CFCST: A Cost-Efficient Spatio-Temporal Coupling Architecture for Multi-Task SHM Systems

Qingzheng Hu[1,2,†], Ningrong Lai[1,†], Yuansu Zou[2], Deshinta Arrova Dewi[1],
Siti Sarah Maidin[1], Luobing Pan[3], Chong Zhang[2,*]
[1]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia
[2]School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu, China
[3]College of Computer and Information Science, Southwest University, Chongqing, China

***Abstract*—As a cornerstone of smart city ecosystems, Structural Health Monitoring (SHM) leverages extensive Internet of Things (IoT) networks to enable real-time health assessment and prediction-based proactive maintenance, ensuring the operational integrity of infrastructure and preventing catastrophic failures. However, achieving reliable SHM within AI-enabled IoT (AIoT) frameworks faces a dual challenge: (1) the prohibitive cost of dense physical sensor deployment required for comprehensive coverage, and (2) the insufficient accuracy of existing time-series forecasting algorithms under complex spatiotemporal dynamics, which leads to unreliable proactive maintenance decisions.**

To address these bottlenecks, we propose CFCST, a unified multi-task deep learning model with a dual-objective architecture that facilitates high-fidelity virtual sensing alongside precise time-series prediction. At its core, the innovatively designed Fourier Channel Topology (FCT) block extracts coupled spatiotemporal features, while the Spatiotemporal Trend-Enhanced Attention (STEA) mechanism reallocates feature weights across tasks and enhances robustness against environmental noise. Evaluated on real-world infrastructure datasets, this single-model approach resolves identified constraints, offering a scalable, resource-aware AIoT solution that delivers both cost-efficient sensing and high-accuracy forecasting for modern SHM.

***Index Terms*—Smart City, Structural Health Monitoring, Internet of Things, Virtual Sensing, Time-series Prediction, Spatiotemporal Features**

I. Introduction

Structural Health Monitoring (SHM) ensures the operational integrity of smart city infrastructure. Driven by AI-enabled Internet of Things (AIoT) frameworks, these systems enable real-time health assessment and predictive maintenance to prevent catastrophic failures [1], [2]. However, large-scale SHM applications face two major bottlenecks.

First, data analytics challenges—including data complexity, distortion, and the inability of existing models to capture long-term dependencies and integrate multiple data types—reduce time-series prediction accuracy, ultimately leading to unreliable maintenance decisions [3], [4].

Second, deploying dense physical sensor networks is financially infeasible, severely restricting early-stage big data collection. Industrial-grade hardware and installation costs are exorbitant; for instance, monitoring bridges like the Bill Emerson and Tsing Ma required tens of thousands of dollars per sensor [5], [6]. Consequently, comprehensive monitoring is rare, leaving vast infrastructures—like most of California's 22,000 bridges—unmonitored due to prohibitive expenses [7].

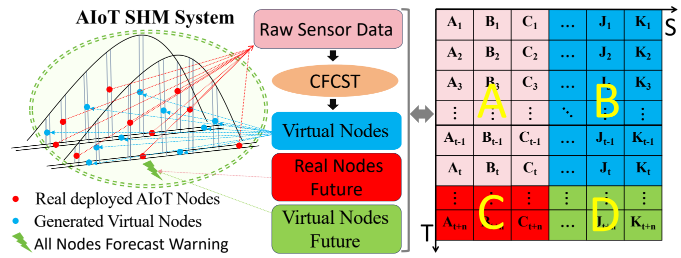

Fig. 1. Core idea of our CFCST design

To address these gaps, we propose the Convolutional Fourier Channel Spatiotemporal Trend-Enhanced (CFCST) model, a novel multi-task deep learning architecture (Fig. 1). CFCST simultaneously enhances time-series forecasting accuracy and reduces deployment costs via reliable virtual sensing. Unlike isolated virtual sensing approaches [8]–[13], CFCST efficiently integrates temporal prediction and spatial generation into a single architecture, enabling future-state prediction for virtual locations. The main contributions are:

1) Unified Multi-Task Architecture: A shared-parameter model integrating spatial, temporal, and spatiotemporal prediction. Incorporating an Uncertainty Weighting Loss (UWL) [14], it effectively couples spatiotemporal features, mitigates task interference, and reduces computational overhead.

2) Fourier Channel Topology (FCT) Block: We propose the FCT block to transform 1D time-series into 2D representations, utilizing convolutions to simultaneously capture intricate periodic temporal patterns and rich spatial features, providing a deeply coupled foundation for high-fidelity multi-task outputs.

3) Feature Weight Re-allocation and Noise Robustness: Since diverse tasks require distinct feature representations, we propose the Spatiotemporal Trend-Enhanced Attention (STEA) mechanism. STEA reallocates spatiotemporal feature weights to meet specific task demands, ensuring optimal feature utilization and reliable virtual sensor generation. Additionally, STEA incorporates a Temporal Trend Enhancement component that utilizes Causal Conv1D to match data points based on underlying temporal trends rather than raw values, thereby significantly improving the model's robustness against environmental noise.

