ST-CA2: Lightweight CSI Input Refinement for Efficient and Tiny Wi-Fi Sensing Models

Jiangyang Liu, Fei Ge*, Gaoming Kang
School of Computer Science, Central China Normal University, Wuhan, China
Emails: liujiangyang@mails.ccnu.edu.cn, feige@ccnu.edu.cn, kanggaoming@mails.ccnu.edu.cn
*Corresponding author

***Abstract*—Wi-Fi sensing on resource-constrained platforms calls for efficient and tiny neural networks, yet CSI pipelines face a recurring trade-off between preserving the full multi-antenna CSI tensor and compressing the input by simple heuristics. We propose ST-CA2 (SpatioTemporal Channel–Antenna Attention), a lightweight input refinement module that performs subcarrier-wise channel attention and antenna-pair-wise attention to output a compact C-by-T representation. Under a fixed downstream recognizer and in a door-opening person-identification setting, ST-CA2 achieves 99.20% OA with 6.208 GFLOPs and 2.006 GB peak GPU memory, compared with 97.20% OA with 53.232 GFLOPs and 8.601 GB for full-input. These results indicate a favorable accuracy–efficiency trade-off for compact CSI input refinement in the studied setting. We also provide analyses under signal-plus-noise assumptions and evaluate temporal statistical priors (GTA, variance, and RMS) for weight generation.**

***Index Terms*—TinyML, efficient and tiny neural networks, Wi-Fi sensing, CSI, input refinement, attention, subcarrier reweighting, antenna-pair fusion**

I. Introduction

Wi-Fi sensing provides a non-intrusive alternative for human-centric recognition in smart environments. Among wireless modalities, Channel State Information (CSI) offers fine-grained measurements at the subcarrier level in MIMO-OFDM systems, enabling a wide range of recognition tasks, including activity recognition and person identification [1], [2]. However, practical CSI pipelines are often challenged by the high dimensionality and strong redundancy of CSI acquired across multiple antenna pairs and subcarriers, as well as noise and multipath-induced distortions. These factors complicate direct full-input learning and motivate principled input refinement methods.

In many deployments, CSI sensing models are expected to run under strict memory/latency constraints. Even when the downstream recognizer is compact, the CSI input and its preprocessing can dominate computation and memory, limiting the use of efficient and tiny neural networks. From the perspective of input usage, existing CSI-based recognition pipelines can be grouped into two categories:

1) *Full-input usage*: the full CSI tensor is fed into the downstream recognizer. This preserves information but incurs redundancy, noise propagation, and compute/memory overhead that scales with the number of antenna pairs and subcarriers.
2) *Input compression*: the CSI input is reduced before learning, e.g., by arithmetic averaging across antenna pairs or random selection. These heuristics improve efficiency but can discard discriminative components.

To bridge this tension, we propose ST-CA2 (SpatioTemporal Channel–Antenna Attention), a lightweight input refinement module that sequentially applies subcarrier-wise channel attention (CA) and antenna-pair-wise attention (AA). CA generates subcarrier weights from temporal statistical priors, while AA performs quality-aware antenna-pair fusion to reduce the antenna-pair dimension.

The contributions of this paper are summarized as follows:

- We propose ST-CA2, a two-stage attention module for subcarrier reweighting and antenna-pair fusion, yielding a compact $C \times T$ representation from an $A \times C \times T$ CSI tensor.
- We provide mathematical analyses for CA subcarrier reweighting and AA antenna-pair fusion under common signal-plus-noise assumptions.
- We evaluate the accuracy–efficiency trade-off under a fixed downstream recognizer against heuristic input compression baselines, and study different temporal priors (GTA, variance, and RMS) for weight generation.

Data/code availability: Due to privacy constraints, the door-opening dataset collected in this work cannot be publicly released. To support reproducibility, we will release the implementation of ST-CA2, the preprocessing settings, and the fold-construction procedure upon acceptance, and will add the repository link in the camera-ready version: `[URL to be added upon acceptance]`.

The remainder of this paper is organized as follows. Section II reviews related work and Section III presents the proposed method and analysis. Section IV reports experimental results and discussion, and Section V concludes the paper.

