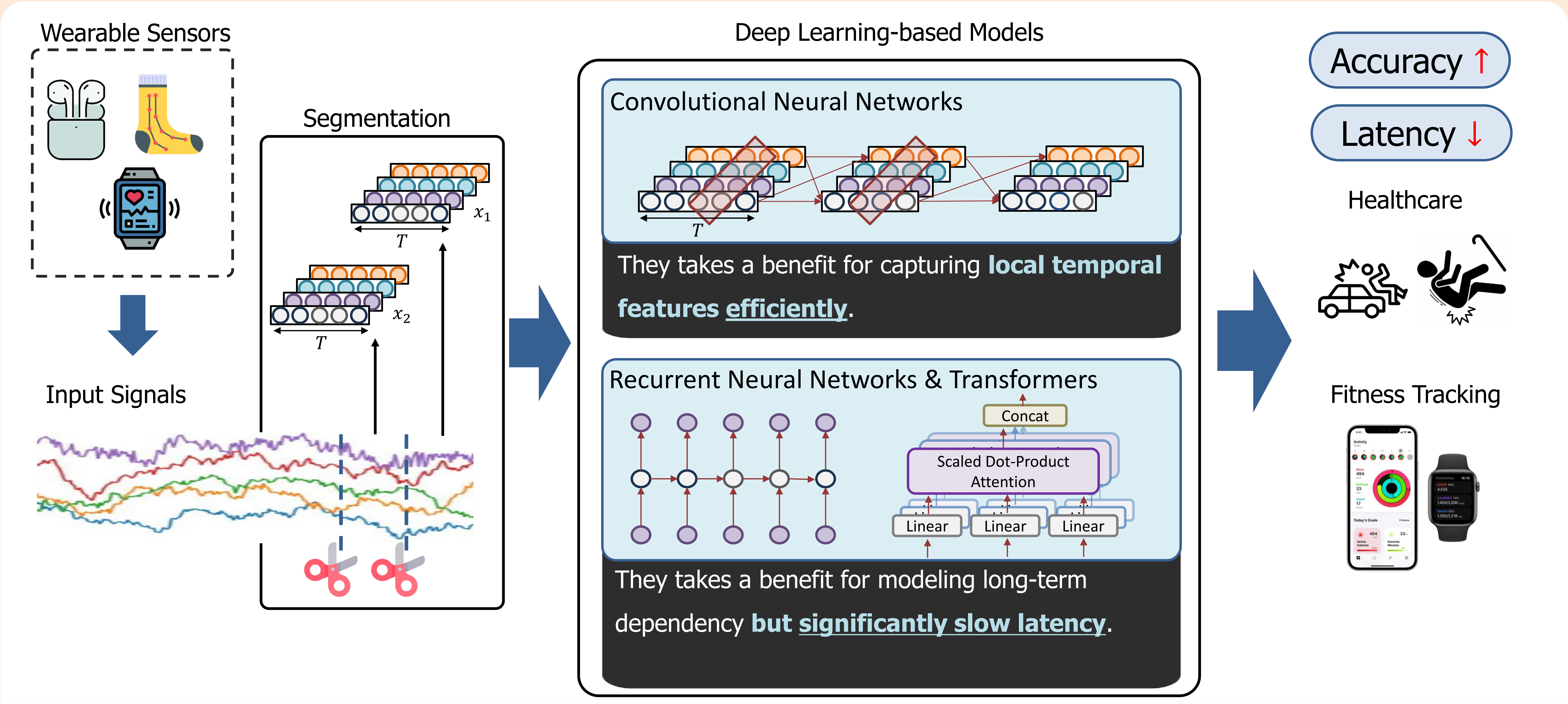
CALANet: Cheap All-Layer Aggregation Network for Human Activity Recognition

Jaegyun Park[1], Dae-Won Kim[1], Jaesung Lee[2,3]

[1]School of Computer Science and Engineering, Chung-Ang University
[2]Department of Artificial Intelligence, Chung-Ang University
[3]AI/ML Innovation Research Center, Chung-Ang University

What is Human Activity Recognition?