†Qingzheng Hu (i24028296@student.newinti.edu.my) and Ningrong Lai (ronglnr@163.com) contributed equally to this work.
*Corresponding Author: Chong Zhang (zhangchong92@swpu.edu.cn).

II. Related Work

Recent studies aiming to enhance SHM prediction accuracy and reduce costs primarily follow two technological directions.

A. Time-Series Forecasting in SHM

Data-driven models forecast future structural states to enable early warnings. Existing methods employ techniques like variational mode decomposition [15], data mining [3], and big data integration [16] to enhance monitoring efficiency. Additionally, various computational models—including ARIMA [17], seasonal ARIMA [18], convolutional neural networks (CNNs) [19], and CNN-LSTM hybrids [4]—have been widely adopted to forecast future structural data trends.

B. Virtual Sensor Data Generation

Virtual sensing substitutes costly physical hardware. Early studies focused on specific regional conversions, such as deriving displacement from acceleration [8], mapping environmental correlations [9], or using physics-informed neural networks [10]. Although models like FAST [11] enable high-volume sensing, they lack multi-sensor versatility. Later approaches, such as CBLA [12] and CBLA+ [13], have introduced multi-category generation but still yield suboptimal data quality. Crucially, these spatial models lack the time-series forecasting capabilities essential for proactive maintenance.

C. Research Gap and Motivation

Current research remains bifurcated: temporal models neglect spatial correlations, limiting accuracy, while spatial models lack temporal foresight and noise robustness, producing unreliable virtual data. To bridge this gap, the CFCST model introduces a unified multi-task architecture that deeply couples spatiotemporal features and enhances noise resistance, resolving the conflict between high-precision forecasting and cost-effective virtual node generation.

III. Methodology

A. Problem Definition and Overview

The proposed CFCST architecture simultaneously optimizes forecasting precision and cost-effectiveness. As illustrated in Fig. 1, axes S and T represent the spatial and temporal dimensions, respectively. Let $\{A, B, \ldots, L\}$ denote the sensor nodes, where $\mathcal{X}_B = \{B_1, B_2, \ldots, B_t\}$ represents historical data collected by node B up to time t. Dataset A aggregates all historical data from physically deployed sensors. CFCST processes Dataset A to produce three functional outputs:

1) Spatial Virtual Sensing (Task(h)): Generates concurrent data for non-instrumented positions (Dataset B), substituting hardware to reduce costs.

2) Temporal Forecasting (Task(t)): Predicts future states of real sensors (Dataset C) for predictive maintenance.

3) Spatiotemporal Prediction (Task(t+h)): Integrates both to forecast future states of virtual sensors (Dataset D), enabling early warnings in non-instrumented regions.

This multi-task architecture leverages FCT and STEA modules (Fig. 2) to achieve a synergistic breakthrough in monitoring coverage and reliability.

B. Spatiotemporal Feature Extraction

We adopt the core module of TimesNet (i.e., TimesBlock) [20] as the backbone for our FCT block. Leveraging its 2D mapping strategy, FCT uses Fast Fourier Transform (FFT) to reshape 1D sequences into 2D space by dominant periodicities:

$$X[k] = \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{2\pi}{N}kn}, \quad k = 0, \ldots, N-1. \tag{1}$$

Since standard Inception blocks miss fine-grained cross-channel correlations, FCT replaces them with CTRGC blocks [21]. Without explicit graph construction, this substitution efficiently captures periodic and structural patterns, yielding robust spatiotemporal features.

C. Spatiotemporal Trend-Enhanced Attention

Coupled spatiotemporal features from FCT are routed to Task(t) and Task(h) (pink $\mathbf{t}$ and blue $\mathbf{h}$ in Fig. 2). However, the spatial Task(h) is limited by FCT's temporal bias and vulnerable to SHM noise. To address these limitations, we propose the CBAM-inspired [22] STEA module. Specifically, an Unsqueeze block first adds a channel dimension to feature $\mathbf{h}$, enabling STEA to optimize Task(h) via two mechanisms:

1) Spatiotemporal Feature Weight Re-allocation: STEA redistributes weights across channel and spatial dimensions to align with Task(h)'s interpolation needs. Channel Attention (M_c) captures inter-channel dependencies via parallel pooling layers: $M_c = \mathrm{Sigmoid}\big(W_1 \cdot \mathrm{ReLU}(W_0 \cdot \mathrm{Maxpool}(X)) + W_1 \cdot \mathrm{ReLU}(W_0 \cdot \mathrm{Avgpool}(X))\big)$. Subsequently, Spatial Attention (M_s) identifies salient regions using a 7×7 convolution on channel-concatenated pooling outputs: $M_s = \mathrm{Sigmoid}(\mathrm{Conv2D}(\mathrm{Concat}(\mathrm{Maxpool}(X'), \mathrm{Avgpool}(X'))))$. This dual-stage refinement reorients features to be predominantly spatial, enhancing virtual node generation.

