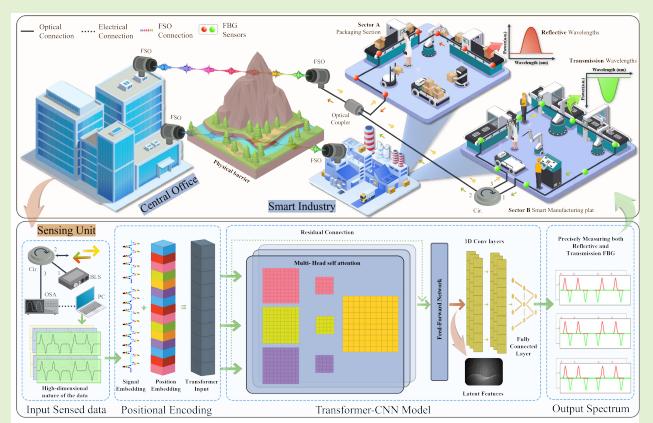


# Integration of FBG Reflection and Transmission Spectra for Sensing Capacity Enhancement With Transformer CNN Demodulation

Michael Augustine Arockiyadoss<sup>ID</sup>, Cheng-Kai Yao<sup>ID</sup>, Pradeep Kumar<sup>ID</sup>, Yao-Jen Chung, Amare Mulatie Dehnaw<sup>ID</sup>, and Peng-Chun Peng<sup>ID</sup>

**Abstract**—We present a novel fiber Bragg grating (FBG) sensor system that employs both reflective and transmission wavelengths to significantly boost capacity within a single network. By integrating free-space optics (FSOs) for remote communication, this architecture operates beyond physical constraints where laying fiber is not feasible, thereby enhancing deployment flexibility. To address the overlap of Bragg wavelengths commonly encountered in dense FBG networks, we utilize a Transformer convolutional neural network (CNN) that leverages positional encoding to more effectively learn data patterns. Preliminary results demonstrate a significant enhancement in sensor capacity and detection accuracy. Moreover, the proposed Transformer CNN reduces both training and inference times by approximately 2.24× and 1.43×, respectively, compared to traditional Transformer models. By leveraging both reflective and transmission FBG readouts within a fixed spectral window, the system supports nearly double the number of sensors, offering a scalable and efficient solution well-suited for high-density, real-time industrial sensing applications.



**Index Terms**—Convolutional neural network (CNN), fiber Bragg grating (FBG), free-space optics (FSOs), overlapping Bragg wavelengths.

## I. INTRODUCTION

FIBER Bragg grating (FBG) sensors [1], [2] are extensively utilized for monitoring physical parameters such as strain, temperature, and vibration [3], [4], [5], [6], [7], [8] by reflecting specific Bragg wavelengths. Traditionally, these sensors rely solely on the reflective Bragg wavelength to identify applied constraints. This approach, however, limits the number of sensors in a network due to potential wavelength overlap when multiple sensors are deployed, making it challenging to distinguish individual sensor signals accurately. In this study, we introduce a novel architecture that leverages both the reflective and transmission wavelengths

of FBG sensors to significantly increase the overall capacity within a single network. While most previous research has focused on reflective wavelengths alone, our approach harnesses the otherwise underutilized transmission spectrum, effectively doubling the number of sensing points without complex multiplexing techniques. This enhancement directly addresses the growing demand for dense sensor networks in applications such as structural health monitoring and industrial process control. To further expand the system's deployment flexibility, we integrate free-space optics (FSOs) into the architecture [9]. FSO enables wireless optical communication between remote sensing sites and a central monitoring station, especially in harsh or geographically challenging environments where laying fiber cables is impractical or cost-prohibitive. By overcoming these location-based constraints, FSO ensures robust data transfer and broadens the range of potential monitoring applications [10], [11], [12].

A significant challenge in densely concentrated FBG sensor networks is the overlapping of Bragg wavelengths when strain is applied, making it difficult to accurately detect and distinguish individual sensor signals [13], [14], [15]. To tackle this issue, we propose a Transformer convolutional

Received 12 April 2025; revised 9 May 2025; accepted 10 May 2025. Date of publication 23 May 2025; date of current version 2 July 2025. This work was supported by the National Science and Technology Council, Taiwan, under Grant NSTC 112-2221-E-027-076-MY2. The associate editor coordinating the review of this article and approving it for publication was Dr. Qiang Wu. (Corresponding author: Peng-Chun Peng.)

The authors are with the Department of Electro-Optical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan (e-mail: pcpeng@ntut.edu.tw).

Digital Object Identifier 10.1109/JSEN.2025.3571065

neural network (CNN). CNNs effectively capture local spectral patterns, while the Transformer's self-attention mechanisms [16], [17], [18] resolve long-range dependencies in overlapping wavelengths. Our hybrid design improves detection accuracy and reduces computational costs compared to traditional methods, ensuring efficient and precise FBG signal analysis. Previous methods, such as wavelength-division multiplexing (WDM) [19] and intensity WDM (IWDM) [20], [21], have attempted to improve sensor capacity but still face limitations regarding scalability and complexity. Our system stands out by simultaneously leveraging reflection and transmission FBG sensing. It also integrates FSO for remote communication and employs a Transformer CNN for rapid, accurate resolution of overlapping Bragg wavelengths. As illustrated in the schematic of the proposed reflective and transmission FBG sensor system with FSO transmission, a monitoring office equipped with a broadband light source (BLS) communicates via FSO links with smart industry, including a packaging sector using reflective-wavelength FBGs and smart manufacturing plants utilizing transmission-wavelength FBGs within the smart industry. The combined reflected and transmitted signals are sent back through the FSO link, where an optical spectrum analyzer (OSA) and computer system process the data. In summary, our primary objectives revolve around three key thrusts.

- 1) Maximize sensor capacity by exploiting both reflection and transmission spectra of FBGs, thereby allowing high-density sensor networks.
- 2) Extend deployment flexibility through the integration of FSO, which ensures reliable data transfer in difficult or remote locations where conventional fiber-based solutions are impractical.
- 3) Improve both speed and accuracy of wavelength detection using a Transformer CNN, specifically reducing training/testing time compared to the original Transformer architecture while enhancing overall detection reliability.

By fulfilling these objectives, the proposed system effectively addresses the challenges posed by overlapping Bragg wavelengths and remote monitoring requirements, paving the way for scalable and versatile sensor networks across diverse industrial and infrastructural applications. Following the style of our previous works, the remainder of this article is organized as follows. Section II presents the experimental setup of our FSO-based reflective and transmission FBG sensor system. Section III details our methodology, including data collection and an overview of the proposed Transformer CNN. Section IV discusses the results and provides analysis. Finally, Section V concludes this article and suggests future research directions.

