

Article

A Hybrid Soot-MixFormer-Based Reconstruction Model for 2D Soot Spatial Distribution Inversion

Zhijie Huang¹ , Xiansong Fu², Shouxiang Lu¹ and Wenbin Yao^{1,*}¹ State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230009, China; hzj1241122657@mail.ustc.edu.cn (Z.H.); sxlu@ustc.edu.cn (S.L.)² Yichang Fire and Rescue Division, Hubei Provincial Fire and Rescue Department, Yichang 443000, China; 18771875312@163.com

* Correspondence: wbyao@ustc.edu.cn

Abstract

Accurate measurement of the 2D soot spatial distribution is vital for optimizing combustion efficiency and reducing pollutant emissions. While 1D laser extinction (LE) is robust and cost-effective, it provides only line-of-sight integrated information, lacking the spatial resolution required to resolve complex soot topologies. We propose Soot-MixFormer, a hybrid deep learning model designed for the high-fidelity inversion of 2D soot distributions from 1D extinction data. The architecture integrates CNN-based local feature extraction with Transformer-based global dependency modeling. Key innovations include a dynamic decoupled generation head and a Dual-Axial Gated Refinement (DAGR) module coupled with a physical hard constraint layer to ensure mass conservation and physical consistency. Experimental results demonstrate that Soot-MixFormer significantly outperforms baseline MLP and CNN models, achieving a Structural Similarity Index (SSIM) of 0.800 and a Pearson Correlation Coefficient (PCC) of 0.915, and a highly suppressed Root Mean Square Error (RMSE) representing less than 10% relative error in high-concentration zones. Furthermore, the model exhibits exceptional robustness, maintaining a cosine similarity above 0.72 even under 10% simulated measurement noise. The model is highly efficient, with only 0.97 M parameters and a real-time inference speed of ~246 FPS. This study provides a novel, low-cost diagnostic paradigm for real-time, high-accuracy monitoring of soot fields in industrial combustion environments, effectively bridging the gap between simple 1D sensing and complex 2D spatial reconstruction.

Keywords: soot; 2D reconstruction; laser extinction; deep learning; physical constraint

Academic Editor: Ziyu Wang

Received: 9 March 2026

Revised: 7 April 2026

Accepted: 16 April 2026

Published: 27 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Soot, a major product of incomplete combustion under oxygen-limited conditions, plays a critical role in combustion heat transfer, pollutant formation, and flame radiation. Its concentration and spatial distribution directly affect radiative characteristics, thermal efficiency, and emission behavior in practical combustion systems [1–4]. Therefore, accurate characterization of soot spatial distributions is essential for combustion diagnostics, kinetic-model validation, and the development of cleaner and more efficient combustion technologies.

Traditional soot diagnostic techniques can generally be classified into intrusive sampling methods and non-intrusive optical methods. Intrusive approaches inevitably disturb the local flow and temperature fields [5,6], making them unsuitable for accurate in situ

measurements. By contrast, optical diagnostics infer soot properties from light absorption, scattering, or radiation and are therefore more widely adopted. Among them, laser-induced incandescence (LII) provides high sensitivity and high spatial-temporal resolution [7,8]. In addition, soot radiation reconstruction methods, such as dual-color, multi-color, and computed tomography (CT), have been developed to recover multidimensional flame structures from multi-angle projections [9–12]. However, these methods usually require complex optical configurations, multiple viewing angles, and computationally expensive inversion procedures, which limit their deployment in dynamic or industrial environments. Even recent advances, such as deep-learning-assisted 3D soot reconstruction and tomographic single-shot time-resolved LII, still rely on sophisticated multi-view measurement systems [13,14].