2) Temporal Trend Enhancement: Lacking historical context, Task(h) suffers from noise. Standard self-attention [23] exacerbates this via point-wise matching, which often confuses noise fluctuations with stable patterns. Inspired by trend-aware mechanisms [24], STEA generates Query and Key representations using multi-scale Causal Conv1D instead of point-wise projections. With embedded local temporal trends, data points perceive contextual environments, effectively filtering noise and ensuring robust accuracy in non-instrumented regions.

D. Multi-Task Learning

Unlike standard uniform models, CFCST accommodates divergent data dimensions for heterogeneous tasks. Task(h) processes current-step inputs $(N, 1, C)$ into outputs (S, C), whereas Task(t) maps historical contexts (N, T, C) to predictions (N, P, C) (N/S: real/virtual nodes; T/P: historical/prediction steps; C: channels).

Due to these structural differences, a Fixed Weight Loss (FWL) would cause the task with the larger loss magnitude to dominate training. We employ an Uncertainty Weighting Loss (UWL) approach to dynamically balance task contributions:

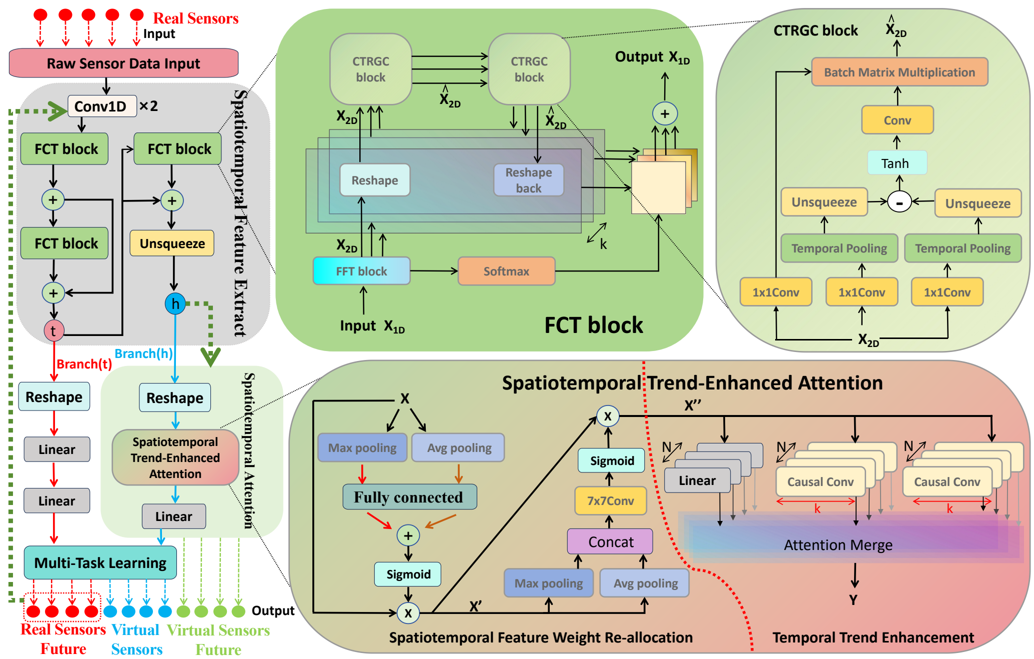

Fig. 2. Overview of the CFCST architecture (left) and detailed configurations of its key modules (right).

$$\mathcal{L} = \sum_{i \in \{h,t\}} \left(\frac{1}{2\sigma_i^2} \mathrm{MSE}(Y_{\text{pred}}^i, Y_{\text{true}}^i) + \frac{1}{2} \log \sigma_i^2 \right) \tag{2}$$

where MSE is the mean squared error for Task(i), and $\log \sigma_i^2$ is an adaptive logarithmic variance parameter (initially 0). By adjusting σ_i^2 during optimization, UWL penalizes disproportionately large losses, preventing single-task dominance and ensuring effective fitting across objectives.

IV. Experimental Validation and Performance Evaluation

A. Experimental Setup

Dataset: We utilize two real-world SHM datasets—the Swedish Steel Bridge (SSB) [25] and the Z24 Bridge AVT (Z24) by KU Leuven [26]—as benchmarks to verify CFCST's performance. To specifically capture the short-term dynamic behaviors of the structures, our experiments incorporated 40,000 continuous data points per sensor from 30 sensors in Z24, and 50,000 continuous data points per sensor from 20 sensors in SSB, providing a comprehensive array of signals including vibration acceleration, strain, and inclination.