## II. EXPERIMENTAL SETUP OF FSO-BASED REFLECTIVE AND TRANSMISSION FBG SENSOR SYSTEM

This section describes the overall configuration for utilizing both reflective and transmission FBGs in conjunction with FSO, focusing primarily on strain measurements conducted in a laboratory environment. The system is designed to investigate overlapping Bragg wavelengths in high-density

**TABLE I**  
EXPERIMENTAL SCENARIOS FOR TESTING MODEL EFFECTIVENESS  
AND EXPLORING TRANSMISSION WAVELENGTH SENSOR SYSTEM

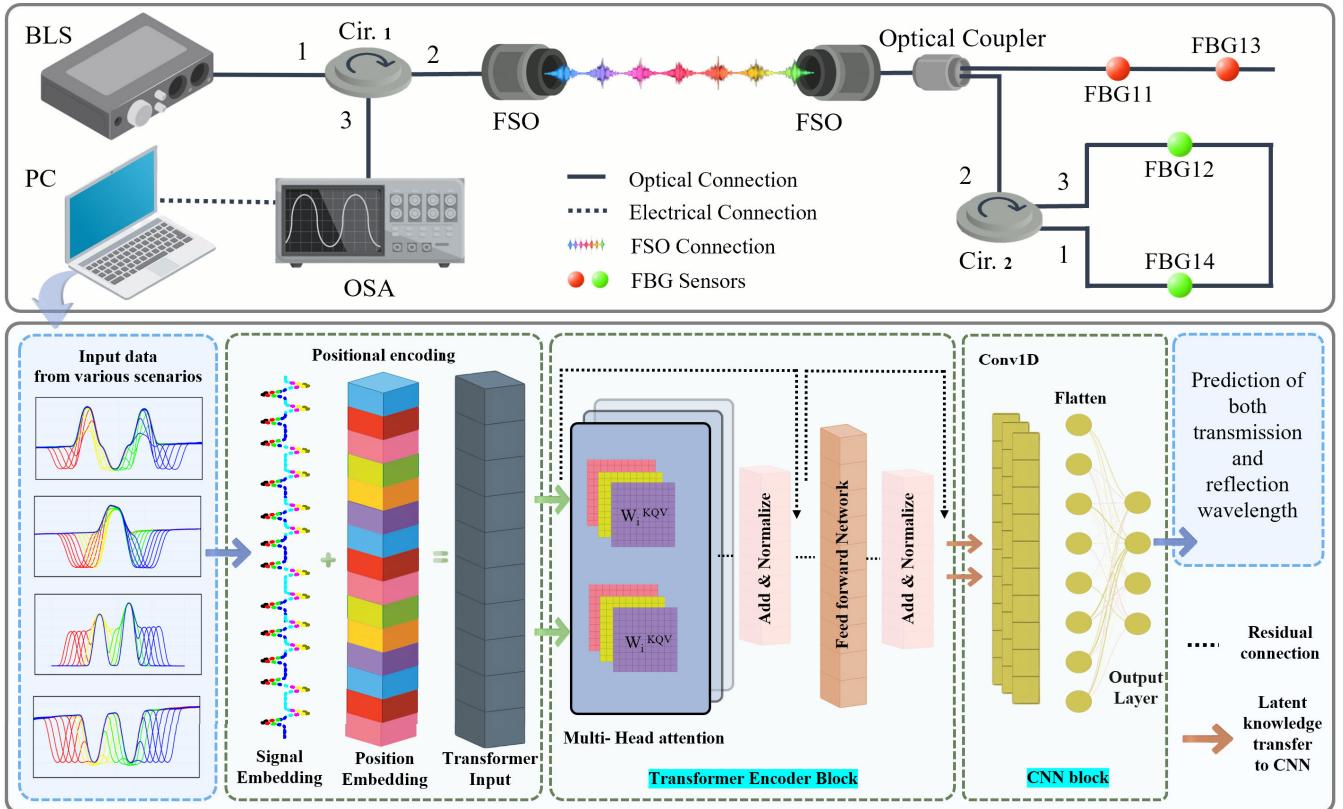
Parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Number of FBG Sensors	4	2	3	3
Reflectivity FBG	2	1	3	0
Transmission FBG	2	1	0	3
Number of Strain Steps	21	20	20	20
Shift per Step (nm)	0.06	0.03	0.06	0.06
Wavelength Shift Range (nm)	1544.7–1545.9	1544.8–1545.4	1544.7–1545.9	1544.7–1545.9

The table summarizes the experimental scenarios used to evaluate both the model's effectiveness and the sensor system's performance, varying by sensor count, reflectivity, transmission, and strain steps.

FBG networks by collecting data under multiple strain levels, resulting in shifts in Bragg wavelengths that can lead to spectral overlap. A schematic illustration of this setup is provided in Fig. 1, and the different experimental scenarios appear in Table I. In high-density environments, the use of both reflected and transmitted FBG wavelengths may lead to partial spectral overlap or noise, particularly under atmospheric turbulence or complex multiplexing conditions. In such cases, machine learning techniques have been shown to effectively mitigate interference and enhance the accuracy of wavelength demodulation [22], [23].

At the monitoring station, a BLS is connected to an optical circulator (Cir. 1), which directs the outgoing light to an FSO transmitter. The beam travels through a short free-space path to an FSO receiver on the sensor side, providing robust and flexible communication without the need for extended fiber runs. In this controlled laboratory setup, environmental factors (e.g., temperature and weather) were not a major concern though alignment of the FSO link was carefully maintained to minimize signal loss. While our work focused on machine learning for FBG wavelength demodulation, we note that recent studies have explored the use of machine learning to enhance FSO alignment and link stability under dynamic conditions [24].

Upon reaching the sensor side, the received light is directed to an optical coupler, which splits the beam into two primary paths. In the reflective wavelength path, FBG11 and FBG13 are attached at one coupler terminal, reflecting specific portions of the incident spectrum back through the FSO link. Meanwhile, in the transmission wavelength path, the other terminal of the optical coupler connects to an optical circulator (Cir. 2), leading to a ring-type arrangement with FBG12 and FBG14. These sensors allow modified spectra to pass through before returning to the monitoring station via the same FSO link. By monitoring both reflective and transmission paths, the system effectively doubles the potential sensing points compared to a purely reflective setup. After the signals from both paths recombine, they are routed back through the FSO receiver at the monitoring station and directed via the optical



**Fig. 1.** Experimental setup of proposed reflective and transmission FBG sensor system with FSO transmission and architecture of Transformer CNN (BLS: broadband light source, PC: personal computer, OSA: optical spectrum analyzer, Cir: circulator, and FSO: free-space optics).

circulator (Cir. 1) to an OSA. Its output is then collected by a computer for wavelength detection and strain analysis. Because all FBG sensors are mounted on a fixed structure, no additional calibration or baseline procedure was performed in this study.

This configuration serves as the foundation for evaluating how overlapping Bragg wavelengths, an inherent challenge in dense FBG networks, are handled under varied strain conditions. Although only a small free-space distance was used here, the integration of FSO demonstrates how this approach can extend beyond the laboratory and into remote or distributed sensing applications where installing long fiber runs might be impractical. As summarized in Table I, several scenarios were tested (reflective-only, transmission-only, or mixed arrays) to illustrate how the architecture adapts to different sensor configurations and strain conditions. These details set the stage for Section III, where data collection procedures, preprocessing steps, and the proposed Transformer CNN approach for resolving overlapping spectra will be discussed. Ultimately, these findings underline the system's adaptability and potential for large-scale, real-world monitoring applications.

### III. METHODOLOGY

#### A. Data Collection

Before detailing the experimental setup, we offer a concise overview of the data collection approach. In this procedure, strain is systematically applied to specific FBG sensors, while other sensors remain unstrained, producing a spectrum of

overlapping signals. These varied conditions facilitate capturing a comprehensive range of data for model training and validation. Fig. 1 illustrates the configuration used to gather spectral data from four FBG sensors, comprising two reflective sensors (FBG11 and FBG13) and two transmission sensors (FBG12 and FBG14). In Scenario 1, strain was selectively applied to FBG11 in 21 increments, shifting its wavelength from 1544.7 to 1545.9 nm at a step size of 0.06 nm, while FBG13, FBG12, and FBG14 remained at their respective wavelengths with minimal or no strain. As the strain on FBG11 increased, partial or complete overlaps were observed between its Bragg peak and those of the other three sensors, posing significant detection and demultiplexing challenges. Although the reflection spectrum of each FBG can be approximated by a Gaussian profile, we consider the intensity of an FBG given [9] by the following equation:

$$I(\lambda, \lambda_{B_i}) = I_p \exp \left[ -4 \ln(2) \left( \frac{\lambda - \lambda_{B_i}}{\Delta \lambda_{B_i}} \right)^2 \right] \quad (1)$$

where  $I_p$  denotes the peak power of the FBG,  $\lambda_{B_i}$  is its central (Bragg) wavelength,  $\lambda$  represents the wavelength range of interest, and  $\Delta \lambda_{B_i}$  corresponds to the full-width at half-maximum (FWHM). In this study, we did not focus on establishing a correlation between  $\Delta \lambda_{B_i}$  and induced strain increments. Data acquisition in the laboratory was carried out without specifying a fixed sampling frequency or measurement duration for each strain increment, and no filtering or