II. Related Work

A. CSI sensing and preprocessing

Wi-Fi sensing commonly relies on either RSSI or CSI. Compared with RSSI, CSI provides fine-grained channel responses at the subcarrier level and is thus more sensitive to human-induced propagation changes [3]. In a MIMO-OFDM system, CSI can be described by

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{X}_i + \mathbf{N}_i \tag{1}$$

where $\mathbf{H}_i$ denotes the subcarrier-wise channel response and $\mathbf{N}_i$ denotes measurement noise [4]. Existing CSI-based recognition methods typically rely on preprocessing (e.g., denoising, segmentation, or time–frequency transforms) before feeding CSI into learning models [5], [6].

B. CSI input usage: full-input vs compression and the gap

With multiple antenna pairs and subcarriers, CSI contains rich diversity but also redundancy and noise. Full-input usage directly feeds the full tensor into a recognizer, which preserves information but increases compute/memory and can amplify nuisance variations. Input compression reduces the antenna/subcarrier dimensions by heuristics such as arithmetic mean fusion or random selection; these strategies are lightweight but can attenuate discriminative temporal variations (feature fading) or yield unstable performance.

Beyond heuristic compression, lightweight learning-based reweighting modules such as SE blocks [7] and CBAM-style attention [8] are widely used to improve compact models. However, such modules are typically inserted into learned feature backbones after feature extraction, whereas ST-CA2 is designed as an input-stage module that operates directly on raw CSI organization and specifically targets subcarrier reweighting and antenna-pair fusion before a fixed downstream recognizer.

For efficient and tiny networks, reducing the *input* cost is as important as compressing the backbone. The above tension motivates ST-CA2: a lightweight input refinement and fusion module that performs (i) subcarrier-wise reweighting to suppress noisy/redundant components and (ii) antenna-pair-wise fusion to reduce the antenna-pair dimension. ST-CA2 outputs a compact $C \times T$ representation, enabling plug-and-play integration with compact downstream recognizers.

III. Proposed Method

ST-CA2 is a lightweight module for CSI input refinement and fusion. It targets redundancy/noise across subcarriers and redundancy/imbalance across antenna pairs, with the goal of reducing the input burden for efficient and tiny downstream models. The module applies CA for subcarrier reweighting and AA for antenna-pair fusion, producing a compact representation.

Fig. 1 summarizes the computation flow of ST-CA2. Given $X \in \mathbb{R}^{A\times C\times T}$, CA performs element-wise gating $X^C_{a,c,t} = X_{a,c,t} \odot W^C_{a,c}$, and AA fuses antenna pairs to obtain $X_{\text{fuse}}(c,t) = \sum_{a=1}^{A} \alpha_{c,a} X^C_{a,c,t} \in \mathbb{R}^{C\times T}$.

A. Notation and problem setup

For each antenna pair $a \in \{1,\dots,A\}$, we denote the CSI measurement as a stack of C subcarrier time series:

$$X_a \in \mathbb{R}^{C\times T},$$

where each row $X_{a,c,:}$ is the length-T temporal sequence of subcarrier c. The full multi-antenna-pair CSI input is formed by stacking $\{X_a\}_{a=1}^{A}$:

$$X = \{X_a\}_{a=1}^{A} \in \mathbb{R}^{A\times C\times T}.$$

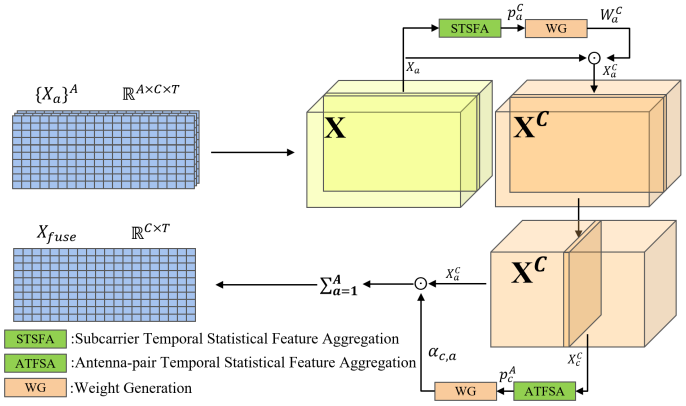

Fig. 1. Overview of ST-CA2.