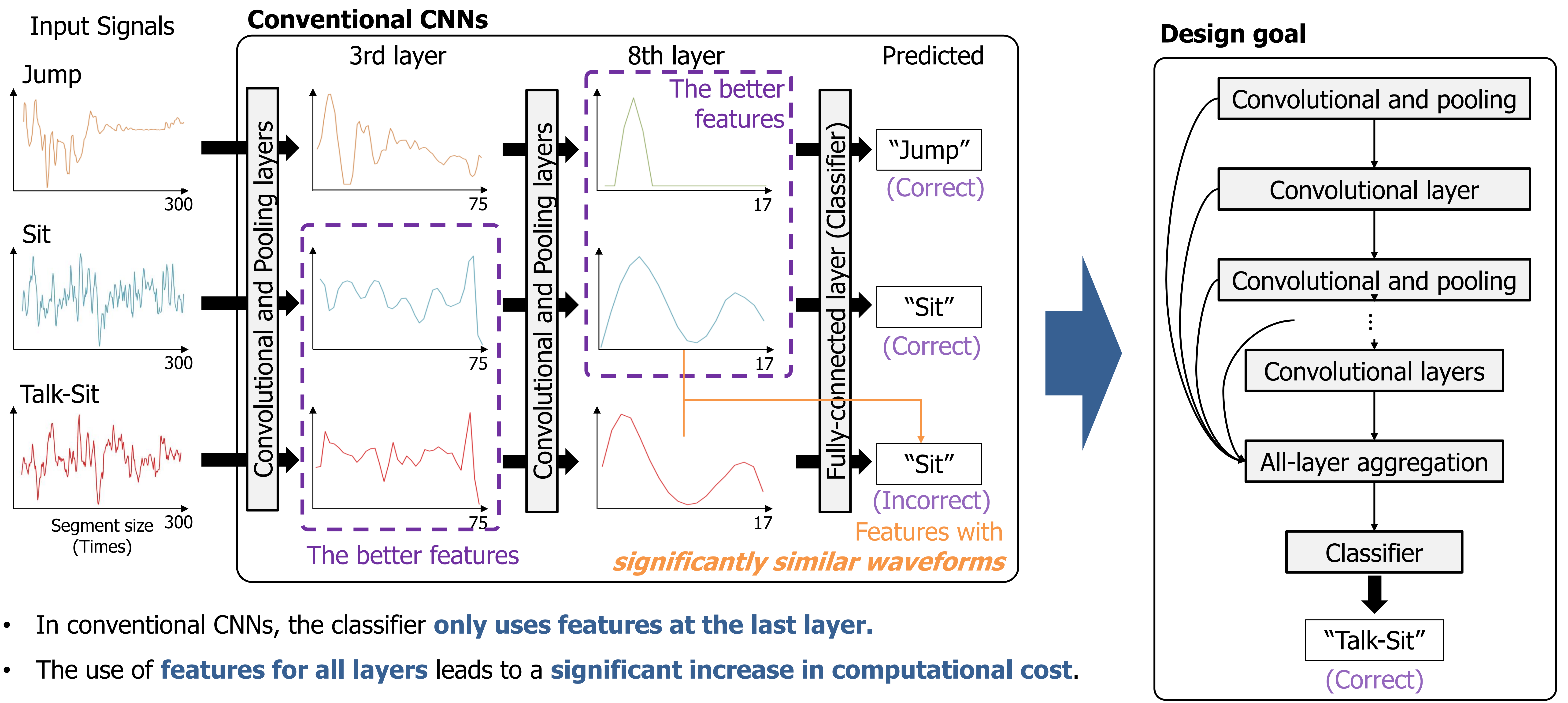
- **Human Activity Recognition (HAR)** aims to record people's behaviors and allows computing systems to monitor, analyze, and assist their daily lives.