Implementation Details: To evaluate virtual sensor reliability, we tested real-to-virtual sensor ratios of 2:1, 3:1, 4:1, and 5:1, corresponding to hardware cost savings of 33.3%, 25%, 20%, and 16.7% against full physical deployment. The dataset was split into training, validation, and testing sets with a ratio of 64:16:20. We trained the model for 400 epochs using the Adam optimizer (learning rate: 0.0001). Historical and prediction time steps were fixed at 96 and 36, respectively.

Evaluation Criteria: We evaluate the model across three objectives: virtual node generation (Task(h)), future data forecasting for real sensors (Task(t)), and future data forecasting for virtual nodes (Task(t+h)). For Task(t), we employ Mean Absolute/Squared Error (MAE/MSE), which are normalized (NMAE/NMSE) for the SSB dataset due to varying physical scales. For Task(h) and Task(t+h), we additionally report Curve Similarity (Similarity) [12], [13], a metric quantifying temporal trend alignment between predicted and actual signals:

$$\begin{aligned} S_{\text{similarity}} &= \left(1 - \frac{|S_1 - S_2|}{\max(S_1, S_2)}\right) \frac{t_1}{t} \\ &+ \left(1 - \frac{|S_3 - S_4|}{\max(S_3, S_4)}\right) \left(1 - \frac{t_1}{t}\right) \end{aligned} \tag{3}$$

where S_1 and S_2 denote the positive region areas of the predicted and true curves, respectively, while S_3 and S_4 represent their negative counterparts. To emphasize positive intervals, the similarity is weighted by t_1/t, where t_1 is the number of positive points in the true sequence of length t.

Baselines: We evaluate CFCST against landmark models: the classic Transformer Encoder (Transformer) [23], TimesNet [20], ModernTCN [27], and CBLA+ [13]. Adapted from official codebases, all baselines are tailored for our dual-task framework—spatial Task(h) and temporal Task(t). For single-step spatial inputs, TimesNet omits RevIN, and ModernTCN's large kernels degenerate into point-wise convolutions.

Hardware: The model was evaluated on an edge server equivalent (Intel i7-10870H CPU, NVIDIA RTX 3060 Laptop 6 GB GPU). The multi-task model CFCST employs a "compute-on-demand" strategy: Task(h) (real-time virtual sensing) requires only 9.81 MFLOPs with 0.33 ms latency, while Task(t) (future-state warning) consumes 601.02 MFLOPs with 0.80 ms latency. With a compact footprint of 7.08 M parameters and 58.89 MB of peak memory usage, the architecture effectively balances predictive accuracy and hardware efficiency, meeting industrial real-time requirements for AIoT edge gateways.

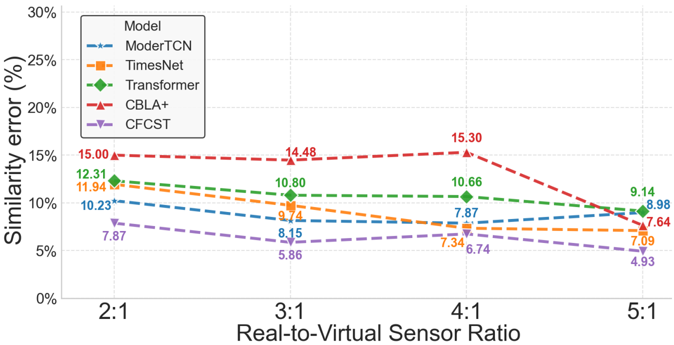

Fig. 3. Similarity error test results under different ratios

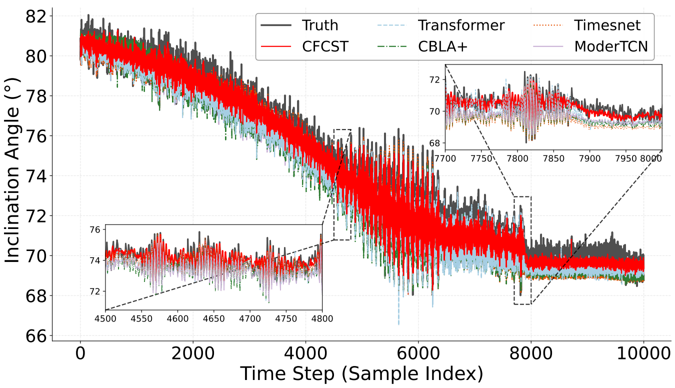

Fig. 4. Test results of different models

B. *Performance Evaluation of Virtual Node Generation*

Fig. 3 compares the Similarity error (the complement of Similarity, reflecting prediction-to-ground-truth discrepancy) of virtual nodes generated by five models at various ratios on Z24. Notably, CFCST consistently achieves the lowest error, reaching a minimum of 4.93% at the 5:1 ratio. At this ratio, 3 virtual sensors were generated from 15 real sensors for SSB, and 5 from 25 for Z24. The corresponding performance results are summarized in Table I, where the values represent the average metrics obtained from multiple sensor tests.