ST-CA2 outputs a compact fused representation

$$X_{\text{fuse}} \in \mathbb{R}^{C\times T},$$

which reduces the antenna-pair dimension while preserving subcarrier-wise temporal structure.

ST-CA2 uses a temporal statistical operator $S(\cdot \mid M)$ to extract a low-dimensional prior from a time series, where $M \in \{\text{GTA}, \text{Var}, \text{RMS}\}$ denotes the choice of statistic. We do not assume a fixed M; its impact is evaluated empirically in Section IV-D.

B. Stage-I: subcarrier-wise Channel Attention (CA)

Channel Attention aims to highlight discriminative subcarriers by learning a subcarrier weight vector for each antenna pair. For antenna pair a and subcarrier c, we first extract a temporal statistical prior:

$$p^C_{a,c} = S\left(X_{a,c,1:T} \mid M\right), \qquad p^C_a \in \mathbb{R}^C. \tag{2}$$

Then a two-layer MLP produces subcarrier weights:

$$W^C_a = \sigma\left(W^C_2 \cdot \delta\left(W^C_1 \cdot p^C_a + b^C_1\right) + b^C_2\right), \qquad W^C_a \in (0,1)^C, \tag{3}$$

where $\delta(\cdot)$ is a ReLU-type nonlinearity and $\sigma(\cdot)$ is the sigmoid function. The original CSI is then reweighted (element-wise gating):

$$X^C_{a,c,t} = X_{a,c,t} \odot W^C_{a,c}, \qquad X^C \in \mathbb{R}^{A\times C\times T}. \tag{4}$$

Here $\odot$ denotes element-wise multiplication with broadcasting of $W^C_{a,c}$ along the temporal dimension t.

C. Stage-II: antenna-pair-wise Attention and Fusion (AA)

Antenna Attention aims to fuse complementary information across antenna pairs and reduce the antenna-pair dimension. For each subcarrier c, we extract an antenna-pair prior from the CA-refined tensor:

$$p^A_{c,a} = S\left(X^C_{a,c,1:T} \mid M\right), \qquad p^A_c \in \mathbb{R}^A. \tag{5}$$

We then generate antenna-pair weights:

$$W^A_c = \sigma\left(W^A_2 \cdot \delta\left(W^A_1 \cdot p^A_c + b^A_1\right) + b^A_2\right), \qquad W^A_c \in (0,1)^A. \tag{6}$$

Algorithm 1 ST-CA2 Fusion Module (Forward Pass)

Require: CSI tensor $X \in \mathbb{R}^{A\times C\times T}$, statistic $S(\cdot \mid M)$, MLP parameters Θ_C, Θ_A, small constant $\varepsilon > 0$
Ensure: Fused representation $X_{\text{fuse}} \in \mathbb{R}^{C\times T}$
{Stage-I: subcarrier-wise CA}
for antenna pair $a = 1$ to A **do**
  $p_a^C[c] \leftarrow S(X_{a,c,1:T} \mid M)$ for $c = 1, \ldots, C$
  $W_a^C \leftarrow \mathrm{MLP}_C(p_a^C; \Theta_C)$ {Sigmoid output in $(0,1)^C$}
  $X_{a,:,:}^C \leftarrow X_{a,:,:} \odot W_a^C$ {broadcast along T}
end for
{Stage-II: antenna-pair AA and fusion}
for subcarrier $c = 1$ to C **do**
  $p_c^A[a] \leftarrow S(X_{a,c,1:T}^C \mid M)$ for $a = 1, \ldots, A$
  $W_c^A \leftarrow \mathrm{MLP}_A(p_c^A; \Theta_A)$
  $\alpha_{c,a} \leftarrow W_{c,a}^A / (\sum_{j=1}^A W_{c,j}^A + \varepsilon)$
  $X_{\text{fuse}}[c,:] \leftarrow \sum_{a=1}^A \alpha_{c,a} X_{a,c,:}^C$
end for
return X_{fuse}

To perform fusion, we normalize weights across antenna pairs to obtain $\alpha_{c,a}$ and compute the fused output:

$$\alpha_{c,a} = \frac{W_{c,a}^A}{\sum_{j=1}^A W_{c,j}^A + \varepsilon}, \qquad X_{\text{fuse}}(c,t) = \sum_{a=1}^A \alpha_{c,a} X_{a,c,t}^C. \tag{7}$$

The output $X_{\text{fuse}} \in \mathbb{R}^{C\times T}$ preserves subcarrier-wise temporal structure while reducing the antenna-pair dimension from A to 1.