◆ We found that CNNs are sufficient to extract meaningful information from the HAR signals with short segmentation lengths.

Motivation

➢ For the HAR, **the loss of detailed information over layers** makes it **challenging to classify the analogous activities**, such as “Sit” and “Talk-Sit” of KU-HAR dataset.

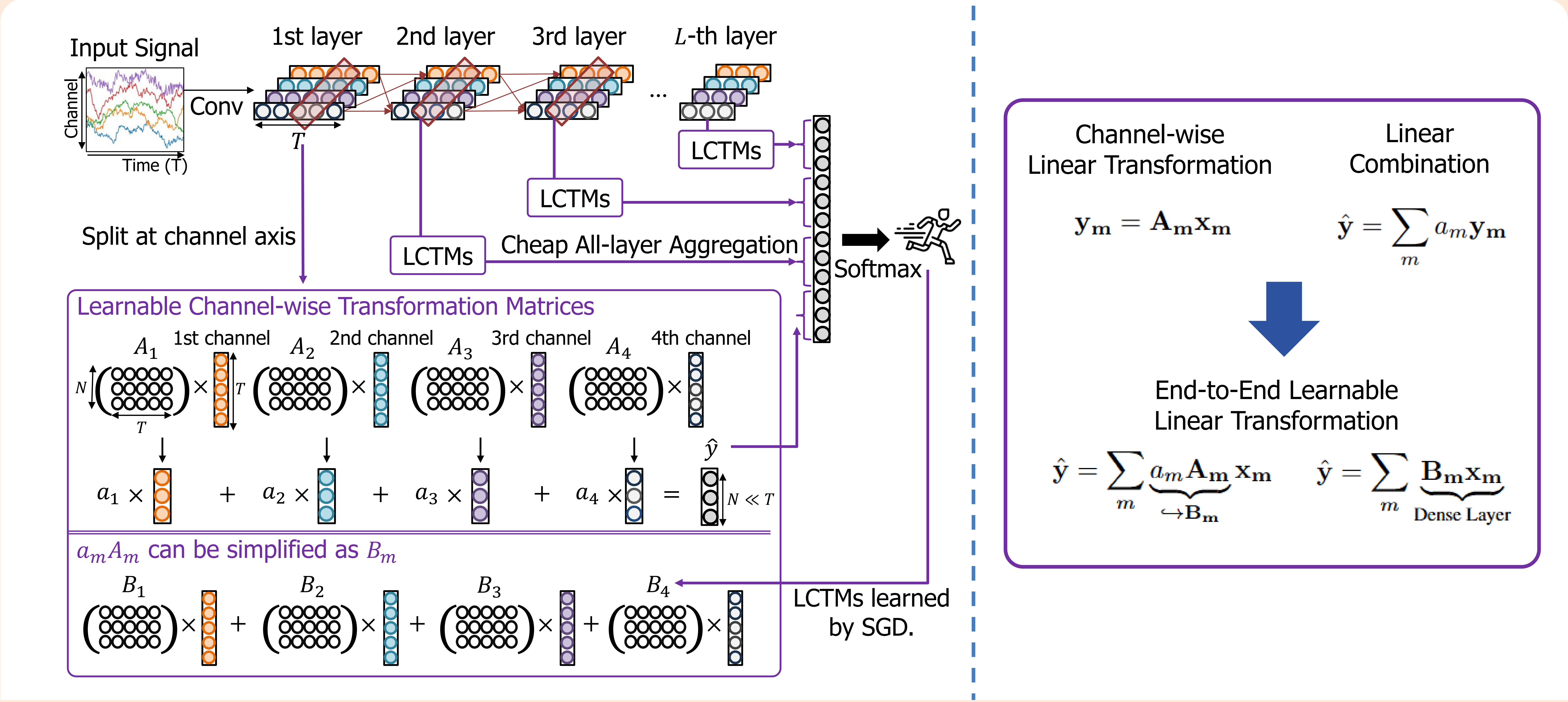


- In conventional CNNs, the classifier **only uses features at the last layer.**
- The use of **features for all layers** leads to a **significant increase in computational cost**.

