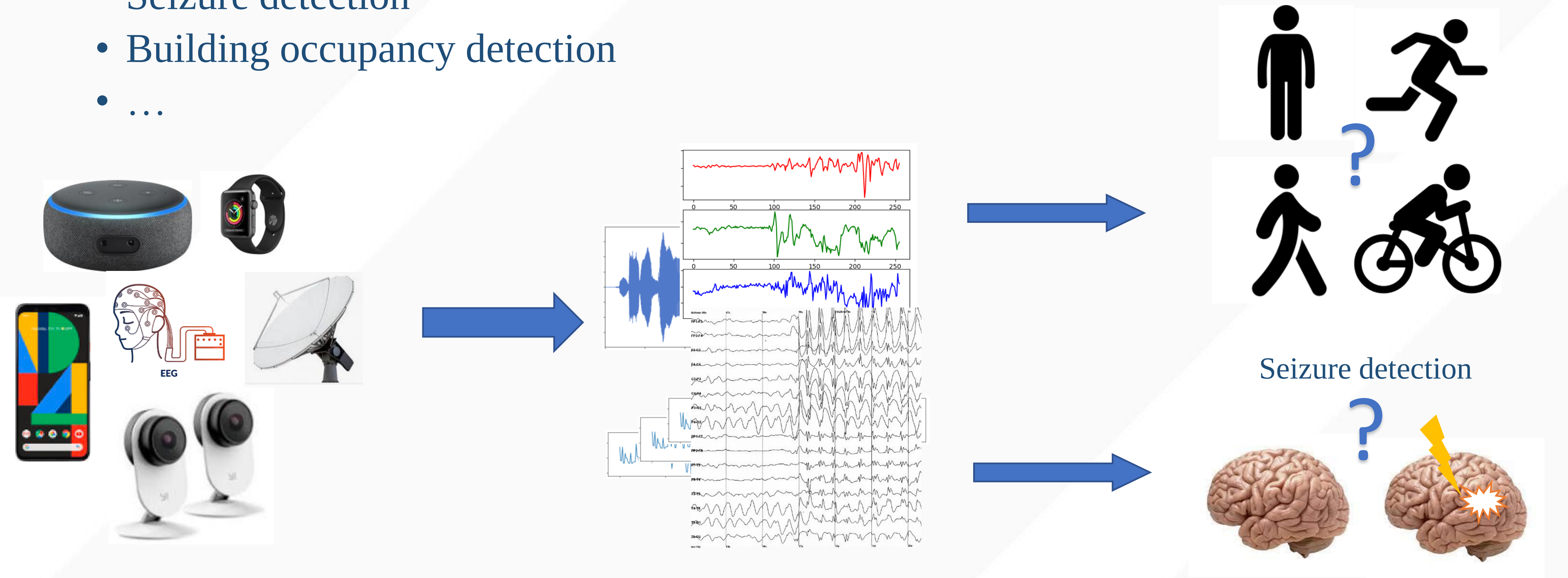
UniTS: Short-Time Fourier Inspired Neural Networks for Sensory Time Series Classification

Shuheng Li, Ranak Roy Chowdhury,
Jingbo Shang, Rajesh K. Gupta, Dezhi Hong
University of California, San Diego

Introduction

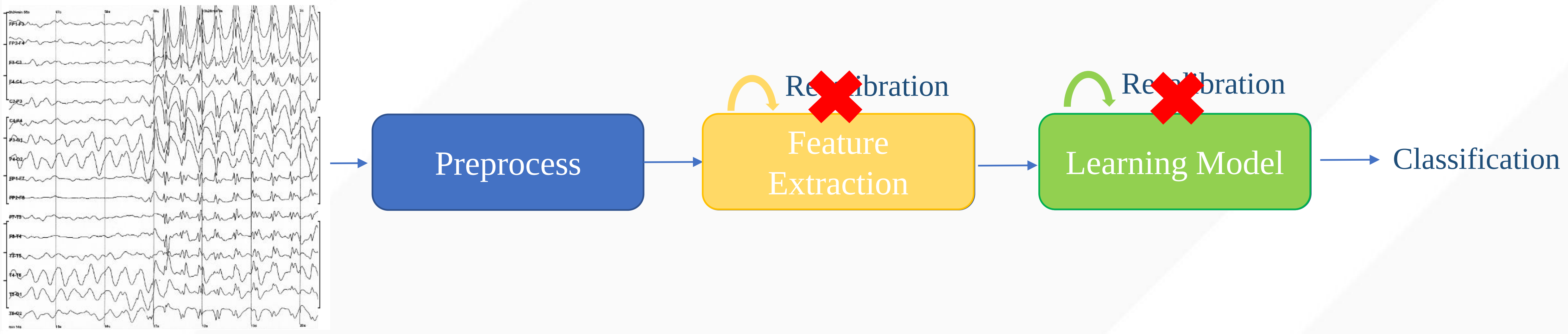
- Sensory time series classification is a crucial problem
  - Human activity recognition
  - Seizure detection
  - Building occupancy detection
  - …

Human activity recognition



Related works – Traditional methods

- Traditional workflows require heavy duplicated manual labor
  - Environment change: recalibration
  - Data/Task change: design from scratch



Related works – Deep neural networks

UC San Diego

- Deep learning techniques are potentiated to generalize well
  - But recent works mainly focus on just one or two types of data