In contrast, laser extinction (LE) is attractive because of its simple hardware, low cost, and strong robustness [15,16]. Nevertheless, a single-path extinction measurement only provides line-of-sight integrated soot information and cannot directly resolve the transverse spatial distribution [17,18]. As a consequence, reconstructing a two-dimensional soot field from a one-dimensional extinction signal is a severely ill-posed inverse problem, since multiple 2D distributions may correspond to the same integrated measurement and small perturbations can be greatly amplified during inversion. To obtain reliable paired data for this challenging task, experimentally calibrated LE–LII measurements have been employed in soot studies, and the dataset used in this work is built upon the experimental framework reported by Yao [19]. Moreover, the combination of LII and LE has been widely recognized as an effective strategy for quantitative soot volume fraction measurement and calibration [20]. Therefore, an important unresolved problem is whether a simple and robust LE signal can be used as the front-end measurement while a data-driven model reconstructs the corresponding high-fidelity 2D soot distribution.

In recent years, deep learning has shown strong potential for solving inverse problems in science and engineering [21–23]. In combustion diagnostics, convolutional neural networks (CNNs) [24] can learn mappings from limited observations to high-dimensional physical fields [25]. For example, Zhang et al. employed a conditional generative adversarial network to predict 2D LII images from line-of-sight soot radiation images [26]. Another study used CNNs to predict radial soot distributions from axis-line extinction signals under an axisymmetric assumption [27]. Simonyan et al. further demonstrated the feasibility of CNN-based flame temperature field reconstruction [28]. More recently, Chen et al. reported machine-learning-based reconstruction of two-dimensional flame temperature and soot concentration fields from partial discrete measurements [29]. These studies confirm the promise of data-driven combustion-field reconstruction; however, most rely on image-based inputs, partial field observations, or restrictive symmetry assumptions, rather than directly addressing general 1D-to-2D soot inversion.

To improve physical consistency, physics-informed deep learning has emerged as an important research direction. Wang et al. combined physics-informed learning with neural operator ideas to reconstruct soot-related fields in acoustically forced laminar sooting flames [30]. Li et al. introduced a physics-informed neural network for nonlinear laser absorption tomography [31], while Molnar et al. developed an unsupervised neural-implicit tomography framework for quantitative imaging of unsteady flames from sparse measurements [32]. Although these studies demonstrate the value of embedding physical knowledge into neural reconstruction, most focus on temperature, species concentration, or tomographic thermochemical-field reconstruction rather than direct recovery of a general 2D soot field from a single 1D extinction vector. In addition, many physics-informed methods impose physical laws through soft penalty terms, which may introduce a trade-off between data fidelity and physics consistency.

To address these limitations, this paper proposes Soot-MixFormer, a hybrid physics-constrained framework for reconstructing two-dimensional soot spatial distributions directly from one-dimensional laser extinction signals. Unlike previous image-based or multi-view approaches, the proposed method explicitly targets the highly ill-posed vector-to-matrix inversion problem. Unlike standard soft-constraint strategies, it embeds a hard physical constraint into the forward process to guarantee projection consistency between the reconstructed 2D soot field and the measured 1D extinction signal. The framework combines CNN-based local feature extraction and Transformer-based global dependency modeling within a multi-scale feature pyramid, and further introduces a low-rank row-column decoupling strategy together with a lightweight Dual-Axial Gated Refinement (DAGR) module to improve structural representation and suppress reconstruction artifacts. Main contributions and innovations of this work are summarized as follows:

- **Novel Inversion Paradigm from Vector to Matrix:** Unlike conventional black-box spatial mapping approaches, we propose a pioneering reconstruction motivation that transforms the traditional 1D-to-2D inversion task into a structured vector-to-matrix generation process. To the best of our knowledge, this is the first model to explicitly decouple 2D soot concentration maps into rank-1 row and column vectors, significantly reducing computational complexity while preserving the inherent spatial correlation of soot topologies.
- **Hybrid Feature Pyramid Backbone and Cross-Attention Decoder:** Overcoming the inherent limitations of purely local CNNs or purely global Transformers, a sophisticated feature extraction backbone is designed by integrating 1D Convolutional Neural Networks (CNNs) with Transformer blocks within a multi-scale feature pyramid framework. This allows the model to simultaneously capture local concentration gradients and global structural dependencies. Furthermore, a novel decoder based on cross-attention mechanisms is specifically developed for the generated row and column vectors, enabling a more expressive and precise mapping from 1D latent features to 2D spatial coordinates.
- **Dual-Axial Gated Refinement and Physics-Constrained Decoding:** We introduce the Dual-Axial Gated Refinement (DAGR) module to selectively enhance spatial features and resolve fine-grained internal textures of the soot. Crucially, distinguishing our approach from standard Physics-Informed Neural Networks (PINNs) that rely on soft penalty terms, the decoding stage is anchored by a physical hard constraint layer derived from mass conservation laws. This integration ensures that the reconstructed soot distributions are not only data-accurate but also physically consistent with the original line-of-sight extinction measurements, effectively suppressing non-physical artifacts.