D. Implementation remarks

Eq. (7) enforces $\alpha_{c,a} \geq 0$ and $\sum_{a=1}^A \alpha_{c,a} \approx 1$, i.e., fusion is (approximately) a convex combination across antenna pairs for each subcarrier. $\varepsilon > 0$ is a small constant for numerical stability; if $\varepsilon \ll \sum_j W_{c,j}^A$, then $\sum_a \alpha_{c,a}$ is close to 1. In Fig. 2, Top-k and Bottom-k are used only for visualization. Algorithm 1 summarizes the forward computation.

E. Theoretical properties

This subsection provides analyses for (i) subcarrier reweighting in CA and (ii) quality-aware fusion in AA.

1) Channel attention analysis:

Proposition 1 (Gradient-aligned tendency)**.** *Let $g(\cdot)$ denote any differentiable downstream recognizer and $\ell(\cdot,\cdot)$ be a differentiable loss. Consider the expected risk*

$$\mathcal{L} = \mathbb{E}_{(X,y)}\big[\ell(g(X^C), y)\big],$$

where X^C depends on W^C through Eq. (4). Then for any antenna pair a and subcarrier c,

$$\frac{\partial \mathcal{L}}{\partial W_{a,c}^C} \;=\; \mathbb{E}_{(X,y)}\left[\sum_{t=1}^T X_{a,c,t} \frac{\partial \ell}{\partial X_{a,c,t}^C}\right]. \tag{8}$$

Proof. By chain rule, $\partial \ell / \partial W_{a,c}^C = \sum_t (\partial \ell / \partial X_{a,c,t}^C)(\partial X_{a,c,t}^C / \partial W_{a,c}^C)$, and $\partial X_{a,c,t}^C / \partial W_{a,c}^C = X_{a,c,t}$. Taking expectation yields Eq. (8). □

Discussion. Eq. (8) characterizes the gradient with respect to the *gate variable* $W_{a,c}^C$. When W^C is parameterized by an MLP (Eq. (3)), gradients to the MLP parameters follow by chain rule, e.g., $\partial \mathcal{L} / \partial \Theta_C = \sum_{a,c} (\partial \mathcal{L} / \partial W_{a,c}^C)(\partial W_{a,c}^C / \partial \Theta_C)$. This proposition is used to interpret the optimization signal acting on the learned gating, rather than to fully characterize training dynamics for a specific choice of the statistic $S(\cdot)$. The equality $\partial \mathbb{E}[\ell] / \partial W = \mathbb{E}[\partial \ell / \partial W]$ holds under standard regularity conditions (e.g., $\ell \circ g$ differentiable in W and dominated by an integrable function) that allow exchanging differentiation and expectation.

Proposition 2 (Statistical prior under a signal-plus-noise model)**.** *Assume a subcarrier observation model for a fixed antenna pair a and subcarrier c:*

$$X_{a,c,t} = s_{t,c}(y) + n_{a,c,t}, \tag{9}$$

where $s_{t,c}(y)$ is a class-dependent signal component and $n_{a,c,t}$ is noise satisfying $\mathbb{E}[n_{a,c,t} \mid y] = 0$ and $\mathrm{Var}(n_{a,c,t} \mid y) = \sigma_{a,c}^2$ (assumed not varying with t). Consider the RMS prior $S(\cdot) = \mathrm{RMS}$:

$$p_{a,c}^{C,\mathrm{RMS}} = \sqrt{\frac{1}{T}\sum_{t=1}^T X_{a,c,t}^2}.$$

Then

$$\mathbb{E}\Big[\big(p_{a,c}^{C,\mathrm{RMS}}\big)^2 \mid y\Big] \;=\; \frac{1}{T}\sum_{t=1}^T s_{t,c}(y)^2 + \sigma_{a,c}^2. \tag{10}$$

Proof. Expanding X^2 and using $\mathbb{E}[n \mid y] = 0$ yields $\mathbb{E}[X^2 \mid y] = s^2(y) + \sigma^2$; averaging over t gives Eq. (10). □