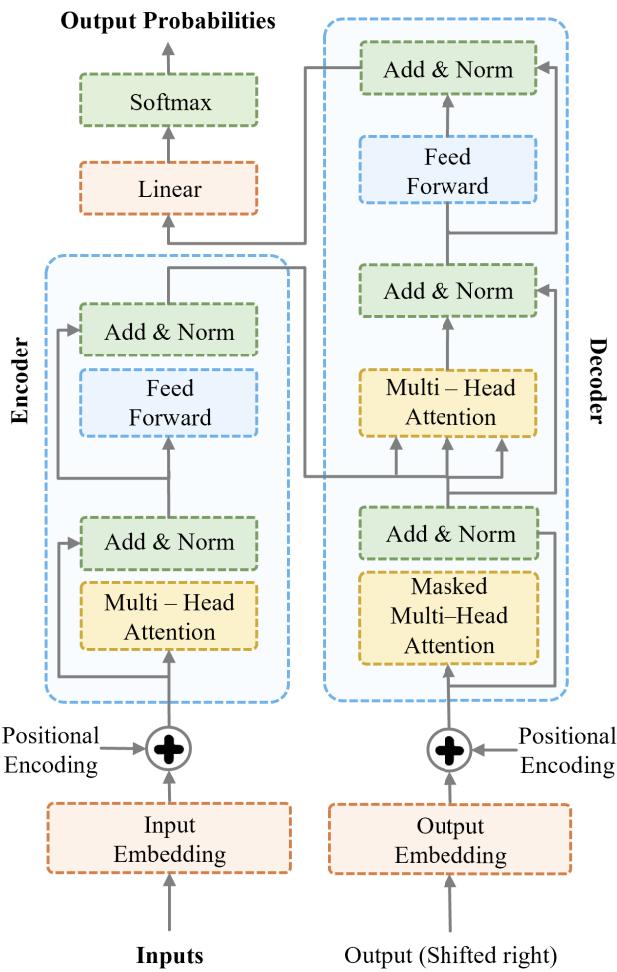


Fig. 2. Transformers architecture [16].

smoothing techniques were applied before the spectra were processed.

In practice, monitoring reflective (FBG11 and FBG13) versus transmission (FBG12 and FBG14) outputs can introduce different spectral behaviors, especially under overlapping conditions. Because both reflective and transmission sensors operate at slightly different wavelengths, the system must accommodate these variations, an aspect that motivated the development of our proposed method. Overlaps caused by strain are ultimately resolved using the Transformer CNN discussed in Section III. Moreover, the overlapped spectra of FBGs cause crosstalk between the FBG sensors, making it challenging to detect the central wavelength of FBGs. Therefore, developing a signal processing method to enhance the accuracy of peak wavelength detection is a crucial issue. Hence, we propose a Transformer CNN-based approach to accurately predict the central wavelength of FBGs. The details of this proposed method are presented in Section III.

### B. Proposed Transformer CNN

To address the challenge of overlapping FBG spectra, we adopt a Transformer-based model enhanced with CNN

layers. A Transformer is a type of neural network architecture that relies entirely on self-attention mechanisms to process sequences of data (Fig. 2) [16]. Unlike traditional sequence models such as RNNs and CNNs, Transformers do not require sequential data to be processed in order. Instead, they process all spectral positions (referred to as “tokens” in the original Transformer literature) in the input sequence simultaneously, allowing for greater parallelization and efficiency during training [16], [25]. Here, each “token” corresponds to a spectral point representing a projected feature of the input spectrum. The Transformer architecture is composed of an encoder and a decoder, both of which are stacks of identical layers. The encoder processes the input sequence and generates a rich representation, while the decoder uses this representation to generate the output sequence, one step at a time [16], [26]. Traditional models such as RNNs suffer from limitations in capturing long-range dependencies due to vanishing gradients and the inherent sequential nature of their computations, which makes parallelization difficult. Transformers address these issues.

- 1) *Utilizing Self-Attention:* Self-attention mechanisms allow the model to weigh the importance of different words in a sequence relative to each other, regardless of their position. This enables the model to capture long-range dependencies more effectively [27].
- 2) *Parallel Processing:* By processing all spectral points simultaneously, Transformers enable faster training times and more efficient use of computational resources [27].
- 3) *Capturing Global Dependencies:* The self-attention mechanism provides a way for the model to consider the entire sequence context for each spectral point, improving the model’s understanding of the language structure [27].

These advantages have established Transformers as the basis for numerous advanced models in natural language processing (NLP), including BERT and the GPT series. Their ability to efficiently handle variable-length sequences without relying on explicit recurrence or fixed-size convolutions has also enabled their application in tasks beyond NLP, including time-series forecasting and analyzing overlapping spectra, where both local peaks and global spectral shifts are significant [27], [28], [29]. However, standard Transformers can be computationally demanding and memory-intensive, especially when processing long sequences, posing challenges for scalability and real-time applications.

Our design focuses primarily on the encoder portion of the Transformer architecture. The encoder’s main function is to transform the input sequence into a rich, contextualized representation that captures the meaning and relationships between spectral points. In our approach, we emphasize key steps critical for effectively encoding overlapping spectral data, ensuring that both local and global dependencies are accurately represented [30], [31].

- 1) *Input Representation:* In our case, each input spectrum is treated akin to a “sequence of spectral points,” where each data point corresponds to an intensity at a given wavelength. Because we handle numeric spectral data

rather than text, no vocabulary or one-hot encoding is used. Instead, raw spectral features are scaled and reshaped into a form suitable for the Transformer (1-D sequence) [32], [33].

- 2) *Positional Encoding*: We employ positional encoding. Transformers process all spectral points simultaneously and, therefore, lack inherent positional information about the order of data points in a sequence. To address this, positional encoding is employed to explicitly add unique positional information to each spectral embedding, enabling the model to understand the relative positions of spectral points across the input sequence. The positional encoding is calculated using sine and cosine functions, ensuring that each position has a distinct representation [34]. For a given position pos (starting from 0) and embedding dimension  $i$  (ranging from 0 to  $d_{\text{model}} - 1$ ), the encoding is defined [16] as follows:

$$\text{PE}(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (2)$$

$$\text{PE}(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right). \quad (3)$$

In these equations, sine functions are applied to even dimensions, while cosine functions are applied to odd dimensions, allowing each position to have a unique encoding. The frequency variation term,  $(1/10000^{(2i/d_{\text{model}})})$ , scales the frequency based on the dimension  $i$ , creating a spectrum of frequencies across dimensions. This variation enables the model to learn complex position-related patterns. The final representation of each spectral point combines its feature projection and positional encoding [16] as follows:

$$x_i = e_i + p_i \quad (4)$$

where  $x_i$  represents the final input for spectral point  $i$ ,  $e_i$  is the feature projection, and  $p_i$  is the positional encoding for the respective position. Adding positional encoding ensures that the model can distinguish between spectral points at different positions and leverage the smooth transitions provided by sine and cosine functions, which ensure that positions close to each other have similar encodings. This mechanism is crucial in tasks such as FBG spectrum analysis, where the relative position of overlapping peaks carries important information for wavelength detection. Positional encoding effectively injects position-dependent information into the model, ensuring that the self-attention mechanism understands the order of spectral samples.