Recent works	Motion data	Medical data	Wave data	Ambient data	Model
MaCNN (IMWUT 18)	O	O	×	O	CNN
Sense-HAR (SenSys 19)	O	×	×	×	CNN
STFNet (WWW 19)	O	×	O	×	STFT + CNN
RFNet (SenSys 20)	×	×	O	×	FFT + LSTM
THAT (AAAI 21)	×	×	O	×	Transformer
LaxNet (WSDM 21)	×	O	×	×	CNN+Attention

Motivation

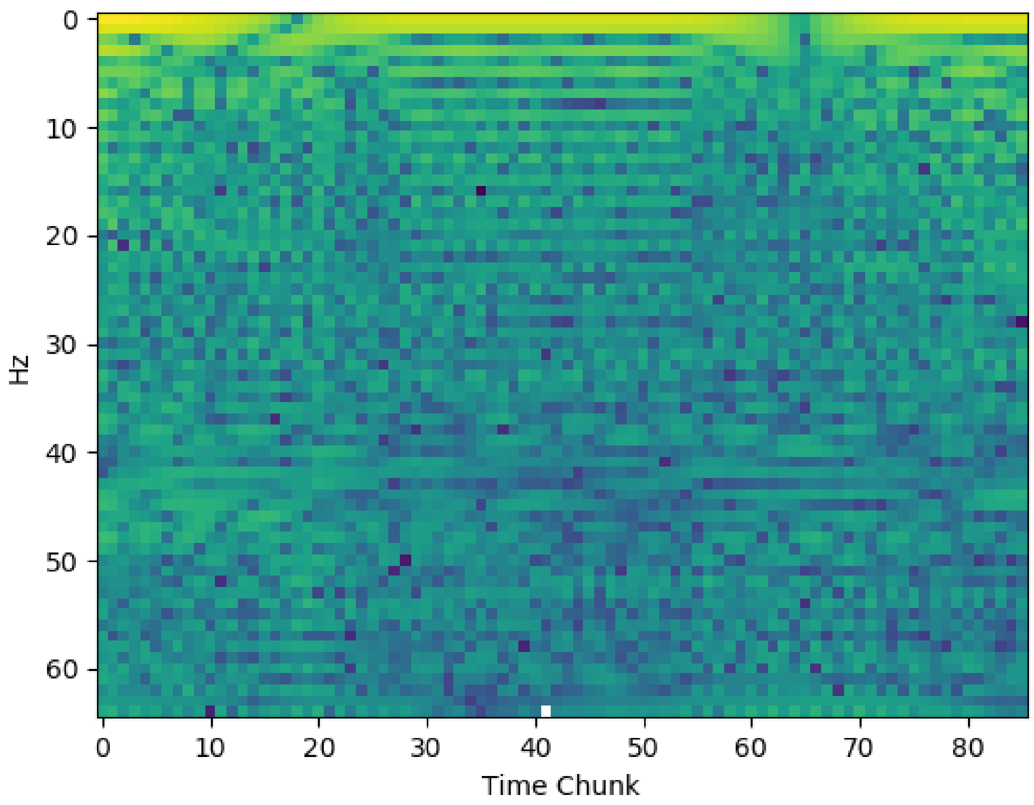
- Can we design a neural network model that is unified to all the tasks?
  - No need for model recalibration
  - Perform as well as domain-specified models
- The issue is
  - Some data suits **temporal domain**: motion data, ambient data
  - Some data suits **frequency domain**: EEG data, wave signals
  - Noise?
  - Sampling rate?

Motivation – Fourier inspired convolution

- **Short time Fourier Transform** conducts **Fourier transform** within a sliding window
  - Fourier weights are explainable and insightful
  - But they are fixed still

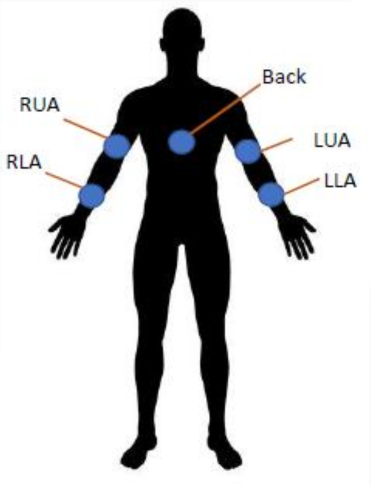
- **Convolution** conducts **linear transform** within a sliding window
  - Convolution weights are learnable
  - But initialized randomly