Discussion. Eq. (10) relates the RMS prior to the signal energy and noise variance (note it is stated for $\mathbb{E}[(\mathrm{RMS})^2]$). If $\sigma_{a,c}^2$ varies with t, the noise term becomes $\frac{1}{T}\sum_t \sigma_{a,c,t}^2$. CA maps p^C to W^C through Eq. (3); the choice of $S(\cdot)$ is evaluated in Section IV-D. We emphasize that this proposition analyzes $\mathbb{E}[(\mathrm{RMS})^2]$, not $\mathbb{E}[\mathrm{RMS}]$; in general $\mathbb{E}[\mathrm{RMS}] \neq \sqrt{\mathbb{E}[(\mathrm{RMS})^2]}$ due to Jensen's inequality.

2) Antenna attention analysis:

Proposition 3 (Optimal fusion and a variance-based reliability proxy)**.** *For a fixed (t,c), assume a standard multi-sensor observation model:*

$$X_{a,c,t}^C = u_{t,c} + \varepsilon_{a,c,t}, \qquad \mathbb{E}[\varepsilon_{a,c,t}] = 0, \qquad \mathrm{Var}(\varepsilon_{a,c,t}) = \sigma_{a,c}^2, \tag{11}$$

where $u_{t,c}$ denotes the latent class-relevant response and $\varepsilon_{a,c,t}$ are independent across antenna pairs a. Consider linear unbiased fusion $\hat{u}_{t,c} = \sum_{a=1}^A \alpha_{c,a} X_{a,c,t}^C$ with $\sum_a \alpha_{c,a} = 1$. The weights minimizing $\mathrm{Var}(\hat{u}_{t,c} - u_{t,c})$ are

$$\alpha_{c,a}^\star = \frac{\sigma_{a,c}^{-2}}{\sum_{j=1}^A \sigma_{j,c}^{-2}}. \tag{12}$$

If, in addition, $u_{t,c}$ is approximately constant within the fusion window (i.e., $\mathrm{Var}_t(u_{t,c}) \approx 0$), then the temporal variance satisfies $\mathrm{Var}_t(X_{a,c,1:T}^C) \approx \sigma_{a,c}^2$, providing a reliability proxy derived from $S(\cdot)$ when $S(\cdot) = \mathrm{Var}$. More generally, such an approximation requires preprocessing that suppresses the temporal dynamics of $u_{t,c}$ (e.g., detrending or differencing),

rather than only removing a constant mean. If $\varepsilon_{a,c,t}$ are correlated across antenna pairs, the optimal weights take the generalized least squares form $\alpha^\star \propto \Sigma^{-1}\mathbf{1}$, where Σ is the antenna-pair noise covariance.

Proof. Under the unbiasedness constraint, the fusion error equals $\sum_a \alpha_{c,a}\varepsilon_{a,c,t}$ and its variance is $\sum_a \alpha_{c,a}^2\sigma_{a,c}^2$. Minimizing this quadratic form subject to $\sum_a \alpha_{c,a} = 1$ via Lagrange multipliers yields Eq. (12). □

Discussion. Eq. (12) assigns larger weights to more reliable antenna pairs (smaller $\sigma_{a,c}^2$). In ST-CA2, the weights are produced from temporal statistics $p_{c,a}^A = S(X_{a,c,1:T}^C \mid M)$ (Eq. (5)) and normalized (Eq. (7)). Under the additional condition above, using variance as $S(\cdot)$ yields a proxy for $\sigma_{a,c}^2$, and the learned mapping can approximate inverse-variance weighting. When $u_{t,c}$ varies strongly within the window, the statistic mixes signal dynamics and noise; this motivates the empirical comparison of GTA/variance/RMS in Section IV-D.

F. Complexity and input-size analysis

1) Input size: Full-input usage has input size $\Theta(ATC)$ per sample. Mean fusion, random selection, and ST-CA2 all produce a compact representation of size $\Theta(TC)$ when the antenna-pair dimension is reduced to 1. Therefore, the antenna-pair dimension reduction yields an input compression ratio of A from full-input to the compact $C\times T$ representation.