2. Data Acquisition and Processing

2.1. Experimental System

To ensure the authenticity of the experimental results and enhance their applicability in real-world scenarios, the proposed model was trained entirely on experimental data. The raw data primarily originated from the experiments conducted by Yao [23], which were rigorously processed prior to model training. The experimental setup comprises five main modules: (1) a turbulent diffusion flame burner, (2) a co-flow laminar burner, (3) a binary fuel mixing system, (4) a gas fuel supply system, and (5) an optical measurement system. Specifically, the burner features a central fuel nozzle with a diameter of 10 mm, surrounded by an air co-flow region to stabilize the flame. In the optical measurement system, the laser source operates at a stable power output of 200 mW, while the high-resolution camera's exposure time is strictly controlled to 1 ms to effectively freeze the instantaneous flame fluctuations. As illustrated in Figure 1, a stable laminar flame was

employed to calibrate the soot measurement system, while the turbulent diffusion soot burner was utilized for the primary experiments. The volumetric flow rates of different gas components entering the mixing chamber were accurately regulated using MKS mass flow controllers (MKS Instruments Inc., GE50A, Andover, MA, USA). The gaseous fuels utilized included methane (CH_4), ethylene (C_2H_4), propylene (C_3H_6), and propane (C_3H_8). During each experimental run, two types of gas fuels were supplied simultaneously. The gas flow rates ranged from 0 to 6.00 L/min and were controlled in fine increments of 0.1 to 0.6 L/min.

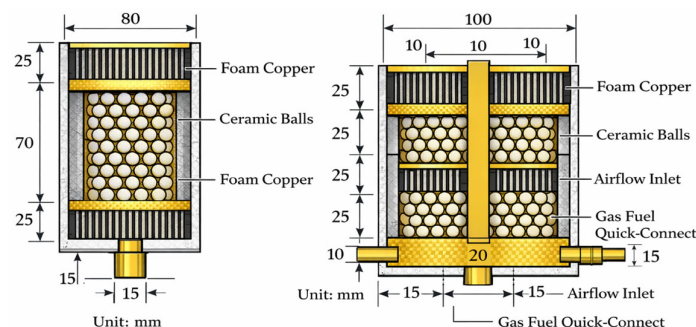


Figure 1. Schematic diagram of the turbulent diffusion soot burner (**left**) and the co-flow laminar burner (**right**).

2.2. Soot Acquisition Method

A Laser-Induced Incandescence (LII) system was employed to measure the two-dimensional (2D) distribution of the soot volume fraction (SVF) within the flame. Concurrently, a Laser Extinction (LE) system was utilized to calibrate the relative SVF values obtained from the LII system into quantitative absolute values.

The equipment for data collection in the experiment is arranged as shown in Figure 2. In the LII setup, a pulsed beam from an Nd: YAG laser (532 nm wavelength) was shaped into a thin laser sheet to illuminate the target soot field. Upon absorbing the laser energy, the soot particles rapidly heated to approximately 4000 K, emitting a strong incandescence signal. An ICCD camera, positioned perpendicular to the laser sheet and equipped with an appropriate bandpass filter, captured this radiation signal to form a 2D image. Within the linear response regime of the LII signal, the recorded signal intensity (grayscale value) is proportional to the local SVF. Because the effective measurement height of the laser sheet (0.04 m) was smaller than the overall flame height (which ranged from 0.1 to 0.5 m), measurements were conducted sequentially at different heights.