◆ How can CNNs aggregate all-layer features cheaply?

CALANet: Cheap All-Layer Aggregation Network

- **Learnable Channel-wise Transformation Matrix (LCTM)** compresses global temporal information in each channel at each layer. (Temporal Resolution $T \to N$, $T \gg N$)



◆ How can CNNs aggregate <u>all-layer</u> features <u>cheaply</u>?

CALANet: Cheap All-Layer Aggregation Network

- **S**calable **L**ayer **A**ggregation **P**ool (**SLAP**) allows CALANet to stack layers without increasing computational cost, improving the effectiveness of all-layer aggregation.

- We **formulated the computational cost** of neural networks as **a function of architectural parameters** in asymptotic notation, i.e., time complexity.

  Proposition 1. The time complexity of CNNs is formalized as: $\mathbb{M} \leq \mathbb{N}(L-1) \Longrightarrow \mathcal{O}(\mathbb{T}\mathbb{D}_k\mathbb{N}^2L)$.

- SLAP is achieved by omitting L from **Proposition 1.**

  Theorem 1. The time complexity of CALANet is reduced to: $\mathcal{O}(\mathbb{T}\mathbb{D}_k\mathbb{N}^2)$.

  - **Corollary 1.** The time complexity of CALANet is equivalent to the shallow CNNs with $L \geq 2$.
  - **Corollary 2.** The time complexity of CALANet is equivalent to the shallow CNNs with $L = 1$ if $\mathbb{M} \approx \mathbb{N}$.

◆ LCTMs and SLAP allow CNNs to aggregate features for all layers while maintaining an efficiency of CNNs.