STFT Spectrogram

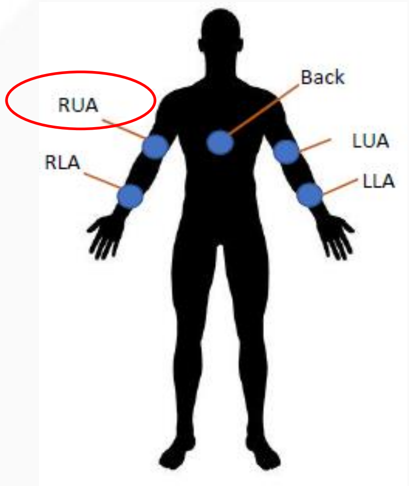


Methodology – Formulation

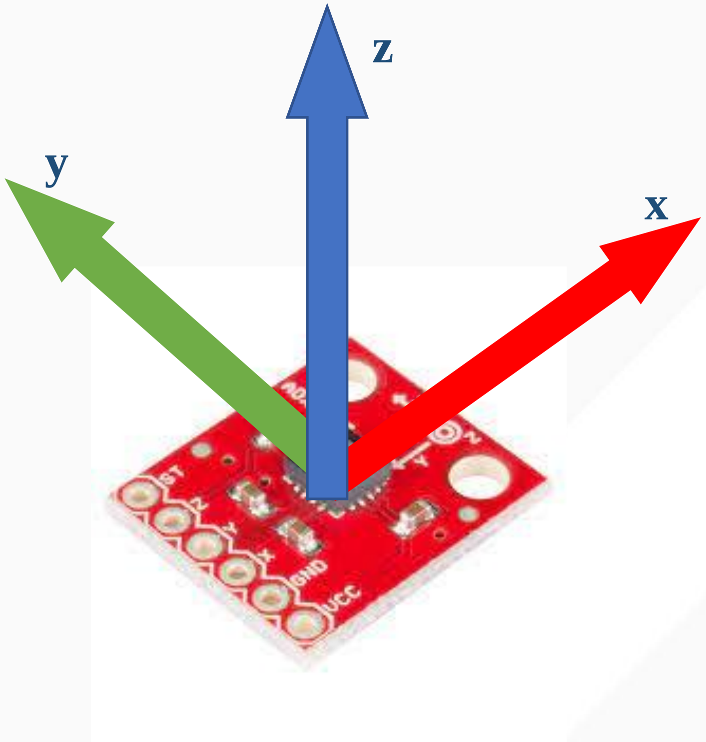
- 5 sets of sensors



- 3 types of sensors
  - 1 Accelerometer
  - 1 Gyroscope
  - 1 Magnetometer



- 3 axes

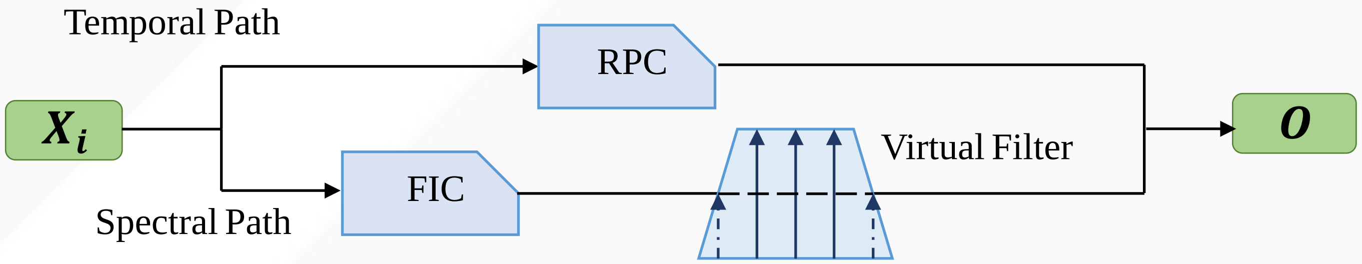


- Each example $\boldsymbol{X} \in \mathbb{R}^{45 \times 256}$ is sampled within fixed time range

Methodology – Temporal Spectral Encoder

UC San Diego

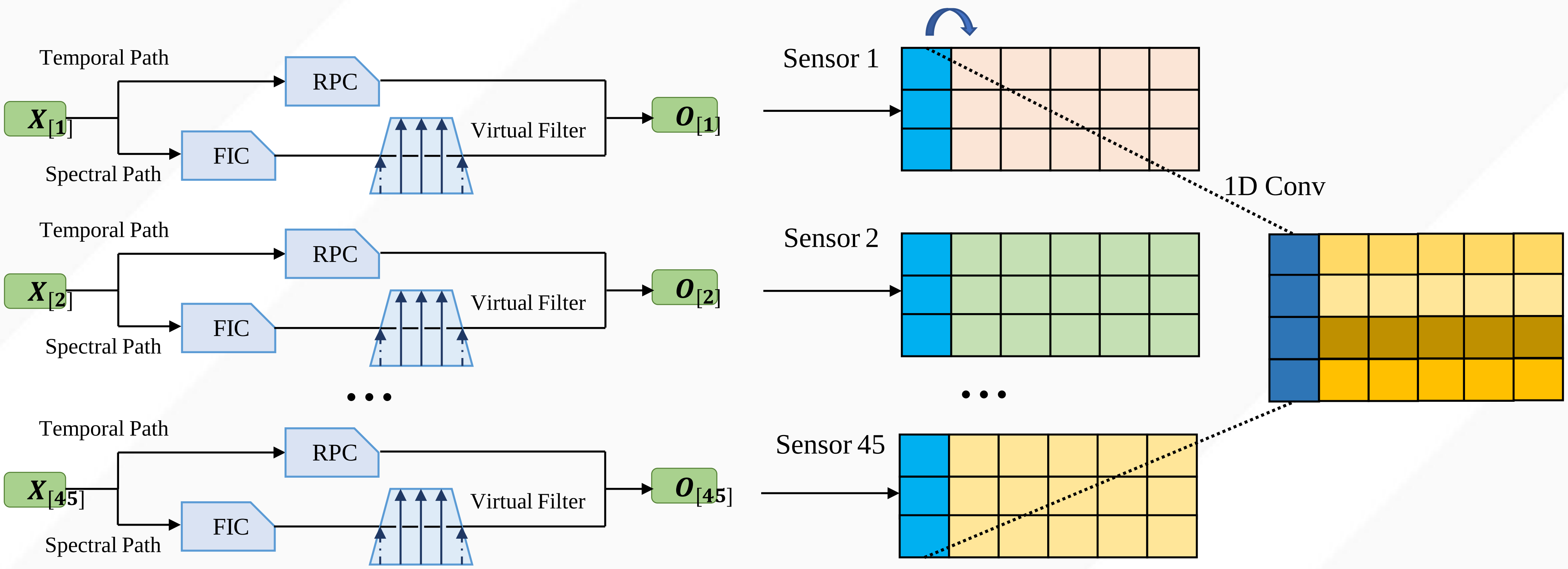
- $\boldsymbol{X_i} \in \mathbb{R}^{256}$ is the i-th channel of $\boldsymbol{X}$