- 3) *Multihead Self-Attention*: The self-attention block [see Fig. 3(a)] uses the number of heads of 2, allowing the model to focus on different parts of the spectrum simultaneously. This is particularly useful for overlapping signals, enabling the model to learn localized peak features while capturing a broader spectral context [30], [35]. For each spectral point, three vectors are computed through linear projections: Query ( $q_i$ ), Key ( $k_i$ ), and

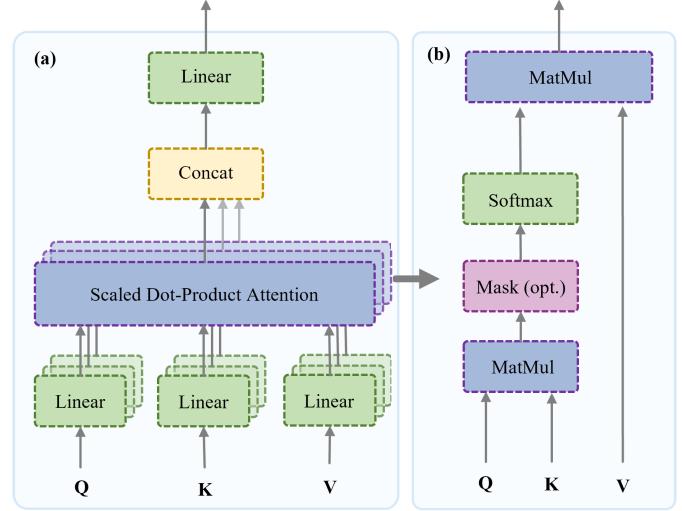


Fig. 3. (a) Multihead attention and (b) scaled dot-product attention [16].

Value ( $v_i$ ) [16], [27]

$$q_i = x_i W^Q, \quad k_i = x_i W^K, \quad v_i = x_i W^V \quad (5)$$

where  $W^Q$ ,  $W^K$ , and  $W^V$  are learnable weight matrices. Queries represent the current spectral point of focus, keys compare against other spectral points, and values aggregate relevant information. In addition, scaled dot-product attention [see Fig. 3(b)] further details the mathematical process by which the attention weights are calculated. The attention mechanism computes a weighted sum of the values, where the weights are determined by the compatibility of queries and keys.

The attention scores matrix ( $S$ ) is calculated [16] as follows:

$$S = \frac{QK^\top}{\sqrt{d_k}} \quad (6)$$

where  $Q$  is the matrix of queries ( $n \times d_k$ ) and  $K^\top$  is the transpose of the keys matrix ( $d_k \times n$ ). This results in an  $n \times n$  matrix, where each element ( $S_{ij}$ ) represents the score between spectral point  $i$  and spectral point  $j$ . To prevent excessively large dot product values that may lead to unstable gradients, the scores are scaled by the square root of the key dimension ( $d_k$ ) following the attention mechanism in [16].

$$S_{\text{scaled}} = \frac{S}{\sqrt{d_k}}. \quad (7)$$

The scaled scores are then passed through a softmax function to normalize them into attention weights ( $A_{ij}$ ) [16]

$$A_{ij} = \text{softmax}(S_{\text{scaled}})_{ij} = \frac{\exp(S_{\text{scaled},ij})}{\sum_{k=1}^n \exp(S_{\text{scaled},ik})}. \quad (8)$$

Finally, the attention output ( $z_i$ ) for each spectral point is computed as a weighted sum of the value vectors ( $v_j$ ) [27]

$$z_i = \sum_{j=1}^n A_{ij} v_j. \quad (9)$$

In this mechanism, the attention weights ( $A_{ij}$ ) indicate the relevance of spectral point  $j$  to spectral point  $i$ . By summing over all spectral points, the model integrates information from the entire sequence, weighted by their computed relevance. This multihead self-attention mechanism enhances the model's ability to analyze both local features and global relationships, making it highly effective for FBG spectrum analysis, where precise detection of overlapping peaks is essential.

- 4) *Add and Norm*: We utilize residual connections and layer normalization to stabilize the training process. Residual connections ensure that the original information from the input is preserved while integrating the outputs from the multihead attention mechanism, which is particularly important for retaining fine-grained peak details in spectral data. After the multihead attention output ( $Z$ ), we apply a residual connection by adding the original input ( $X$ ) back to the attention output [36]

$$X' = X + Z. \quad (10)$$

Residual connections help mitigate the vanishing gradient problem by improving the flow of gradients through the network, which is essential for deep architectures. They also simplify the learning process by allowing the model to focus on learning incremental changes over an identity mapping rather than trying to learn the entire transformation from scratch. Next, we apply layer normalization to the residual output ( $X'$ ) [36]

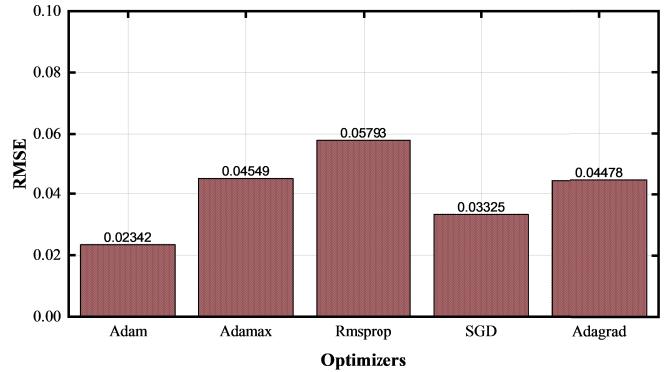
$$X'' = \text{LayerNorm}(X') \quad (11)$$

where  $X''$  is the normalized output fed into the next layer, combining residual and normalization benefits for stable and efficient training. Layer normalization normalizes the inputs across the features (dimensions) for each spectral point, reducing internal covariate shifts and ensuring that the input distribution remains consistent across layers. This normalization step stabilizes the learning process and accelerates convergence, providing a smoother optimization pathway. By combining residual connections and layer normalization, the model effectively balances the integration of attention outputs with original input features, ensuring both stability and efficient gradient flow during training. This approach is critical for maintaining fine-grained details in spectral data analysis while preventing instability in deeper layers.

- 5) *Positionwise FFN*: After the attention mechanism, a feed-forward network (FFN) applies a nonlinear transformation independently to each spectral point. This step refines the attention outputs, enhancing the model's ability to capture complex patterns [37]. The FFN consists of two linear transformations with a rectified linear unit (ReLU) activation in between [16]

$$F(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (12)$$

where  $W_1$  and  $W_2$  are learnable weight matrices, while  $b_1$  and  $b_2$  are bias vectors. The ReLU activation introduces nonlinearity, allowing the network to learn



**Fig. 4.** Performance comparison with other optimizers.

intricate relationships between features [38], [39]. This positionwise transformation complements the attention mechanism by adding depth and refinement to spectral representations. In our design, we use an ReLU activation with a hidden dimension of 102 to balance efficiency and expressiveness.