Experimental Results

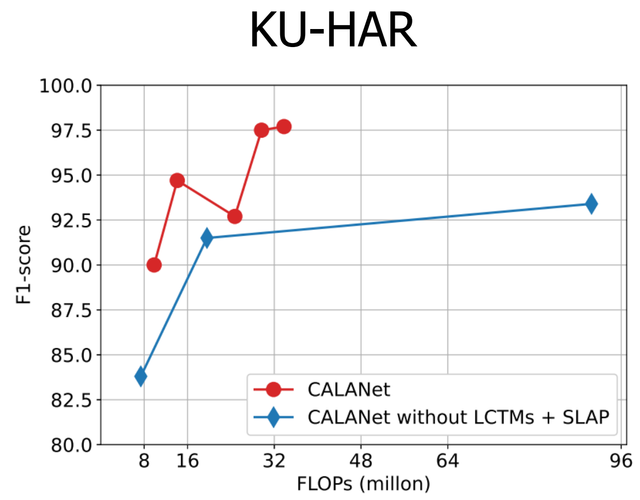
➢ Comparison to SoTA

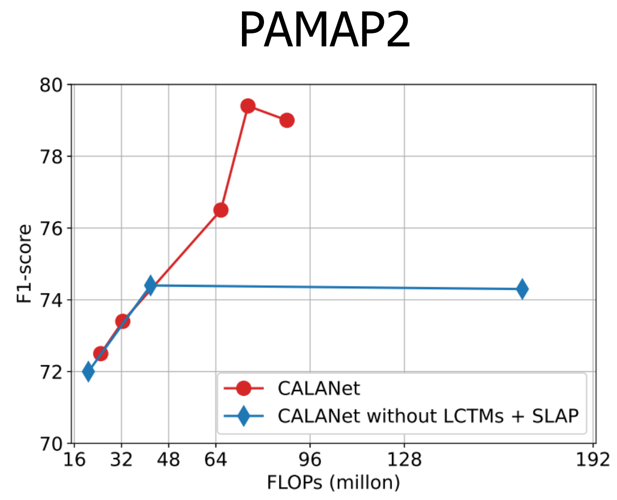
Model	UCI-HAR F1	UCI-HAR FLOPs	UniMiB-SHAR F1	UniMiB-SHAR FLOPs	DSADS F1	DSADS FLOPs	OPPORTUNITY F1	OPPORTUNITY FLOPs
CALANet (**Ours**)	96.1	**7.6M**	78.3	**8.8M**	90.0	**8.5M**	81.6	**19.3M**
Shallow ConvNet [23]	92.5▼	17.9M	72.2▼	18.2M	85.6▼	48.5M	79.5▼	74.3M
RepHAR [49]	95.1▼	31.8M	71.6▼	37.3M	85.5▼	32.9M	80.0▼	26.0M
DeepConvLSTM [35]	91.4▼	67.2M	71.6▼	80.4M	85.5▼	68.3M	62.0▼	50.4M
Bi-GRU-I [50]	94.6▼	46.1M	75.2▼	54.0M	85.6▼	48.7M	77.2▼	39.8M
RevAttNet [40]	95.1▼	143.1M	76.7▼	168.7M	87.6▼	140.2M	78.6▼	101.5M
IF-ConvTransformer [63]	95.4	209.8M	77.0▼	183.5M	87.5▼	628.4M	82.2	986.2M
T-ResNet [54, 12]	95.3	123.2M	76.5▼	145.5M	87.3▼	125.8M	80.9	96.9M
T-FCN [54, 12]	95.8	68.9M	76.9▼	80.6M	86.7▼	76.1M	76.2▼	65.8M
MILLET [11]	94.7▼	111.6M	81.4△	129.9M	84.3▼	132.8M	82.3	125.0M
DSN [55]	95.4	270.8M	79.8	320.0M	86.4▼	265.7M	71.8▼	192.1M

Model	KU-HAR F1	KU-HAR FLOPs	PAMAP2 F1	PAMAP2 FLOPs	REALDISP F1	REALDISP FLOPs
CALANet (**Ours**)	97.5	**29.6M**	79.4	**74.9M**	98.2	**56.7M**
Shallow ConvNet [23]	77.9▼	41.6M	67.4▼	151.8M	95.9▼	209.9M
RepHAR [49]	93.4▼	74.4M	73.0▼	131.9M	94.7▼	72.7M
DeepConvLSTM [35]	93.5▼	169.1M	77.3▼	303.9M	91.7▼	156.9M
Bi-GRU-I [50]	94.9▼	108.0M	71.0▼	194.1M	96.1▼	111.3M
RevAttNet [40]	97.7	335.3M	79.7	573.5M	98.5	282.1M
IF-ConvTransformer [63]	96.4▼	491.7M	80.1	1.7G	97.4	3.0G
T-ResNet [54, 12]	95.0▼	290.0M	71.4▼	506.1M	96.0▼	270.1M
T-FCN [54, 12]	92.5▼	161.7M	72.5▼	298.5M	95.9▼	184.5M
MILLET [11]	97.8	262.5M	80.2	509.5M	95.1▼	352.9M
DSN [55]	97.1	634.8M	68.8▼	1.08G	97.5	532.7M

➢ The breakdown effect of CALANet

Networks	L	KU-HAR F1	KU-HAR FLOPs	PAMAP2 F1	PAMAP2 FLOPs
CALANet with LCTMs + SLAP	9	97.5	29.7M	79.4	74.9M
CALANet with LCTMs only	4	93.8▼	60.0M	73.1▼	113.3M
CALANet with ALA only	4	95.0▼	577.9M	72.8▼	1.7G





➢ Real-Time Activity prediction

Model	Inference Time (ms / window) Min	Mean	Max
CALANet	1.59ms	2.25ms	3.40ms
Shallow ConvNet	1.57ms	2.15ms	3.48ms

Thanks!

Contact Us:

Linkedin | Scholar | AI Lab. | AutoML Lab.

Email: jgp0566.cau@gmail.com