- Fourier Inspired Convolution (FIC)
  - Convolution initialized by Fourier weights
- Virtual Filter
  - Select the most important k virtual frequency channels
- Resolution Preserve Convolution (RPC)
  - Initialized randomly
  - Same size as FIC

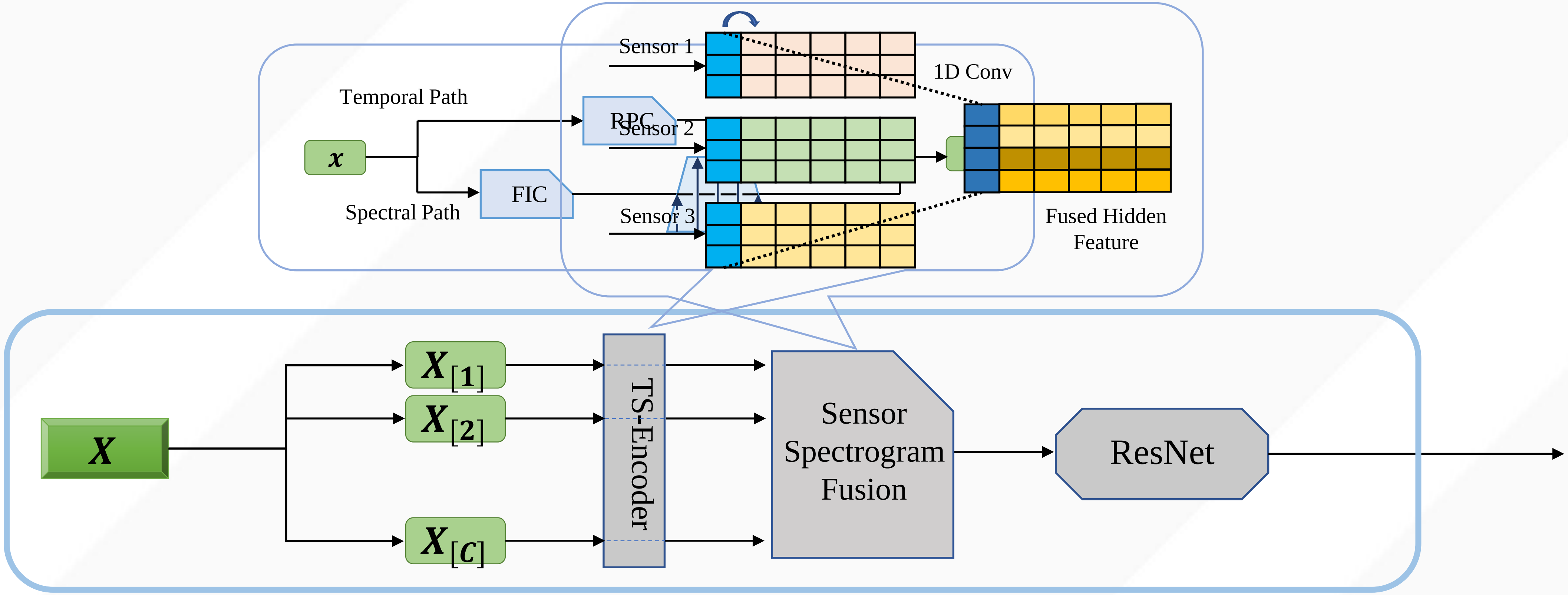
Methodology – Virtual spectrogram fusion