Table I reveals that ModernTCN achieves high Similarity on Z24 but generalizes poorly to SSB (evidenced by a sharp Similarity drop). Conversely, the Transformer's strong Similarity on SSB drops sharply on Z24. Although CBLA+ and TimesNet yield respectable Similarity scores, their MAE and MSE remain suboptimal. In contrast, CFCST consistently excels across both datasets, validating its robust generalization.

For visual clarity, Fig. 4 plots the predictions against the ground truth for one of the three virtual sensors generated on SSB at this ratio, directly aligning with the quantitative

TABLE I: EXPERIMENTAL RESULTS FOR TASK(H) AT 5:1 RATIO

Dataset	Metric	Transformer	TimesNet	ModernTCN	CBLA+	CFCST
Z24	Similarity(%)	90.86	92.91	91.02	92.36	**95.07**
	MAE/10^{-4}	0.181	0.167	0.162	0.151	**0.126**
	MSE/10^{-8}	0.071	0.067	0.065	0.070	**0.055**
SSB	Similarity(%)	94.43	95.26	81.29	93.26	**97.45**
	MAE/10^{-2}	2.160	2.467	2.432	2.575	**2.038**
	MSE/10^{-3}	1.168	1.431	1.324	1.425	**1.128**

TABLE II: EXPERIMENTAL RESULTS FOR TASK(H) AT 4:1 RATIO

Dataset	Metric	Transformer	TimesNet	ModernTCN	CBLA+	CFCST
Z24	Similarity(%)	89.34	92.66	92.13	84.70	**93.26**
	MAE/10^{-4}	0.241	0.246	0.262	0.227	**0.158**
	MSE/10^{-8}	0.122	0.134	0.154	0.118	**0.072**
SSB	Similarity(%)	91.75	90.35	77.41	89.17	**92.02**
	MAE/10^{-2}	2.481	2.501	2.586	2.737	**2.270**
	MSE/10^{-3}	1.549	1.562	1.608	1.747	**1.412**

TABLE III: ABLATION STUDY FOR TASK(H) AT 5:1 RATIO

Model	Z24 Dataset			SSB Dataset		
	Similarity (%)	MAE / 10^{-4}	MSE / 10^{-8}	Similarity (%)	MAE / 10^{-2}	MSE / 10^{-3}
TimesNet	92.91	0.167	0.067	95.26	2.467	1.431
CTI	93.06	0.145	0.060	91.16	2.390	1.162
CFC	88.64	0.148	0.062	90.71	2.406	1.278
CFCS	92.93	0.139	0.058	91.75	2.163	1.206
CFCST(h)	**97.97**	**0.124**	**0.054**	95.36	**1.902**	**1.063**
CTIST(h)	93.19	0.156	0.064	94.14	2.132	1.149
CFCST(1:1)	92.88	0.175	0.071	88.60	2.924	1.564
CFCST	95.07	0.126	0.055	**97.45**	2.038	1.128

results in Table I. While ModernTCN maintains a closer spatial distance to the ground truth than CBLA+, its temporal trajectory is poorly aligned (lower MAE and MSE but inferior Similarity). Conversely, TimesNet tracks temporal fluctuations well (high Similarity) but exhibits a significantly larger spatial gap than CFCST, reflected in its higher MAE and MSE. These results highlight the necessity of using a multi-metric evaluation for spatial interpolation tasks: Similarity assesses temporal feature extraction, while MAE and MSE evaluate spatial feature extraction. Relying solely on traditional time-series metrics like MAE and MSE is insufficient for a comprehensive performance assessment.

Fig. 3 shows that prediction errors increase at lower sensor ratios. Table II quantifies this at a 4:1 ratio, where 4 virtual nodes are generated from 16 real sensors for SSB, and 6 from 24 for Z24. While other models degrade significantly as the real-to-virtual ratio decreases, CFCST maintains a 92.64% average Similarity score, confirming its robust feature utilization for cost-effective SHM.

To evaluate key components, we conducted an ablation study at a 5:1 ratio (Table III). To validate our proposed STEA module, we decomposed it into two components for the ablation study: Spatiotemporal Feature Weight Re-allocation (SFWR) and Temporal Trend Enhancement. We defined several variants: CTI (TimesBlock+1D-CNN), CFC (FCT+1D-CNN), CFCS (CFC+SFWR), CTIST(h) (CTI+STEA), and CFCST(h) (CFC+STEA). Specifically, CTIST(h) and CFCST(h) are single-task models corresponding solely to Branch(h) in Fig. 2, excluding Task(t) functionality. For multi-task variants, CFCST(1:1) utilizes the FWL to balance Task(t) and Task(h) with a 1:1 loss ratio, while the final CFCST employs the UWL-based architecture.