After the Transformer encoder output, we apply a Conv1D layer with parameters (filters = 24, kernelsize = 2, and activation = "relu"). This convolutional step focuses on capturing local peak features while condensing the sequence dimension. To improve generalization and address overfitting, BatchNormalization and Dropout (dropoutrate = 0.5) follow the convolutional layer. The CNN output is then flattened and passed through a final dense layer (output size = 4), directly predicting the central wavelengths of the four sensors simultaneously. This integration reduces computational demands while effectively capturing both global relationships and local patterns. To quantify performance, we employ standard error measures for wavelength regression, including mean square error (mse), root mse (RMSE), mean percentage error (MPE), mean absolute error (MAE), and MPE. Among these, RMSE serves as a key metric for evaluating prediction accuracy. Mathematically, RMSE is defined [40], [41] as follows:

$$\text{RMSE} = \sqrt{\frac{1}{d} \sum_{k=1}^d (Y_{\text{est}} - X_{\text{act}})^2} \quad (13)$$

where  $Y_{\text{est}}$  represents the predicted central wavelengths,  $X_{\text{act}}$  denotes the actual central wavelengths, and  $d$  is the number of test data points. This measure provides a clear indication of how closely the model's predictions align with the true values, with lower RMSE values indicating better performance. We tested multiple optimizers, including Adamax, root mean square propagation (RMSprop), stochastic gradient descent (SGD), and adaptive gradient algorithm (Adagrad). Among these, adaptive moment estimation (Adam) consistently yielded the lowest RMSE in this domain, as illustrated in Fig. 4, demonstrating its superior performance in optimizing our model effectively across varying conditions. In parallel, we explored different activation functions (sigmoid, hyperbolic tangent (Tanh), and ReLU) and found that ReLU outperformed the others, especially in scenarios with overlapping signals, where its ability to mitigate vanishing gradient issues and

**TABLE II**  
TRANSFORMERS CNN HYPERPARAMETERS

Parameter	Value
Learning rate	0.0001
Batch size	36
Dropout rate	0.5
L2 Regularization	0.005
Number of attention heads	2
Feed-forward dimension	102
Gradient clipping	1.0
Optimizer	Adam
Embedding dimension	22
Number of transformer blocks	1
Epochs	1000
Activation function	relu

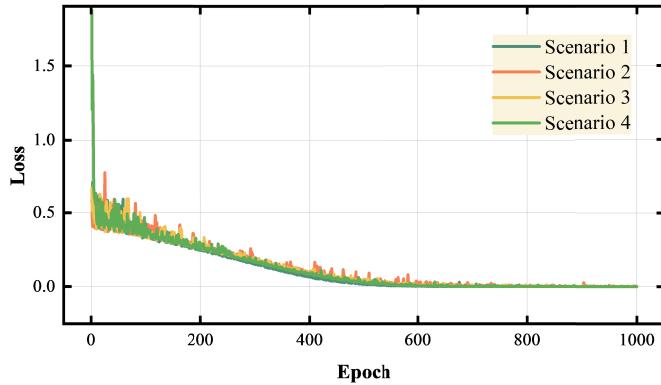


Fig. 5. Loss of Transformer CNN for four scenarios.

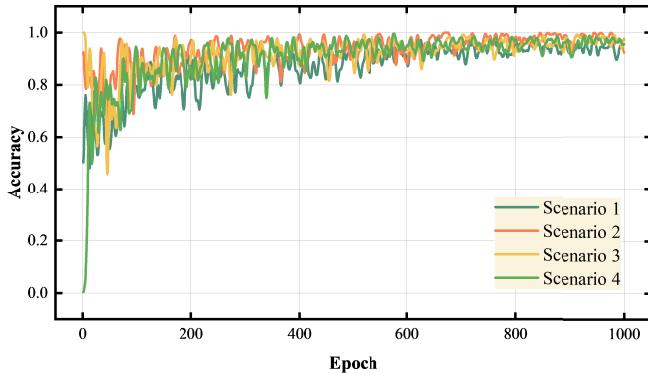


Fig. 6. Accuracy of Transformer CNN for four scenarios.

efficiently handle sparse activations proved highly effective [42]. The final architecture employs ReLU activation functions and the Adam optimizer, with key hyperparameters summarized in Table II. Once trained on the four scenarios described in our experiments, the model yields predictions for each scenario's overlapping spectra. Fig. 5 illustrates the training loss, offering insight into the learning efficiency across scenarios, whereas Fig. 6 displays the accuracy achieved, indicating the model's effectiveness in resolving spectral overlaps with high precision. Additionally, a comparison of inference times across scenarios is presented in Fig. 7, which highlights the execution speed of the Transformer CNN model.

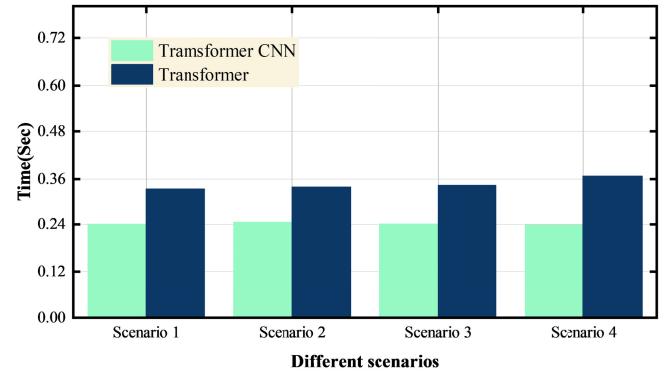


Fig. 7. Testing time comparison.

#### IV. RESULTS AND DISCUSSION

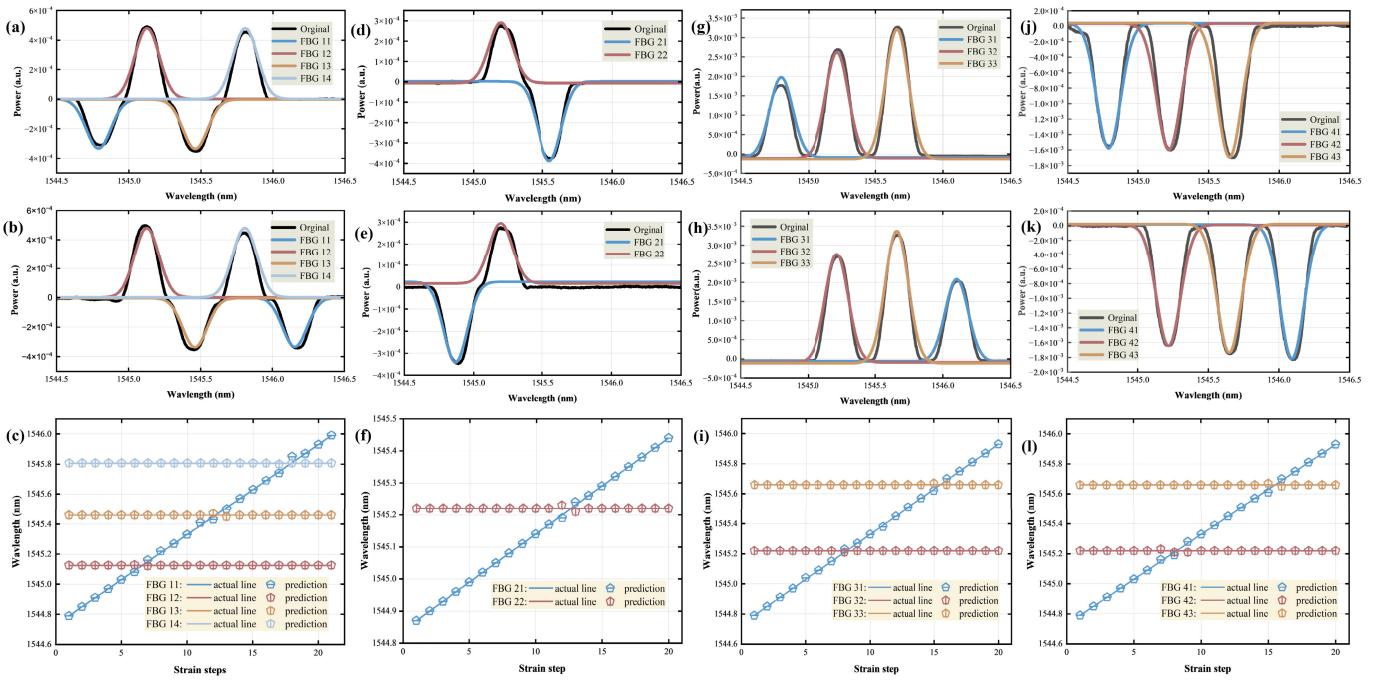
All experiments were conducted on a PC equipped with an Intel<sup>1</sup> Core<sup>2</sup> i7-9700K CPU running at 3.60 GHz (eight cores and eight logical processors) and 16.0 GB of RAM. A single NVIDIA GeForce RTX 2080 Ti GPU (11 GB of dedicated memory) was utilized to accelerate training and inference. Software versions included Python 3.8.20, Keras 2.10.0, and TensorFlow-GPU 2.10.0. We first evaluated the ability of the Transformer CNN model to detect Bragg wavelengths accurately under various strain conditions across four scenarios (Scenarios 1–4). In each scenario, a specific sensor (e.g., FBG11) was strained in multiple increments, creating partial or complete overlaps with other FBG sensors.