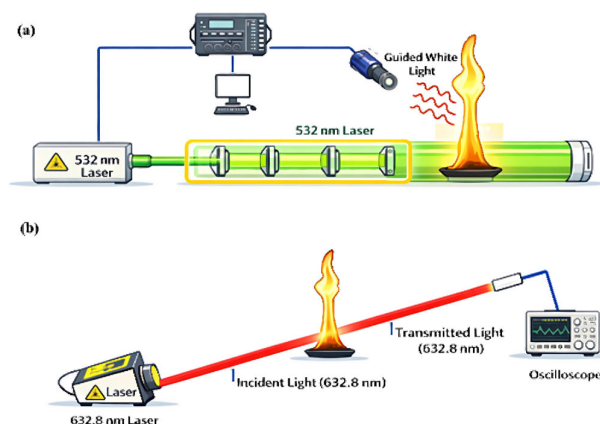


Figure 2. Data acquisition setups: (a) Laser-induced incandescence (LII) system and (b) Laser extinction (LE) system.

Synchronously with the LII measurements, a collimated continuous-wave He-Ne laser beam (632.8 nm wavelength) was directed through a specific chord of the combustion field. According to the Beer–Lambert law, the relationship between the transmitted and incident light intensities is defined as

$$K_{LLEM} = -\ln(I/I_0) \quad (1)$$

where K_{LLEM} is the cumulative SVF along the optical path obtained via the LE method. By correlating these two datasets, the path-integrated LE measurements were used to calibrate the 2D LII signal intensity distributions into absolute SVF fields.

The buoyant diffusion burner utilized in the experiments features a porous structure with a 0.08 m aperture. The 532 nm laser sheet intersected the flame axis to facilitate the measurement of the binary gas fuel buoyant diffusion flame. To cover the entire flame, 500 consecutive frames were acquired for each 0.04 m height segment, and the data for each segment were subsequently processed and analyzed.

2.3. Soot Calibration Method

The raw signals captured by the LII technique represent incandescence intensities, which only reflect relative SVF values. To obtain the 2D distribution of the absolute SVF, a calibration proportionality factor must be calculated by correlating these relative signals with the absolute SVF acquired via the LE technique. By simultaneously measuring the same soot using both LE and LII, the proportionality factor can be derived by comparing the absolute SVF integrated along the laser path (from LE) with the path-integrated incandescence signal (from LII). This factor is subsequently applied to the target experimental flame.

According to Francqueville [20], the absolute SVF measured by LE is the cumulative value along the laser optical path. By integrating the LII signal along the same extinction path, the expression for the calibration proportionality factor is given by:

$$c = \frac{\int_0^L f_v dx}{\int_0^L I_{LII} dx} = \frac{\lambda_{LE}}{k_e} \frac{KL_{LEM}}{KL_{LII}} \quad (2)$$

where K_{LLEM} is the path-integrated SVF obtained from LE, K_{LII} is the integrated incandescence signal along the optical path, λ_{LE} is the LE laser wavelength (632.8 nm in this study), and k_e is the dimensionless extinction coefficient (set to 9.7 in this study [33]). Specifically, KL_{LEM} represents the path-integrated soot optical thickness derived directly from the measured transmitted laser intensity via the Beer–Lambert law. Correspondingly, KL_{LII} is the spatial line integral of the 2D LII relative intensity field captured by the ICCD camera along the exact same optical chord. By extracting and averaging data from the center chord of a standard Santoro flame during the calibration experiment, the experimental ratio of these two path integrals was determined to be $KL_{LII}/KL_{LEM} = 32.7$. In addition, the threshold setting of the LII intensity field determines the effective concentration boundary, which affects the optical path length. The literature usually uses 1–2% of the peak value as the threshold, and 1.5% is taken in this study to balance the signal-to-noise ratio and boundary integrity. Therefore, the calibration coefficient calculated based on the standard flame is:

$$c = \frac{\lambda_{LE}}{k_e} \frac{KL_{LEM}}{KL_{LII}} = \frac{632.8 \times 10^{-9}}{9.7 \times 32.7} = 2.0 \times 10^{-9} \quad (3)$$

The results are associated and transformed using this coefficient. To evaluate the reliability of this quantitative calibration, an uncertainty analysis was performed for the constant c . According to the error propagation law, the total relative uncertainty primarily originates from three sources: (1) the empirical uncertainty of the dimensionless extinction coefficient k_e ($\sim \pm 10\%$); (2) the measurement noise and minor beam steering effects in

KL_{LEM} ($\sim \pm 2\%$); and (3) the integration errors in KL_{LII} due to detector shot noise and threshold sensitivity ($\sim \pm 5\%$). The combined relative uncertainty of c is estimated to be approximately $\pm 11.3\%$, which falls well within the widely accepted range for quantitative soot optical diagnostics.