2) Computation of ST-CA2: Let D denote the hidden dimension of the two-layer MLPs. Computing temporal priors p^C and p^A from X and X^C requires $O(ATC)$ operations for common statistics (GTA/Var/RMS). The CA MLP maps $\mathbb{R}^C \to \mathbb{R}^D \to \mathbb{R}^C$ for each antenna pair, contributing $O(ACD)$. The AA MLP maps $\mathbb{R}^A \to \mathbb{R}^D \to \mathbb{R}^A$ for each subcarrier, contributing $O(CAD)$. Element-wise reweighting and weighted fusion are both $O(ATC)$. Overall, the computational complexity of ST-CA2 is

$$O\big(ATC + ACD + CAD\big),$$

where the ATC term typically dominates when T is large (e.g., hundreds of time steps).

IV. Experimental Results and Discussion

We report four experiments: (i) input usage strategy comparison (Table I), (ii) temporal prior ablation for $S(\cdot)$ (Table II), (iii) CA weight evolution across epochs (Fig. 2), and (iv) signal-level fusion comparison against arithmetic mean (Fig. 3).

A. Data collection and protocol

We evaluate on a CSI-based person identification dataset collected during a door-opening motion (subjects rotate the handle and enter a room). Data were recorded in an indoor corridor environment with existing Wi-Fi.

Two computers were used: one as a Wi-Fi transmitter and the other as a receiver. Both were equipped with an Intel 5300 wireless network interface card supporting 3×3 MIMO. The Linux 802.11n CSI Tool [9] was employed to capture CSI from all three receive antennas simultaneously. The system operated in the 2.4 GHz band, with CSI measurements recorded at a sampling rate of 100 Hz.

A total of nineteen participants were recruited and each performed the door-opening task for 50 trials. Each CSI sample contains $N_a = 9$ antenna pairs, $N_c = 30$ OFDM subcarriers, and $T = 500$ time steps. Raw CSI amplitude data were denoised via a 3-level DWT with the db4 wavelet basis. A 5-fold cross-validation strategy was adopted; results are averaged across folds. A total of nineteen participants were recruited and each performed the door-opening task for 50 trials, yielding 950 samples in total. Each CSI sample contains $N_a = 9$ antenna pairs, $N_c = 30$ OFDM subcarriers, and $T = 500$ time steps. Raw CSI amplitude data were denoised via a 3-level DWT with the db4 wavelet basis. We adopted participant-level 5-fold cross-validation and used the same folds for all compared methods. In each fold, the training and test participant sets were disjoint, and all 50 trials from one participant were assigned to only one side of the split, preventing identity leakage across training and testing. Results are averaged across the five folds.

We report OA, macro Precision/Recall/F1, peak GPU memory, parameters, and GFLOPs.

B. Implementation details (brief)

Each CSI sample is stored as a CSV sequence and standardized by $x \leftarrow (x - 0.0025)/0.0119$ before being reshaped into $(A{\cdot}C)\times T = (270)\times 500$ for full-input usage. For compressed inputs, the 270 channels are split into $A\times C = 9\times 30$ segments and then reduced to a $C \times T = 30 \times 500$ representation by *Average* (mean over antenna pairs), *Random* (uniform selection of one antenna-pair segment per sample), or *ST-CA2* (CA+AA fusion).

For the downstream recognizer, we use a fixed backend that converts each $C \times T$ sequence into a Gramian Angular Summation Field (GASF) image of size $C\times T\times T$ and applies a lightweight CNN classifier. We train with Adam (learning rate 10^{-3}), batch size 16, and cross-entropy loss under 5-fold cross-validation. Unless otherwise specified, the reported peak GPU memory and GFLOPs are measured within this common learning setup after the CSI input has been prepared under each strategy, and are used here as relative comparisons within the same protocol rather than end-to-end edge-device measurements.

C. Input usage strategies

We compare four CSI input usage strategies under the same downstream recognizer and training protocol:

1) *Full*: the full CSI tensor $X \in \mathbb{R}^{A\times C\times T}$ is used as input.
2) *Random*: one antenna-pair segment is uniformly selected for each sample to construct a compact input.
3) *Average*: arithmetic averaging is applied along the antenna-pair dimension to obtain $\bar{X}(c,t) = \frac{1}{A}\sum_{a=1}^{A} X_{a,c,t} \in \mathbb{R}^{C\times T}$.
4) *ST-CA2*: CA+AA fusion produces $X_{\text{fuse}} \in \mathbb{R}^{C\times T}$.