In Scenario 1, Fig. 8(a) shows the initial strain step, where FBG11 is observed as a transmission spectrum. After 21 strain steps, it progressively overlaps with all three FBGs, with Fig. 8(b) illustrating the final strain step, where the overlapping peaks are fully developed. In scenario 2, Fig. 8(d) presents the first strain step, where FBG21 is maintained as a transmission FBG. After 20 strain steps, it fully overlaps with FBG22, following an intermediate overlap with a reflective FBG. The final strain step is depicted in Fig. 8(e), showing the completed overlap. In Scenario 3, Fig. 8(g) represents the initial configuration, where all FBGs are reflective FBGs before strain is applied. As strain increases, overlapping occurs with two FBG sensors, and Fig. 8(h) displays the final step, capturing the resulting spectral overlap. In Scenario 4, Fig. 8(j) illustrates the initial configuration, where all FBGs are transmission FBGs before strain is applied. As the strain progresses, overlapping occurs with two FBG sensors, and Fig. 8(k) represents the final strain step. To present the full progression of strain steps across all scenarios, Fig. 8(c), (f), (i), and (l) collectively illustrates the wavelength detection and the actual line. In each scenario, the Transformer CNN consistently distinguishes overlapping spectra and accurately estimates the central wavelength for each sensor.

Training time was measured over 1000 epochs for both the Transformer CNN and a baseline Transformer model. As summarized in Fig. 9, the Transformer CNN demonstrated

<sup>1</sup>Registered trademark.

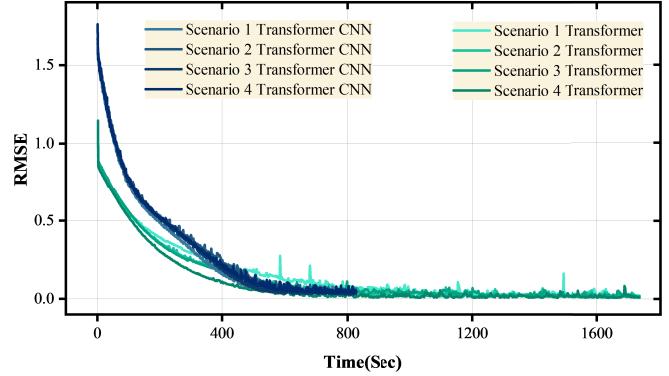
<sup>2</sup>Trademarked.



**Fig. 8.** Progression of strain steps and predicted wavelengths across four scenarios. (a) and (b) Two transmission FBGs and two reflective FBGs first and last step with their original and predicted Bragg wavelengths. (c) Comparison of actual and predicted Bragg wavelengths for Scenario 1. (d) and (e) One transmission FBG and one reflective FBG first and last step with their original and predicted Bragg wavelengths. (f) Comparison of actual and predicted Bragg wavelengths for Scenario 2. (g) and (h) Three reflective FBGs first and last step with their original and predicted Bragg wavelengths. (i) Comparison of actual and predicted Bragg wavelengths for Scenario 3. (j) and (k) Three transmission FBGs first and last step with their original and predicted Bragg wavelengths. (l) Comparison of actual and predicted Bragg wavelengths for Scenario 4.

substantially lower training times across all scenarios. In Scenario 1, Transformer CNN completed training in 711.56 s, while Transformer required 1575.25 s, making it approximately 2.21 times faster. Similarly, in Scenario 2, Transformer CNN finished in 677.28 s, outperforming Transformer's 1535.36 s with an efficiency improvement of 2.27 times. In Scenario 3, Transformer CNN recorded a training time of 684.72 s compared to 1531.11 s for Transformer, reflecting a 2.24 $\times$  improvement. Finally, in Scenario 4, Transformer CNN required 680.48 s, whereas Transformer needed 1531.05 s, achieving a 2.25 $\times$  speedup. These results underscore that integrating CNN layers into the Transformer encoder effectively reduces computational overhead without compromising the model's performance. Fig. 7 compares inference (testing) times for the Transformer CNN and Transformer models across four scenarios. In Scenario 1, Transformer CNN completed testing in 0.24013 s, while Transformer required 0.33477 s, making it approximately 1.39 times faster. In scenario 2, Transformer CNN achieved an inference time of 0.24509 s compared to 0.33910 s for Transformer, reflecting a 1.38 $\times$  improvement. In scenario 3, Transformer CNN recorded 0.24097 s, outperforming Transformer's 0.34311 s with a 1.42 $\times$  speedup. Finally, in scenario 4, Transformer CNN achieved the fastest inference time of 0.23774 s, significantly outperforming Transformer's 0.36714 s with a 1.54 $\times$  improvement. These results highlight the consistent efficiency of the Transformer CNN architecture in reducing inference time, demonstrating its suitability for real-time or rapid deployment applications where detection speed is essential.

Fig. 10 presents the performance metrics of the CNN and Transformer CNN across all four experimental scenarios. In terms of mse, the model achieved 0.00032 in



**Fig. 9.** Training time comparison between Transformer CNN and Transformer across four scenarios.

Scenario 1, 0.00047 in Scenario 2, 0.00019 in Scenario 3, and 0.00017 in Scenario 4. Corresponding MAE values were 0.01361, 0.01563, 0.00666, and 0.01001, while RMSE values were 0.01802, 0.02169, 0.00719, and 0.01309 for Scenarios 1–4, respectively. These consistently low values indicate the model's robustness in accurately resolving overlapping spectral signals and predicting central wavelengths under various strain configurations. Compared with a conventional CNN model, the Transformer CNN demonstrates a significant performance advantage. The CNN's mse values were notably higher across all scenarios: 0.01456 in Scenario 1, 0.00720 in Scenario 2, 0.03817 in Scenario 3, and 0.01704 in Scenario 4. Similarly, in terms of MPE, the Transformer CNN showed smaller and more stable values: -0.00061, -0.00097, -0.00016, and -0.00029 across Scenarios 1–4, respectively. These slightly negative MPE values indicate mild overprediction, which remains well within acceptable bounds and