2.4. Experimental Conditions

The experiments utilized pure methane (nominal purity 99.99%), pure ethylene (99.99%), pure propylene (99.99%), and pure propane (99.9%) as the base fuels for mixing. The total volumetric flow rate of the fuel mixture was strictly maintained at 2.0 L/min to ensure the stability and consistency of the experimental boundary conditions.

A binary mixing strategy was adopted, pairing the four base fuels to create six distinct fuel combinations. These four base fuels exhibit widely varying sooting propensities: non-sooting methane, lightly sooting propane, moderately sooting ethylene, and heavily sooting propylene. The six combinations comprehensively cover three specific categories: alkane + alkane, alkane + alkene, and alkene + alkene. The specific operating condition parameters for the SVF measurement experiments across these binary buoyant diffusion flames are detailed in Table 1.

Table 1. Summary of Experimental Conditions.

No.	Fuel1	Fuel2	Density (kg/m ³)	Calorific Value (MJ/kg)	Volumetric Flow (L/min)
C1–C11	CH ₄	C ₂ H ₄	0.65~1.15	50.03~47.17	2.0
C12–C22	CH ₄	C ₃ H ₆	0.65~1.72	50.03~45.79	2.0
C23–C33	CH ₄	C ₃ H ₈	0.65~1.80	50.03~46.36	2.0
C34–C44	C ₂ H ₄	C ₃ H ₆	1.15~1.72	47.17~45.79	2.0
C45–C55	C ₂ H ₄	C ₃ H ₈	1.15~1.80	47.17~46.36	2.0

2.5. Data Preprocessing

The data collected by LII is in the form of a 1393×1080 matrix. The soot region is selected, and a 400×400 matrix centered on the flame's axis is chosen. The data collected by the LE method was converted and calculated by LII, and presented in the form of a 400×1 column vector. This column vector represents the data after applying the calibration coefficient. Figure 3 shows a pair of raw data sets, including the column vector curve representing the LE data and the heatmap of the cropped LII data matrix.

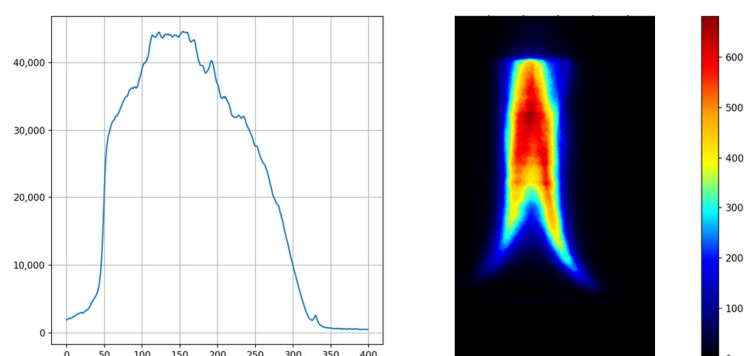


Figure 3. Raw Data Pair: Extinction Data Column Vector Curve (Left) and LII Data Matrix Heatmap (Right). Note: The values displayed on the y-axis (0–40,000) represent the raw relative intensity counts (or scaled intermediate values) for clear visualization. The absolute soot volume fraction is computed subsequently by multiplying these relative values by the calibration coefficient of 2.0×10^{-9} .