- Channel-wise virtual spectrograms are concatenated and convolved for fusion



Methodology – Overview 1

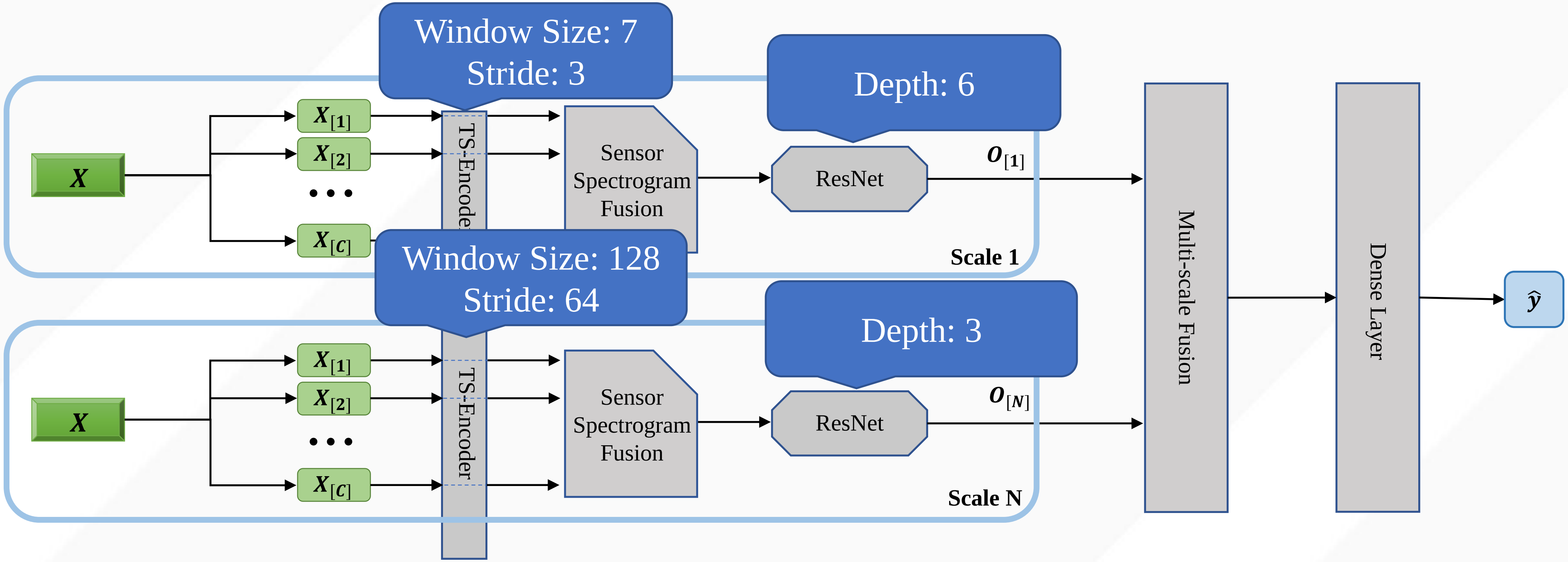
- We then use ResNet[1] to model the fused hidden feature



[1] He, Kaiming, et al. Deep residual learning for image recognition

Methodology – Overview 2

- We adopt the idea of Inception network[2] and design multi-scale parallel branches with different **TS-Encoder** size and **ResNet** depth.



[2] Szegedy, Christian, et al. Inception-v4, inception-resnet and the impact of residual connections on learning

Evaluation – Datasets and metrics

- Motion
  - Motion sensor dataset
  - Activity recognition
- Seizure
  - Brain sculp EEG dataset
  - Seizure detection
- WiFi
  - WiFi dataset
  - Activity recognition
- KETI
  - Building ambient dataset
  - Occupancy detection

Dataset	Sampling Rate	#Instances	#Classes
Motion	30 Hz	11,073	4
Seizure	256 Hz	21,436	2
WiFi	1000 Hz	8,099	7
KETI	0.1 Hz	16,921	2

- Metrics
  - Accuracy
  - F1 score
    - tp : true positive
    - fp: false positive
    - fn: false negative

$$F_1 = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$$

Evaluation – Main result

UC San Diego

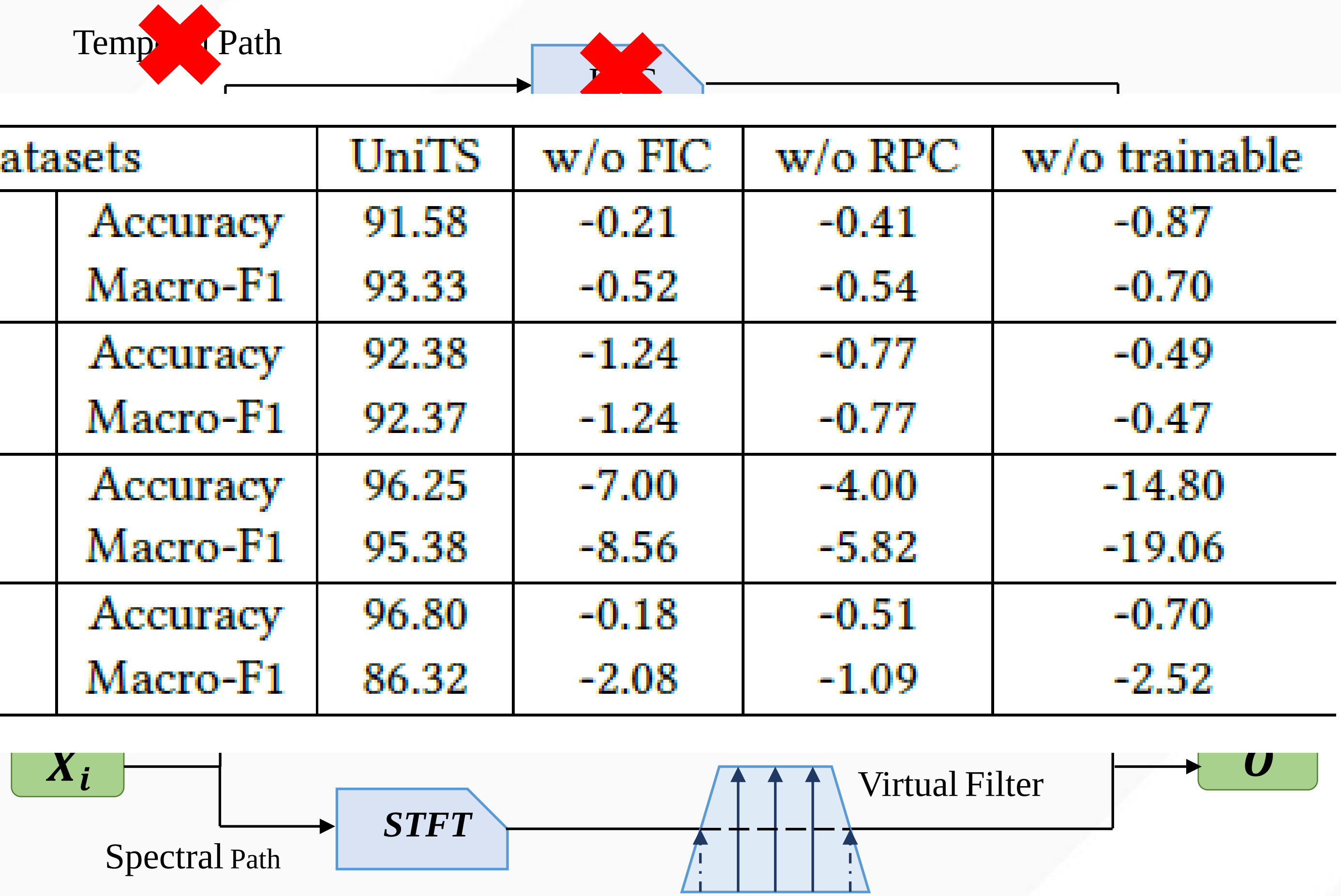
- Our model achieves the best performance in average