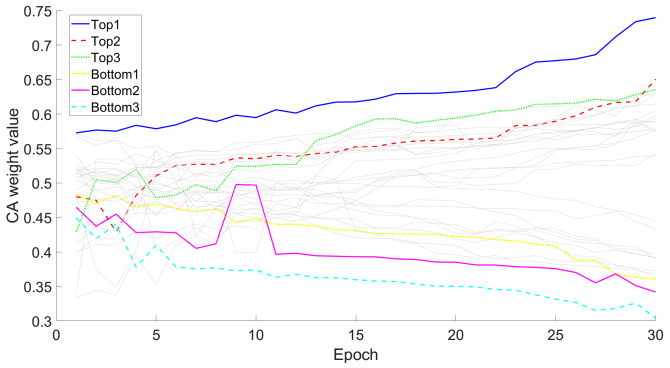

Fig. 2. CA weight evolution across epochs (Top-k and Bottom-k highlighted).

Table I reports recognition performance and efficiency. ST-CA2 achieves the best OA (99.20%) with 6.208 GFLOPs, improving over full-input (97.20%, 53.232 GFLOPs) while using a compact input. Compared with Average and Random, ST-CA2 improves accuracy with similar compute. This reduction in input-side compute and memory directly supports deploying efficient and tiny downstream models without relying on full-input processing.

D. Temporal prior $S(\cdot)$ ablation

The temporal statistical operator $S(\cdot)$ in ST-CA2 determines the priors p^C and p^A used for weight generation. We compare three typical scalar extraction methods: Global Temporal Average (GTA), Variance, and Root Mean Square (RMS).

For a subcarrier time series $X_{a,c,1:T}$, GTA is

$$p_{a,c}^{GTA} = \frac{1}{T}\sum_{t=1}^{T} X_{a,c,t},$$

Variance is

$$p_{a,c}^{Var} = \frac{1}{T-1}\sum_{t=1}^{T}\left(X_{a,c,t}-\mu_{a,c}\right)^2, \quad \mu_{a,c} = \frac{1}{T}\sum_{t=1}^{T} X_{a,c,t},$$

and RMS is

$$p_{a,c}^{RMS} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} X_{a,c,t}^2}.$$

Note that Var uses the sample factor $1/(T-1)$, while GTA/RMS use $1/T$; this constant scaling does not affect the relative ranking of antenna-pair "reliability" when variance is used as a proxy.

Table II reports the effect of GTA/variance/RMS as $S(\cdot)$. RMS achieves the best OA (99.20%), while variance and GTA lead to lower accuracy.

E. CA weight evolution

We record the CA weight vector $W^C \in (0,1)^C$ at the end of each training epoch (fixed antenna pair, fixed input batch). For visualization, Top-k and Bottom-k subcarriers are highlighted and the remaining curves are shown in gray (Fig. 2).

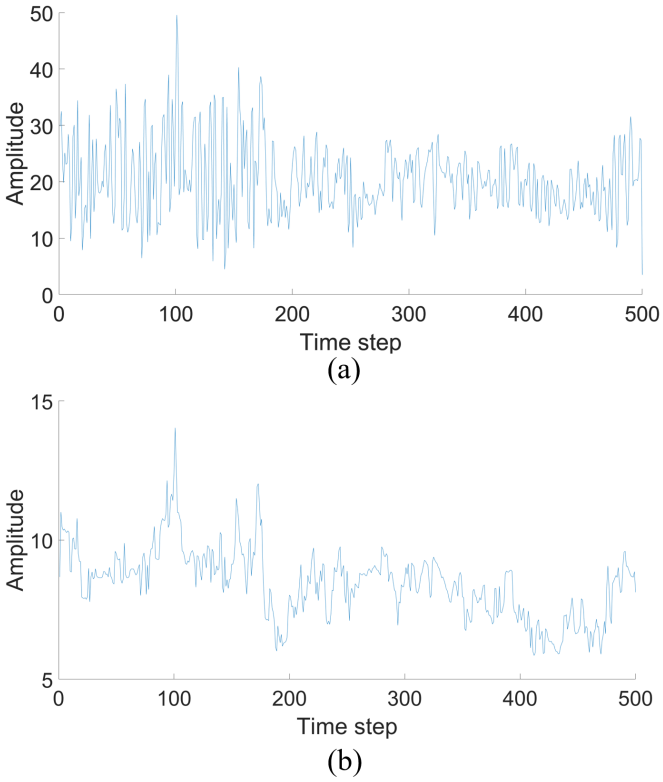

Fig. 3. Signal-level fusion comparison: (a) ST-CA2 vs. (b) arithmetic mean.