Data cleaning is performed to remove outliers caused by laser energy fluctuations or camera noise. In total, 693,600 valid data sets are collected at a frequency of 10 Hz. Although the data is collected under quasi-steady soot conditions, due to the inherent fluctuations of the flame, the data is averaged every 50 sets to ensure reliability and generalizability. The final dataset used for model training consists of 13,872 sets, with the training, validation, and test sets split in an 8:1:1 ratio. This results in 11,098 training sets and 1387 sets each for validation and testing.

All samples in this dataset originate from a single experimental system [19]. This constraint is intrinsic to the 1D-to-2D optical inversion problem: because forward projection physics are strictly coupled to hardware geometry (e.g., optical path length and sensor resolution), directly mixing cross-laboratory datasets without rigorous geometric recalibration is mathematically invalid. To prevent overfitting to this fixed setup, the dataset is meticulously designed to encompass highly diverse combustion regimes. By varying binary hydrocarbon mixtures and flow rates, it yields highly complex and asymmetric soot topologies. While adapting the model to entirely different burner geometries remains a direction for future research, this extensive internal diversity is sufficient to validate the generalizability of the proposed Soot-MixFormer.

Regarding data distribution, the 13,872 samples cover a broad and balanced range of soot volume fractions (SVF), from non-sooting boundaries to dense cores (up to 10 ppm), preventing model bias toward specific flame states. Furthermore, the system's inherent measurement uncertainty was quantitatively evaluated. Despite laser and sensor fluctuations, the optical system maintains a high Signal-to-Noise Ratio (SNR) of approximately 35 dB. Based on error propagation from the Beer–Lambert law, the overall measurement uncertainty for the path-integrated SVF is estimated within $\pm 4.5\%$. This establishes a reliable physical baseline, verifying the model's robustness against real-world experimental noise.

3. Methodology and Network Architecture

To explicitly outline the proposed hybrid Soot-MixFormer framework, the model is built upon a highly customized encoder–decoder architecture. As illustrated in the following subsections, the encoder is designed to extract deep latent representations from the 1D laser extinction signals, while the decoder is responsible for progressively reconstructing the high-fidelity 2D soot spatial distribution. A core innovation within this architecture is the integration of advanced attention mechanisms, specifically the Dual-Axial Gated Refinement (DAGR) module detailed in Section 3.5, which models long-range global dependencies to capture complex soot topologies. Furthermore, the optimization of this hybrid network is guided by a meticulously designed composite loss function—incorporating structural similarity and physical constraints—which is explicitly defined and comprehensively evaluated in Section 4.1.1.

3.1. Mathematical Formulation of the Ill-Posedness and Regularization

From a rigorous mathematical perspective, the 1D-to-2D soot inversion is a severely ill-posed inverse problem. Let $X \in \mathbb{R}^{H \times W}$ denote the target 2D soot distribution and $V \in \mathbb{R}^W$ denote the 1D line-of-sight measurement. The forward projection can be modeled as a linear operator A :

$$V = AX \quad (4)$$

where $A \in \mathbb{R}^{W \times (HW)}$ represents the spatial integration matrix.

According to Hadamard's criteria, this problem violates both uniqueness and stability:

Non-uniqueness: Because the number of unknown variables ($H \times W$) vastly exceeds the number of known measurements (W), the forward operator A possesses a highly non-trivial null space, denoted as $\mathcal{N}(A)$. For any arbitrary vector $X_{null} \in \mathcal{N}(A)$, $A(X +$

$X_{null}) = AX = V$. Consequently, infinitely many 2D spatial distributions can yield the exact same 1D projection, making the unconstrained inverse problem fundamentally non-unique.

Instability: The condition number of A is infinite, meaning small perturbations or measurement noise δV in the 1D signal can be catastrophically amplified, leading to unbounded errors in the reconstructed 2D field without proper constraints.

To mathematically stabilize this ill-posedness, classical approaches rely on Tikhonov regularization:

$$\hat{X} = \operatorname{argmin}_X \|AX - V\|_2^2 + \lambda \mathcal{R}(X) \quad (5)$$

where $\mathcal{R}(X)$ relies on hand-crafted priors (e.g., Total Variation). In our framework, the proposed Soot-MixFormer functions as an advanced, data-driven regularizer. Instead of penalizing the residual, our physical hard constraint mathematically restricts the solution manifold such that the data consistency term is strictly zero ($\|AX - V\|_2^2 = 0$). Simultaneously, the hybrid CNN-Transformer architecture, optimized over the extensive dataset $\mathcal{D}$, acts as a powerful implicit prior $\mathcal{R}_\theta(X)$. It restricts the solution space from the infinite mathematically possible solutions to the highly constrained subspace of physically and thermodynamically plausible flame topologies, effectively restoring uniqueness and stability to the inversion process.