Datasets		DWT	ResNet	MaCNN	SenseHAR	STFNets	RFNet-base	THAT	LaxCat	UniTS
Motion	Accuracy	87.43	89.54	88.32	88.90	88.98	90.22	<u>90.35</u>	60.19	**91.58**
	Macro-F1	89.01	91.36	90.32	88.60	88.52	<u>92.28</u>	92.12	41.47	**93.33**
Seizure	Accuracy	90.48	88.98	80.92	85.10	82.82	**92.71**	89.77	87.03	<u>92.38</u>
	Macro-F1	90.44	88.92	80.77	85.10	82.75	**92.70**	89.76	86.95	<u>92.37</u>
WiFi	Accuracy	55.00	88.00	84.09	<u>94.90</u>	79.75	83.88	93.00	74.63	**96.25**
	Macro-F1	47.79	86.30	84.05	<u>92.80</u>	74.29	79.31	90.00	71.09	**95.38**
KETI	Accuracy	95.56	96.05	93.83	96.50	89.18	95.63	<u>96.61</u>	91.28	**96.80**
	Macro-F1	82.14	83.12	73.54	84.90	69.32	82.97	<u>86.21</u>	71.13	**86.32**
average	Accuracy	82.12	90.64	86.79	91.35	85.18	90.61	<u>92.43</u>	78.28	**94.25**
	Macro-F1	77.35	87.43	82.17	87.85	78.72	86.82	<u>89.52</u>	67.66	**91.85**

Bold: Best <u>Underline: Second Best</u>

Evaluation – Ablation study

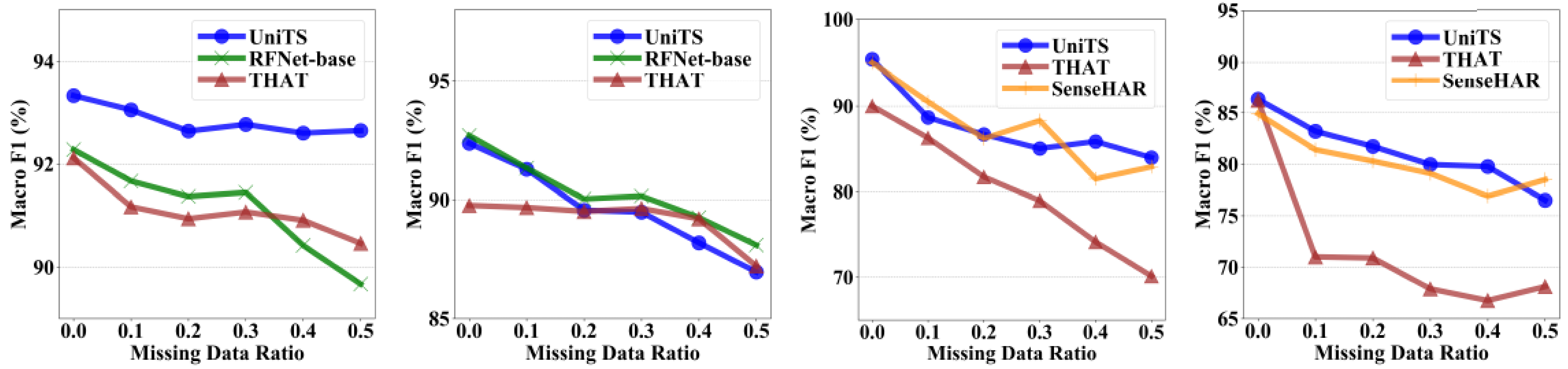
- We consider the following variants for ablation study
- w/o R
- w/o F
- w/o trainable