The ablation results reveal the following.

1) TimesNet excels at temporal extraction but lacks spatial precision. Even after the integration of a 1D-CNN (CTI), its MAE and MSE remain suboptimal, suggesting the TimesBlock

TABLE IV: EXPERIMENTAL RESULTS FOR TASK(T)

Dataset	Metric	Ratio	Transformer	TimesNet	ModernTCN	CBLA+	CFCST
Z24 Dataset	MAE/10^{-5}	2:1	1.339	1.378	1.446	2.194	**1.279**
		3:1	1.392	1.420	1.589	2.256	**1.331**
		4:1	1.428	1.486	1.653	2.361	**1.405**
		5:1	1.546	1.617	1.666	2.420	**1.459**
	MSE/10^{-9}	2:1	0.423	0.439	0.468	0.896	**0.397**
		3:1	0.449	0.456	0.530	0.932	**0.431**
		4:1	0.488	0.503	0.582	1.001	**0.479**
		5:1	0.578	0.575	0.610	1.081	**0.515**
SSB Dataset	MAE/10^{-2}	2:1	1.409	1.537	1.316	1.529	**1.232**
		3:1	1.596	1.579	1.531	1.667	**1.391**
		4:1	1.616	1.662	1.579	1.858	**1.325**
		5:1	1.596	1.579	1.531	1.667	**1.343**
	MSE/10^{-4}	2:1	3.849	4.207	2.939	4.588	**2.802**
		3:1	6.001	6.034	4.970	6.551	**5.050**
		4:1	6.037	6.917	5.573	7.429	**5.160**
		5:1	6.001	6.034	4.970	6.551	**4.979**

TABLE V: ABLATION STUDY FOR TASK(T) AT 5:1 RATIO

Dataset	Metric	Ratio	TimesNet	CTI	CFCST(t)	CFCST(1:1)	CFCST
Z24 Dataset	MAE/10^{-5}	2:1	1.378	1.367	**1.262**	1.420	1.279
		3:1	1.420	1.394	**1.319**	1.457	1.331
		4:1	1.486	1.465	**1.376**	1.609	1.405
		5:1	1.617	1.532	**1.420**	1.576	1.459
	MSE/10^{-9}	2:1	0.439	0.439	**0.391**	0.448	0.397
		3:1	0.456	0.458	**0.424**	0.467	0.431
		4:1	0.503	0.492	**0.464**	0.580	0.479
		5:1	0.575	0.538	**0.497**	0.563	0.515
SSB Dataset	MAE/10^{-2}	2:1	1.537	1.373	**1.184**	1.462	1.232
		3:1	1.579	1.495	**1.319**	1.899	1.391
		4:1	1.662	1.428	**1.299**	1.553	1.325
		5:1	1.579	1.495	**1.319**	1.523	1.343
	MSE/10^{-4}	2:1	4.207	3.076	**2.771**	4.134	2.802
		3:1	6.034	5.180	**4.771**	7.534	5.050
		4:1	6.917	5.123	**4.922**	6.034	5.160
		5:1	6.034	5.180	**4.771**	5.422	4.979

architecture is ill-suited for spatial interpolation.

2) Replacing TimesBlock with our FCT block (CFC) initially couples richer spatiotemporal features, but the resulting mixture is temporally dominant. Consequently, CFC's metrics temporarily decline compared to CTI. Introducing SFWR for weight re-allocation significantly improves all three metrics (CFCS vs. CFC), validating the STEA module's effectiveness in reconstructing spatiotemporal features.

3) The full single-task model, CFCST(h), adds the Temporal Trend Enhancement component. This improves robustness by mitigating noise, confirming STEA's role in enhancing both stability and performance.

4) The superior performance of CTIST(h) over CTI further confirms the effectiveness of our STEA module and demonstrates its robust generalization capability.

5) The improvement from CFC to CFCST(h) far exceeds that of CTI to CTIST(h). As STEA is a feature representation module, its efficacy depends on the richness of extracted features. This confirms that our proposed FCT block possesses superior spatial feature extraction capabilities.

6) CFCST(1:1) performs poorly due to loss imbalances between Task(h) and Task(t). By employing the UWL algorithm, CFCST effectively couples spatiotemporal features for high performance. Although its spatial accuracy is slightly lower than the single-task CFCST(h) due to multi-task parameter sharing, CFCST(h) lacks the predictive functionality essential for SHM. Thus, CFCST is superior for comprehensive monitoring, whereas CFCST(h) is suitable only when future prediction is not required.