3.2. Overall Architecture Overview

Mathematically, reconstructing a two-dimensional (2D) asymmetric soot spatial distribution directly from a single-view, one-dimensional (1D) laser extinction signal is a severely ill-posed inverse problem. To overcome the “curse of dimensionality” and physical inconsistency inherent in traditional deep learning approaches, this paper proposes a novel, end-to-end hybrid neural network architecture termed Soot-MixFormer (Soot spatial distribution reconstruction via Mixed CNN-Transformer architecture).

Instead of employing a brute-force dense mapping strategy, Soot-MixFormer innovatively integrates the local feature extraction capability of Convolutional Neural Networks (CNNs) with the global dependency modeling of Transformers, further combined with low-rank spatial decoupling and physical conservation laws. As illustrated in Figure 4, the complete forward inference pipeline is highly streamlined and consists of four synergistic modules:

1. **Physical Prior Embedding and Feature Preprocessing:** Compresses raw 1D measurement signals non-linearly and injects spatial position priors to establish a robust mathematical foundation.
2. **Multi-Scale CNN-Transformer Pyramid Backbone:** Employs a U-Net-style encoder-decoder structure. It leverages stride 1D convolutions for efficient local downsampling and Transformer blocks to hierarchically extract global macroscopic semantics, perfectly fusing high-frequency local details with long-range dependencies.
3. **Dynamic Low-Rank Outer Product Generation Head:** Decouples the 2D matrix generation into 1D row and column feature vectors using a cross-attention mechanism, drastically reducing computational complexity to an $O(L)$ level while preserving spatial continuity.
4. **Dual-Axial Gated Refinement and Physical Constraint:** Eliminates structural artifacts via a lightweight, CNN-based dual-axial refiner (using large-kernel asymmetric convolutions) and rigorously enforces mass conservation along the optical path through mathematical hard constraints.

By seamlessly connecting these four modules, Soot-MixFormer achieves a superior balance among computational efficiency, multi-scale structural perception, and rigorous

physical fidelity. The specific internal mechanisms and mathematical formulations of these modules will be comprehensively detailed in the following subsections.

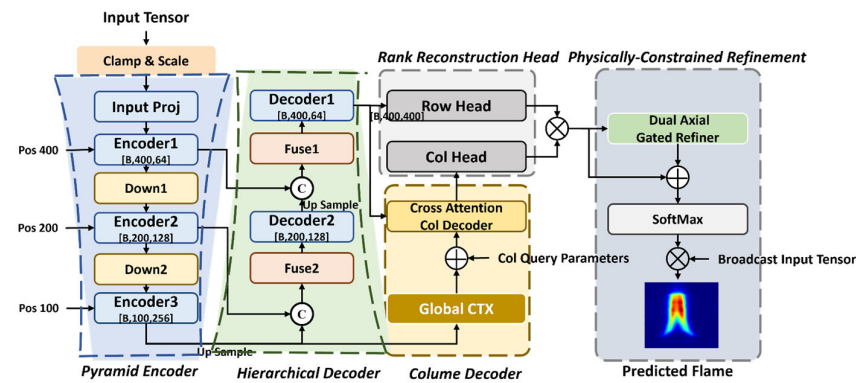


Figure 4. Schematic diagram of the Soot-MixFormer architecture for 2D soot spatial distribution reconstruction.

3.3. Physical Prior Embedding and Feature Preprocessing

Raw 1D extinction signals from optical sensors are inevitably contaminated by background noise and exhibit vast dynamic ranges across different soot zones. To process these signals robustly, Soot-MixFormer employs a physical prior embedding and feature preprocessing module at its frontend.