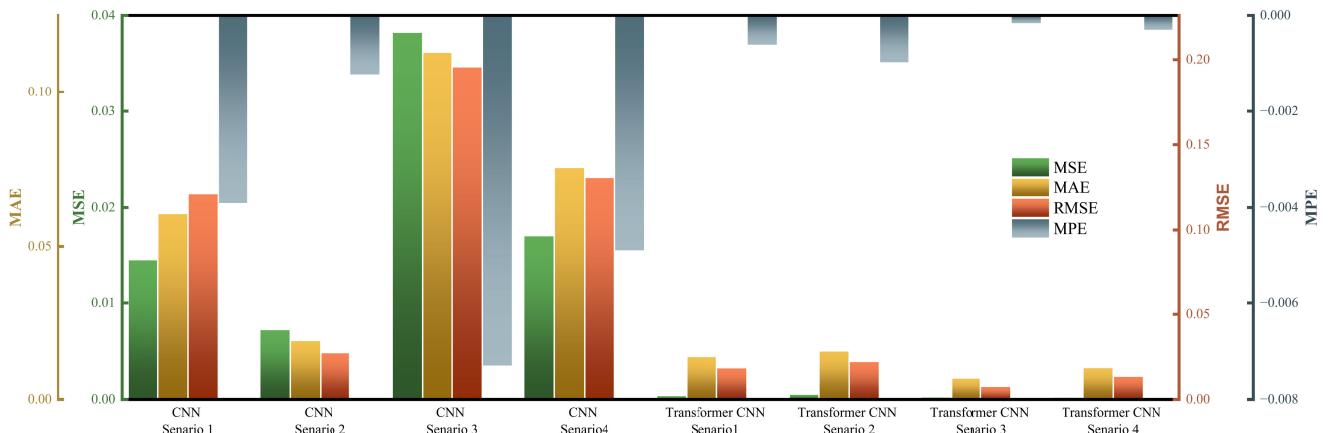


Fig. 10. Performance metrics for CNN and Transformer CNN across scenarios.

demonstrates better consistency compared to the CNN model. In contrast, the CNN exhibited larger deviations, with MPEs of  $-0.00390$ ,  $-0.00122$ ,  $-0.00730$ , and  $-0.00487$ . These results clearly highlight the superior accuracy and reliability of the Transformer CNN in high-density FBG configurations. Moreover, when benchmarked against other deep learning models such as GRU and LSTM, the Transformer CNN still performs better, showing significantly lower mse values across all tested scenarios. The GRU model reported an mse of 0.03389, while the LSTM achieved 0.02162. In comparison, the Transformer CNN achieved a much lower average mse of 0.00029, demonstrating its enhanced ability to demodulate overlapping spectra efficiently and accurately.

The Transformer CNN hybrid architecture combines self-attention for capturing long-range dependencies in overlapping FBG spectra and CNN layers for localized feature extraction, enabling faster and more precise training and inference. As shown in Fig. 9, the training loss curves demonstrate faster convergence and stable learning behavior, reflecting the synergy between these components. Among the optimizers tested, Adam was chosen for its stable convergence and consistent performance. These architectural choices reduce computational resource requirements, making the model suitable for large-scale deployments and real-time wavelength detection in scenarios with heavy spectral overlap.

## V. CONCLUSION

In this work, we presented a novel FBG sensor architecture that leverages both reflective and transmission wavelengths to significantly increase network capacity within a single optical fiber link. By integrating FSO, the system overcomes traditional geographical constraints and enables reliable long-distance data transmission, making it suitable for remote and distributed sensing applications. The overlapping Bragg wavelength problem, common in dense FBG networks, was addressed by implementing a Transformer CNN that combines the self-attention capabilities of a Transformer encoder with convolutional layers for local feature extraction. Extensive testing across multiple scenarios, each applying incremental strain to one or more FBGs, demonstrated that the Transformer CNN consistently outperforms a purely Transformer-based model. It not only reduces training and inference time by up to

a factor of 2 in some cases but also achieves lower mse values, indicating more accurate wavelength detection when spectra overlap. These performance gains are attributed to the model's ability to capture both global spectral dependencies and localized peak structures within overlapping signals. Overall, the proposed methodology offers a scalable and efficient solution for advanced FBG sensing. By utilizing both reflective and transmission modes, the system effectively doubles the number of deployable sensors within a given spectral window without compromising response time or precision. In addition, the proposed Transformer CNN demodulation architecture significantly improves detection accuracy and reduces computational time, addressing the challenges posed by overlapping Bragg wavelengths in dense FBG networks. Future implementations stand to benefit from this architecture's reduced overhead and high scalability, paving the way for real-time and large-scale deployments in industrial process control, structural health monitoring, and other applications where dense and flexible optical sensor networks are required.

## REFERENCES

- [1] G. Meltz, W. W. Morey, and W. H. Glenn, "Formation of Bragg gratings in optical fibers by a transverse holographic method," *Opt. Lett.*, vol. 14, no. 15, pp. 823–825, 1989.
- [2] Y. J. Rao, "Recent progress in applications of in-fibre Bragg grating sensors," *Opt. Lasers Eng.*, vol. 31, no. 4, pp. 297–324, Apr. 1999.
- [3] Y. C. Manie et al., "Intensity and wavelength division multiplexing FBG sensor system using a Raman amplifier and extreme learning machine," *J. Sensors*, vol. 2018, pp. 1–11, Sep. 2018.
- [4] S. Xu, F. Xing, H. Yang, and G. Li, "Multiple compensation and linear fitting demodulation for Raman distributed temperature sensors," *IEEE Sensors J.*, vol. 24, no. 23, pp. 39002–39010, Dec. 2024.
- [5] W. Naku, C. Zhu, A. K. Nambisan, R. E. Gerald, and J. Huang, "Machine learning identifies liquids employing a simple fiber-optic tip sensor," *Opt. Exp.*, vol. 29, no. 24, pp. 40000–40014, Nov. 2021.
- [6] Z. Zhang, H. Wu, C. Zhao, and M. Tang, "High-performance Raman distributed temperature sensing powered by deep learning," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 654–659, Jan. 15, 2021.
- [7] M. A. Arockiyadoss et al., "Self-healing fiber Bragg grating sensor system using free-space optics link and machine learning for enhancing temperature measurement," *Electronics*, vol. 13, no. 7, p. 1276, Mar. 2024.
- [8] P. Kumar, C.-K. Yao, M. Augustine Arockiyadoss, A. Mulatje Dehnaw, C.-Y. Chang, and P.-C. Peng, "YOLO-v7 improved with adam optimizer: Realizing orphaned fiber Bragg grating to sense superimposed personalized dynamic strain," *IEEE Sensors J.*, vol. 24, no. 23, pp. 39923–39933, Dec. 2024.