C. Performance Evaluation of Real Node Future Prediction

To evaluate Task(t), we forecasted future data for real sensors across various real-to-virtual ratios. For Z24, ratios of 5:1, 4:1, 3:1, and 2:1 correspond to 25, 24, 21, and 20 real sensors, respectively; for SSB, these correspond to 15, 16, 15, and 12. Results in Table IV represent the average MAE and MSE across all real sensors.

Table IV shows that CFCST consistently yields lower MAE and MSE than other models across all ratios. This demonstrates that CFCST not only excels in virtual node generation but also rivals or surpasses classic time-series models, making it suitable for proactive early warning in environmental monitoring. Conversely, while CBLA+ performs adequately in virtual data generation, its time-series prediction capability is significantly weaker.

To evaluate module contributions to Task(t), an ablation study was conducted (Table V). Specifically, CFCST(t) (FCT+1D-CNN) is a single-task model corresponding solely to Branch(t) in Fig. 2. The ablation results reveal the following.

1) CFCST(t) outperforms CTI in time-series prediction, demonstrating the superior temporal extraction of our FCT block. Combined with the Task(h) ablation results (Point 5), this further validates the FCT block's ability to capture rich spatiotemporal features simultaneously.

2) Consistent with Task(h) Point 6, CFCST(1:1) performs poorly due to loss imbalances, whereas CFCST overcomes this limitation to achieve predictive performance near that of CFCST(t). Notably, while CFCST(t) excels in accuracy, it lacks the virtual sensor generation capability necessary to reduce SHM hardware costs, making the multi-task CFCST a more comprehensive solution.

D. Performance Evaluation of Virtual Node Future Prediction

Having validated Task(h) and Task(t), we now evaluate their integration: predicting future data for virtual nodes (Task(t+h), i.e., A→D in Fig. 1), which is vital for long-term SHM stability. We compare two logic paths: 1) A→B→D: generating virtual history then predicting its future; 2) A→C→D: predicting real-node futures then generating virtual futures.

Method 1 fails because it generates virtual data point-wise using only instantaneous spatial features, ignoring the spatial feature evolution over time. This leads to a temporal distribution that deviates from the original series, preventing the model from learning its underlying dynamics. Consequently, errors accumulate over the thousands of steps required for long-term structural monitoring, resulting in numerical instability (tending toward infinity). Therefore, we adopt Method 2 for Task(t+h), as illustrated by the green data flow arrows in Fig. 2.

TABLE VI: ABLATION STUDY FOR TASK(T+H) AT 5:1 RATIO

Model	Z24 Dataset			SSB Dataset		
	Similarity (%)	MAE / 10^{-4}	MSE / 10^{-8}	Similarity (%)	MAE / 10^{-2}	MSE / 10^{-3}
CTIST	89.96	0.202	0.093	91.33	2.829	1.746
CFCST(1:1)	91.47	0.209	0.096	85.05	2.907	1.604
CFCST	**93.15**	**0.197**	**0.092**	**94.73**	**2.712**	**1.590**

This approach first predicts future real-node states by coupling current and historical data, establishing a foundation that preserves both temporal and spatial distributions. Generating virtual future states from this robust basis subsequently yields high-precision predictions.

Since no existing publicly available models support this function, this section only presents an ablation study comparing model performance. To ensure consistency with the Task(h) and Task(t) evaluations, the experiment was conducted at a 5:1 ratio (Table VI). Here, we defined CTIST as the multi-task baseline (TimesBlock+1D-CNN+STEA+UWL) to be compared against our FCT-based variants, CFCST(1:1) and CFCST. The ablation results reveal the following.

1) Consistent with the ablation results for Task(h) and Task(t), CFCST(1:1) again performs poorly due to the aforementioned loss imbalances.

2) The CFCST outperforms CTIST across all metrics, further validating that the proposed FCT block possesses superior spatiotemporal feature extraction capabilities compared to TimesBlock.

In summary, comprehensive evaluations on real-world datasets confirm that CFCST effectively addresses core SHM data challenges. The results demonstrate that the model achieves high-precision forecasting while producing high-fidelity virtual sensor data. By successfully executing Task(t+h), CFCST maintains long-term operational stability and expands sensing coverage, proving the synergistic efficiency of our spatiotemporal architecture and uncertainty-based balancing.

V. Conclusion

In this paper, we introduced CFCST, a novel spatiotemporal computing paradigm for cost-effective SHM within AIoT frameworks. Experiments demonstrate that CFCST resolves the conflict between monitoring precision and hardware costs via reliable virtual sensing. By reducing reliance on dense physical deployments, it lowers economic barriers to large-scale data acquisition, providing a critical technical foundation for resilient urban infrastructure and smart city evolution.

References

[1] E. Aktan, I. Bartoli, B. Glišić, and C. Rainieri, "Lessons from bridge structural health monitoring (shm) and their implications for the development of cyber-physical systems," *Infrastructures*, vol. 9, no. 2, 2024.