Fig. 2 shows non-uniform subcarrier weights across epochs, with stable separation between Top-k and Bottom-k trajectories.

F. Signal-level fusion comparison

To examine AA-based antenna-pair fusion, we compare the fused time series produced by ST-CA2 and arithmetic mean on a representative subcarrier (Fig. 3).

Fig. 3 compares fused time series. ST-CA2 preserves larger dynamic range and sharper motion-induced variations than arithmetic mean.

G. Discussion

ST-CA2 is an input-stage module and can be placed before different downstream recognizers. From an efficient-and-tiny perspective, ST-CA2 reduces the CSI input to a compact $C \times T$ form with small overhead, allowing the downstream model capacity to be allocated to recognition rather than handling redundant input. The evaluation here uses a fixed downstream recognizer and limited ablations. The current results are therefore best interpreted for the studied door-opening dataset and protocol. Broader comparisons with other learning-based lightweight modules and end-to-end measurements on CPU/MCU-class devices remain future work.

V. Conclusion

We presented ST-CA2, a two-stage module for CSI input refinement and fusion toward efficient and tiny Wi-Fi sensing

Method	OA (%)	F1 Score (%)	Precision (%)	Recall (%)	memory usage (GB)	parameters	GFLOPs
Full	97.20	97.38	97.72	97.32	8.601	1,450,451	53.232
Random	90.40	88.66	90.05	88.83	**1.446**	**697,811**	**6.192**
Average	93.20	90.91	93.27	91.46	1.447	697,811	6.192
ST-CA2	**99.20**	**99.25**	**99.36**	**99.19**	2.006	702,138	6.208

TABLE I
COMPARISON OF CSI INPUT USAGE STRATEGIES UNDER A FIXED DOWNSTREAM RECOGNIZER.

Method	OA (%)	F1 Score (%)	Precision (%)	Recall (%)
GTA	96.80	95.93	96.43	96.00
Variance	98.00	97.40	97.95	97.18
RMS	**99.20**	**98.92**	**98.97**	**99.00**

TABLE II
PERFORMANCE OF DIFFERENT SCALAR EXTRACTION STRATEGIES IN THE ST-CA2 MODULE

models. CA performs subcarrier reweighting and AA performs antenna-pair fusion to produce a compact $C \times T$ representation. Within the studied door-opening identification setting and under a fixed downstream recognizer, the analyses and experiments indicate an improved accuracy–efficiency trade-off compared with full-input and heuristic compression baselines.

ACKNOWLEDGEMENTS

We thank colleagues for helpful comments and suggestions that improved this manuscript. The research is partially supported by the grant from National Natural Science Foundation of China (62173157) and the Cross Research Project of Fundamental Research Funds for Central China Normal University.

REFERENCES

[1] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, jun 2019.

[2] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "Fila: Fine-grained indoor localization," in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2210–2218.

[3] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, dec 2013.

[4] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wichase: A wifi based human activity recognition system for sensorless environments," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2017, pp. 1–6.

[5] H. Lee, C. R. Ahn, and N. Choi, "Fine-grained occupant activity monitoring with wi-fi channel state information: Practical implementation of multiple receiver settings," *Advanced Engineering Informatics*, vol. 46, p. 101147, 2020.

[6] W. Zhang, J. Li, F. Ge, J. Hu, Z. Dai, X. Cao, Z. Yang, and X. Shuai, "Sewi: A framework enhancing csi-based human activity recognition," in *Advanced Intelligent Computing Technology and Applications*, D.-S. Huang, X. Zhang, and J. Guo, Eds. Singapore: Springer Nature Singapore, 2024, pp. 164–175.

[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018, pp. 7132–7141.

[8] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018, pp. 3–19.

[9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.