- [9] Y. C. Manie, C.-K. Yao, T.-Y. Yeh, Y.-C. Teng, and P.-C. Peng, "Laser-based optical wireless communications for Internet of Things (IoT) application," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24466–24476, Dec. 2022.
- [10] H. Kaushal and G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 57–96, 1st Quart., 2017.
- [11] A. Carrasco-Casado and R. Mata-Calvo, "Free-space optical links for space communication networks," 2020, *arXiv:2012.13166*.
- [12] C.-K. Yao et al., "Fiber/Free-space optics with open radio access networks supplements the coverage of millimeter-wave beamforming for future 5G and 6G communication," *Fibers*, vol. 13, no. 4, p. 39, Apr. 2025. [Online]. Available: <https://www.mdpi.com/2079-6439/13/4/39>
- [13] S. T. Hayle et al., "Hybrid of free space optics communication and sensor system using IWDM technique," *J. Lightw. Technol.*, vol. 40, no. 17, pp. 5862–5869, Sep. 1, 2022.
- [14] P. Liu, Z. Xu, Y. Wang, L. Jiang, and M. Wu, "Optimized demodulation of highly overlapped fiber Bragg grating sensor networks using two-stage deep learning approach," *Opt. Eng.*, vol. 63, no. 4, Apr. 2024, Art. no. 046103.
- [15] G. Rudloff and M. A. Soto, "Multipeak wavelength detection of spectrally overlapped fiber Bragg grating sensors through a CNN-based autoencoder," *IEEE Sensors J.*, vol. 24, no. 13, pp. 20674–20687, Jul. 2024.
- [16] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, Long Beach, CA, USA, Jun. 2017, pp. 5998–6008.
- [17] Y. Bai, J. Mei, A. Yuille, and C. Xie, "Are transformers more robust than CNNs," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, Nov. 2021, pp. 26831–26843.
- [18] M. Jianwen, M. Lunlin, Y. Hua, L. Leping, and C. Lingping, "CNN with embedding transformers for person reidentification," *Math. Problems Eng.*, vol. 2023, no. 1, Jan. 2023, Art. no. 4591991.
- [19] Y. Yu, L. Lui, H. Tam, and W. Chung, "Fiber-laser-based wavelength-division multiplexed fiber Bragg grating sensor system," *IEEE Photon. Technol. Lett.*, vol. 13, no. 7, pp. 702–704, Jul. 2001.
- [20] Y. C. Manie et al., "Enhancement of the multiplexing capacity and measurement accuracy of FBG sensor system using IWDM technique and deep learning algorithm," *J. Lightw. Technol.*, vol. 38, no. 6, pp. 1589–1603, Mar. 15, 2020.
- [21] P.-C. Peng, J.-H. Lin, H.-Y. Tseng, and S. Chi, "Intensity and wavelength-division multiplexing FBG sensor system using a tunable multiport fiber ring laser," *IEEE Photon. Technol. Lett.*, vol. 16, no. 1, pp. 230–232, Jan. 2004.
- [22] S. D. Bogale et al., "Demodulating optical wireless communication of FBG sensing with turbulence-caused noise by stacked denoising autoencoders and the deep belief network," *Electronics*, vol. 13, no. 20, p. 4127, Oct. 2024. [Online]. Available: <https://www.mdpi.com/2079-9292/13/20/4127>
- [23] S. D. Bogale, C.-K. Yao, Y. C. Manie, Z.-G. Zhong, and P.-C. Peng, "Wavelength-dependent Bragg grating sensors cascade an interferometer sensor to enhance sensing capacity and diversification through the deep belief network," *Appl. Sci.*, vol. 14, no. 16, p. 7333, Aug. 2024. [Online]. Available: <https://www.mdpi.com/2076-3417/14/16/7333>
- [24] B. Tartoussi, N. Tartoussi, B. Altayebrafei, J. Charafeddine, W. M. R. Shakir, and N. Chendeb, "Optimizing FSO system performance: Integrating machine learning for security and reliability," in *Proc. 5th IEEE Middle East North Afr. Commun. Conf. (MENACOMM)*, Byblos, Lebanon, Feb. 2025, pp. 1–8.
- [25] B. Peng et al., "RWKV: Reinventing RNNs for the transformer era," 2023, *arXiv:2305.13048*.
- [26] A. Géron, *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly Media, 2019.
- [27] S. Islam et al., "A comprehensive survey on applications of transformers for deep learning tasks," *Expert Syst. Appl.*, vol. 241, May 2024, Art. no. 122666. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423031688>
- [28] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. NIPS*, vol. 34, Dec. 2021, pp. 22419–22430.
- [29] G. Tucudean, M. Bucos, B. Dragulescu, and C. D. Calleanu, "Natural language processing with transformers: A review," *PeerJ Comput. Sci.*, vol. 10, p. e2222, Aug. 2024.
- [30] H. Guo and W. Liu, "S3L: Spectrum transformer for self-supervised learning in hyperspectral image classification," *Remote Sens.*, vol. 16, no. 6, p. 970, Mar. 2024.
- [31] M. Ahmad, M. Usama, A. M. Khan, S. Distefano, H. A. Altuwaijri, and M. Mazzara, "Spatial-Spectral transformer with conditional position encoding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [32] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5522214.
- [33] S. Varahagiri, A. Sinha, S. R. Dubey, and S. K. Singh, "3D-convolution guided spectral-spatial transformer for hyperspectral image classification," 2024, *arXiv:2404.13252*.
- [34] Q. Zhang et al., "An empirical study on the impact of positional encoding in transformer-based monaural speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 1001–1005.
- [35] T. Zhang, B. Luo, and G. Wang, "Improving vision transformers by overlapping heads in multi-head self-attention," 2024, *arXiv:2410.14874*.
- [36] P. S. Huggins, H. F. Chen, P. N. Belhumeur, and S. W. Zucker, "Finding folds: On the appearance and identification of occlusion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR*, vol. 2, Jun. 2001, pp. 718–725.
- [37] S. Lu, M. Wang, S. Liang, J. Lin, and Z. Wang, "Hardware accelerator for multi-head attention and position-wise feed-forward in the transformer," in *Proc. Int. Syst.-Chip Conf.*, 2020, pp. 84–89.
- [38] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Feb. 2022.
- [39] Y. Lu et al., "Understanding and improving transformer from a multi-particle dynamic system point of view," 2019, *arXiv:1906.02762*.
- [40] L. Hong, Y. Wang, Y. Shi, and J. Cai, "Research on denoising method for low-frequency fiber Bragg grating sensing signal based on ICEEMDAN and SampEn combined with WT," *J. Opt.*, vol. 53, pp. 1–11, Dec. 2024.
- [41] K. Pereira, W. Coimbra, R. Lazaro, A. Frizera-Neto, C. Marques, and A. G. Leal-Junior, "FBG-based temperature sensors for liquid identification and liquid level estimation via random forest," *Sensors*, vol. 21, no. 13, p. 4568, Jul. 2021.
- [42] E. C. Seyrek and M. Uysal, "A comparative analysis of various activation functions and optimizers in a convolutional neural network for hyperspectral image classification," *Multimedia Tools Appl.*, vol. 83, no. 18, pp. 53785–53816, Nov. 2023.



**Michael Augustine Arockiyadoss** received the B.Sc. degree in computer applications and the M.Sc. degree in computer science from the St. Joseph's College, Trichy, India, in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree with the Department of Electro-Optical Engineering, National Taipei University of Technology, Taipei, Taiwan.

His research focuses on optical fiber sensors, optical communication, the IoT, cloud computing, and machine learning.



**Cheng-Kai Yao** received the B.S. degree from the Department of Electro-Optical Engineering, National United University, Miaoli, Taiwan, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electro-Optical Engineering, National Taipei University of Technology, Taipei, Taiwan.

His research interests include optic communication systems and fiber optic sensors.



**Pradeep Kumar** received the B.S. degree in information technology and the M.Tech. degree in computer science and engineering from Guru Gobind Singh Indraprastha University, Delhi, India, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree in electro-optical engineering with the National Taipei University of Technology, Taipei, Taiwan.

His research interests include optical communication, fiber sensors, generative AI, deep learning, machine learning, and image processing.



**Amare Mulatie Dehnaw** received the B.S. degree in information science from Jimma University, Jimma, Ethiopia, in 2012, the M.Sc. degree in information science from Addis Ababa University, Addis Ababa, Ethiopia, in 2015, and the Ph.D. degree in electrical engineering and computer science from the National Taipei University of Technology, Taipei, Taiwan, in 2023.

He is currently a Postdoctoral Researcher with the Department of Electro-Optical Engineering, National Taipei University of Technology. His

research interests include fiber optic sensors, optical wireless communications, and deep learning.



**Yao-Jen Chung** received the B.S. degree from the Department of Electro-Optical Engineering, Feng Chia University, Taichung, Taiwan, in 2024. He is currently pursuing the M.S. degree with the Department of Electro-Optical Engineering, National Taipei University of Technology, Taipei, Taiwan.

His research interests include optic communication systems and fiber optic sensors.



**Peng-Chun Peng** received the Ph.D. degree from the Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 2005.

He is currently a Full Professor and the Chairman with the Department of Electro-Optical Engineering, National Taipei University of Technology, Taipei, Taiwan. He has authored more than 150 journal articles and holds nine U.S. patents. His publications have been cited more than 4200 times, resulting in an H-index of 33.

Dr. Peng was invited to the technical program committees of several prestigious conferences, including the Optical Fiber Communication Conference (OFC), OptoElectronics and Communications Conference (OECC), International Topical Meeting on Microwave Photonics (MWP), Asia Communications and Photonics Conference (ACP), and Wireless and Optical Communications Conference (WOCC).