[2] N. Lai *et al.*, "A comprehensive review of lightweight deep learning models for edge computing with future directions," *Discover Computing*, vol. 29, no. 1, p. 110, 2026.

[3] Q. Jin, Z. Liu, J. Bin, and W. Ren, "Predictive analytics of in-service bridge structural performance from shm data mining perspective: A case study," *Shock and Vibration*, vol. 2019, pp. 1–11, 2019.

[4] X. Zhang *et al.*, "High-speed railway seismic response prediction using cnn-lstm hybrid neural network," *Journal of Civil Structural Health Monitoring*, vol. 14, no. 5, p. 15, 2024.

[5] S. Cho, B. F. Spencer, H. Jo, J. Li, and R. E. Kim, "Bridge monitoring using wireless smart sensors," *Spie Newsroom*, pp. 1–3, 2012.

[6] C. Lau, W. Mak, K. Wong, W. Chan, and K. Man, "Structural health monitoring of three cable-supported bridges in hong kong," *Structural health monitoring*, pp. 450–460, 2000.

[7] University of Georgia SensorWeb Research Laboratory, "A versatile WSN platform for structural health monitoring," *University of Georgia SensorWeb Research Laboratory*. [Online]. Available: https://sensorweb.engr.uga.edu/index.php/shm. [Accessed: Jun. 14, 2025].

[8] S. Cui, T. Hoang, K. Mechitov, Y. Fu, and B. F. Spencer, "Adaptive edge intelligence for rapid structural condition assessment using a wireless smart sensor network," *Engineering Structures*, vol. 326, no. 000, 2025.

[9] A. Kowli, V. Rani, and M. Sanap, "Data-driven virtual sensing for spatial distribution of temperature and humidity," *Journal of Building Engineering*, vol. 67, p. 105726, 2023.

[10] M.-S. Go, J. H. Lim, and S. Lee, "Physics-informed neural network-based surrogate model for a virtual thermal sensor with real-time simulation," *International Journal of Heat and Mass Transfer*, vol. 214, p. 124392, 2023.

[11] C. Zhang *et al.*, "Fast: A ubiquitous inference computation model for temperature and humidity sensing," *IEEE Sensors Journal*, vol. 25, no. 1, pp. 1452–1464, 2024.

[12] Y. Wang *et al.*, "Cbla: Empowering virtual sensor nodes with zero deployment costs for shm systems," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–10.

[13] C. Zhang *et al.*, "A reliable virtual sensing architecture with zero additional deployment costs for shm systems," *IEEE sensors journal*, no. 22, p. 24, 2024.

[14] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," *IEEE*, 2018.

[15] J. Xin *et al.*, "Bridge deformation prediction based on shm data using improved vmd and conditional kde," *Engineering structures*, no. Jun.15, p. 261, 2022.

[16] A. Entezami, H. Sarmadi, B. Behkamal, and S. Mariani, "Big data analytics and structural health monitoring: A statistical pattern recognition-based approach," *Sensors (Basel, Switzerland)*, vol. 20, no. 8, 2020.

[17] S. Bian, J. Zhuo, and L. Zhu, "Strain prediction of bridge shm based on ceemdan-arima model," *IOP Conference Series: Earth and Environmental Science*, vol. 558, no. 3, p. 032036, aug 2020.

[18] B. Qu, P. Liao, and Y. Huang, "Outlier detection and forecasting for bridge health monitoring based on time series intervention analysis," *Structural Durability Health Monitoring (SDHM)*, vol. 16, no. 4, 2022.

[19] B. K. Oh, B. Glisic, Y. Kim, and H. S. Park, "Convolutional neural network-based wind-induced response estimation model for tall buildings," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 10, pp. 843–58, Oct 2019.

[20] H. Wu *et al.*, "TimesNet: Temporal 2D-variation modeling for general time series analysis," in *International Conference on Learning Representations*, 2023.

[21] Y. Chen *et al.*, "Channel-wise topology refinement graph convolution for skeleton-based action recognition," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 13 359–13 368.

[22] 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)*, 2018, pp. 3–19.

[23] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[24] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong, "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5415–5428, 2021.

[25] J. Leander, J. Nyman, R. Karoumi, P. Rosengren, and G. Johansson, "Dataset for damage detection retrieved from a monitored bridge pre and post verified damage," *Data in Brief*, vol. 51, p. 109729, 2023.

[26] J. Maeck, B. Peeters, and G. De Roeck, "Damage identification on the z24 bridge using vibration monitoring," *Smart materials and structures*, vol. 10, no. 3, p. 512, 2001.

[27] D. Luo and X. Wang, "Moderntcn: A modern pure convolution structure for general time series analysis," in *The twelfth international conference on learning representations*, 2024, pp. 1–43.