Datasets		UniTS	w/o FIC	w/o RPC	w/o trainable
Motion	Accuracy	91.58	-0.21	-0.41	-0.87
	Macro-F1	93.33	-0.52	-0.54	-0.70
Seizure	Accuracy	92.38	-1.24	-0.77	-0.49
	Macro-F1	92.37	-1.24	-0.77	-0.47
WiFi	Accuracy	96.25	-7.00	-4.00	-14.80
	Macro-F1	95.38	-8.56	-5.82	-19.06
KETI	Accuracy	96.80	-0.18	-0.51	-0.70
	Macro-F1	86.32	-2.08	-1.09	-2.52

Evaluation – Robustness

- Our model is robust to
  - The change of data sampling rate
  - The change of input example size
  - Random additive noise in data
  - Missing values in data

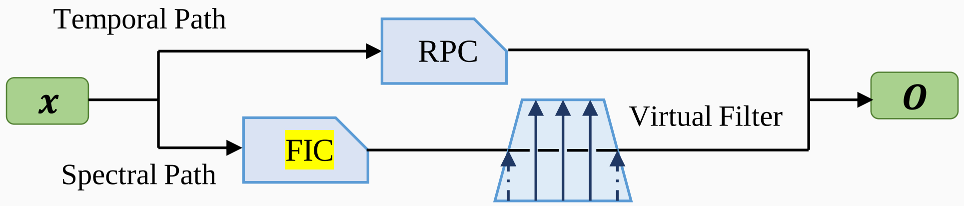


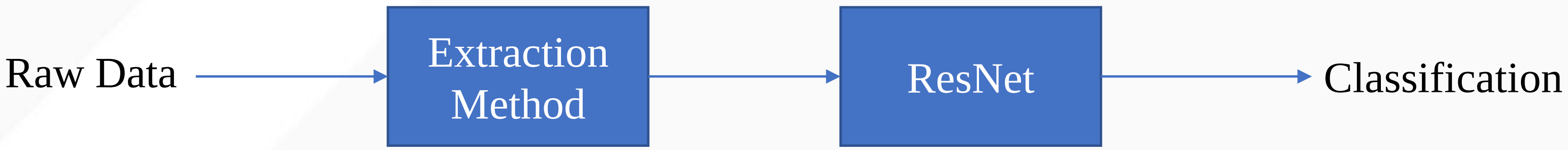
(a) Motion (b) Seizure (c) WiFi (d) KETI

Discussion – FIC for Feature Extraction

UC San Diego

- As a feature extraction module
- The trained FIC model can replace STFT

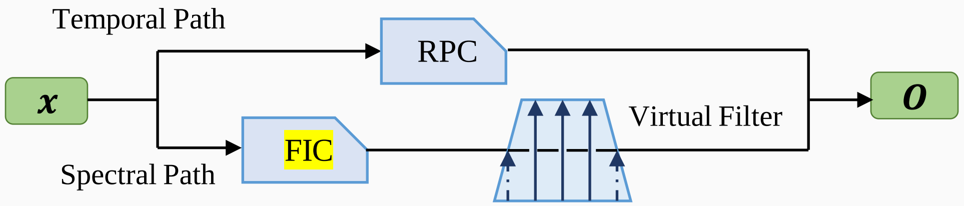


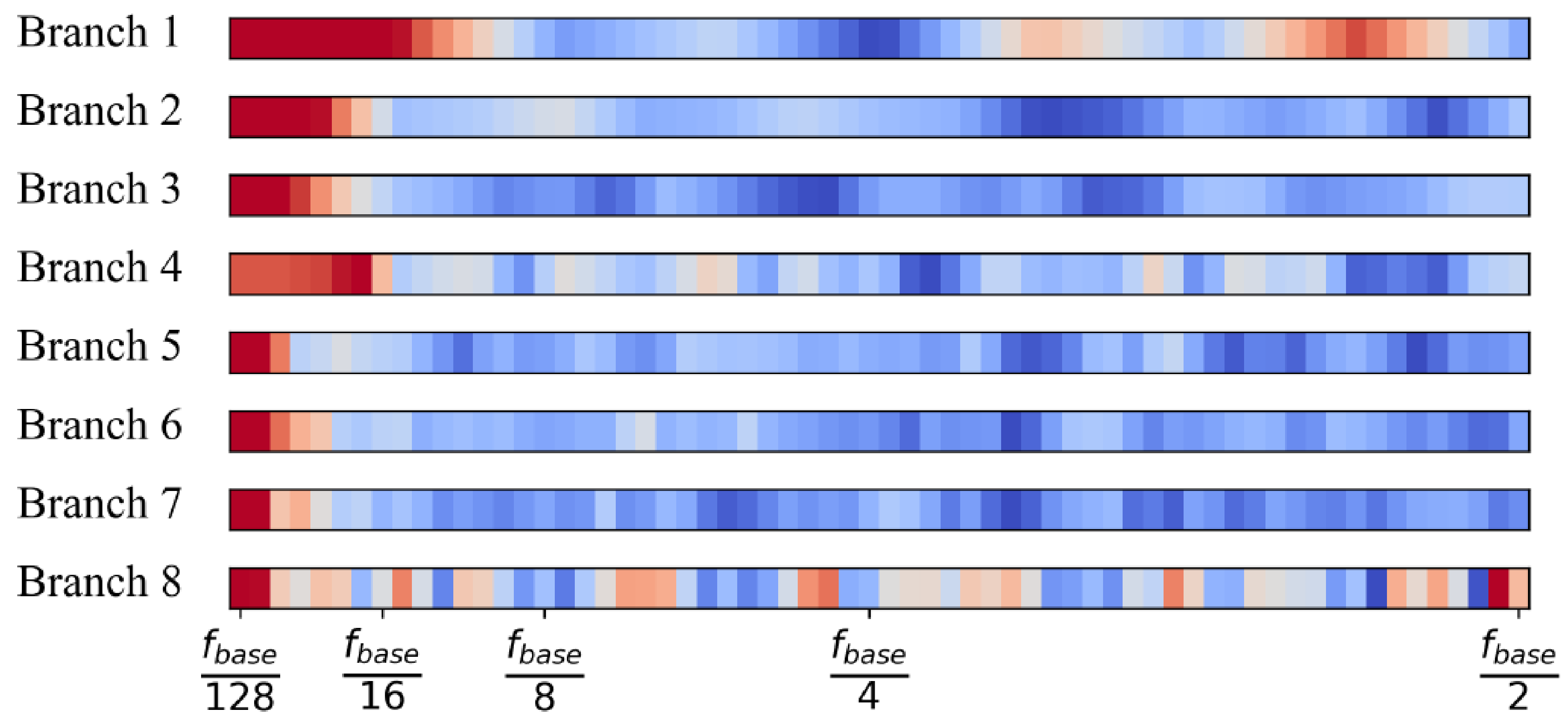


Extraction Method	Motion		WiFi	
	Accuracy	Macro-F1	Accuracy	Macro-F1
STFT	80.50	78.72	55.13	45.34
FIC Layer	**89.40**	**91.53**	**90.88**	**88.96**

Discussion – Explainability

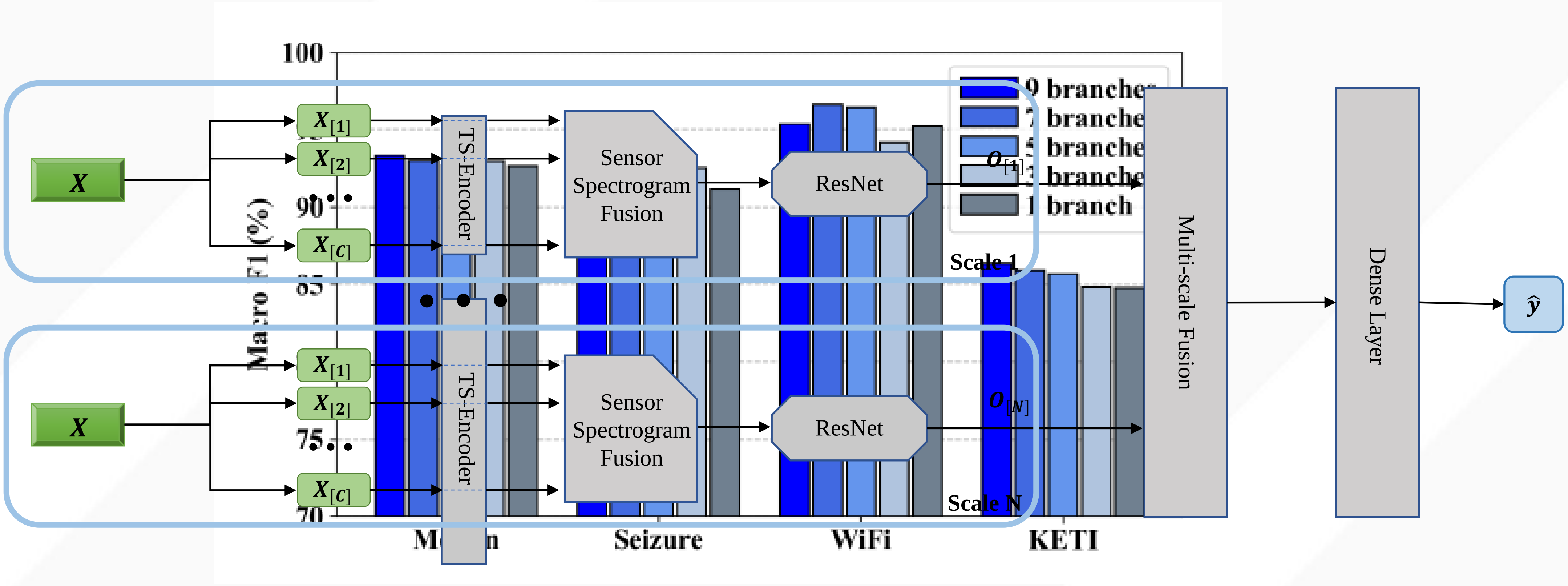
- We study the effect of FIC weights change
- Red indicates FIC **highlights** these channels





Discussion – Hyperparameter tuning

- Multi-scale branches help the model achieves decent performance without the need of exhaustive hyperparameter search



Conclusion

- We proposed UniTS, a Unified framework for sensory Time Series classification
  - TS-Encoder
  - Virtual Spectrogram Fusion
  - Multi-scale
- Evaluated on four different datasets
  - The state-of-the-art performance
  - Ablation study
  - Robustness
- Verified the explainability and hyperparameter tuning issue
  - Use for feature extraction

Thanks for your attention

- https://github.com/Shuheng-Li/UniTS-Sensory-Time-Series-Classification

- Speaker: Shuheng Li
- Email: shl060@ucsd.edu