3.3.1. Physical Clamping and Non-Linear

Let the raw 1D measurement vector be denoted as $\mathbf{x}_{raw} \in \mathbb{R}^L$ (spatial resolution $L = 400$). Because physical path-integrated SVF values are strictly non-negative, a zero-value clamping operation is first applied to eliminate anomalous negative noise caused by baseline fluctuations:

$$\mathbf{x} = \max(\mathbf{x}_{raw}, 0) \quad (6)$$

To accommodate the drastic magnitude variations in the concentration field and prevent computational instability, a logarithmic transformation is then applied to compress the dynamic range:

$$\mathbf{x}_{log} = \ln(1 + \mathbf{x}) \quad (7)$$

The constant 1 avoids singularities and ensures linear approximation near the origin, perfectly preserving weak signals while compressing extremely high-concentration peaks. The resulting feature dimension becomes $\mathbf{x}_{log} \in \mathbb{R}^{L \times 1}$.

3.3.2. High-Dimensional Space Projection

Since standard 1D scalars are incompatible with multi-head self-attention mechanisms, a linear projection layer maps the physical signals into a high-dimensional latent space ($D = 64$). This module sequentially applies a linear transformation, Layer Normalization, and a GELU activation:

$$\mathbf{H}_{proj} = \text{GELU}(\text{LayerNorm}(\mathbf{x}_{log} \mathbf{W}_{in} + \mathbf{b}_{in})) \quad (8)$$

where $\mathbf{W}_{in} \in \mathbb{R}^{1 \times D}$ and $\mathbf{b}_{in} \in \mathbb{R}^D$ represent the weight and bias matrices of the linear projection layer, respectively. Layer Normalization is incorporated to stabilize feature distributions and accelerate model convergence.

3.3.3. Spatial Positional Prior Embedding

Transformers inherently lack spatial awareness due to permutation invariance. However, in tomographic reconstruction, the 1D sequence index strictly corresponds to actual physical coordinates across the soot. To break permutation symmetry and encode these geometric priors, a learnable 1D positional encoding $\mathbf{E}_{pos} \in \mathbb{R}^{L \times D}$ is injected. Unlike rigid sinusoidal encodings, learnable parameters implicitly adapt to soot distribution tendencies (e.g., asymmetry) during backpropagation. The final tensor fed into the backbone is formulated as:

$$\mathbf{H}_{input} = \mathbf{H}_{proj} + \mathbf{E}_{pos} \quad (9)$$

Through these steps, raw 1D signals are cleaned, projected, and spatially encoded, establishing a robust $\mathbf{H}_{input} \in \mathbb{R}^{L \times D}$ foundation for subsequent global feature extraction.

3.4. Multi-Scale Pyramid Transformer Backbone and Fusion

Soot spatial distributions exhibit both macroscopic continuity and severe microscopic gradients. To capture these simultaneously, the backbone adopts a U-Net-style multi-scale pyramid architecture, flawlessly integrating the local inductive bias of Convolutional Neural Networks (CNNs) with the global receptive field of Transformers.

3.4.1. Hierarchical CNN-Transformer Encoder

The encoder progressively expands the receptive field to extract deep global semantics from the preprocessed input $\mathbf{H}_0 \in \mathbb{R}^{L \times D}$ ($L = 400$, $D = 64$). Initially, a multi-head self-attention Transformer block extracts global features at the original resolution, yielding $\mathbf{F}_{400} \in \mathbb{R}^{400 \times 64}$.

To perform spatial downsampling while increasing feature abstraction, a 1D convolution with a stride of 2 ($\text{Conv1D}_{s=2}$) is applied. This captures local gradient correlations between adjacent optical paths. Following the convolution, Layer Normalization (LN) and a GELU activation function (ϕ) are applied. A scale-specific positional encoding ($\mathbf{E}_{pos,200}$) is then added:

$$\mathbf{H}_{200} = \phi(\text{LN}(\text{Conv1D}_{s=2}(\mathbf{F}_{400}))) + \mathbf{E}_{200} \quad (10)$$

$$\mathbf{F}_{200} = \text{TransformerBlock}_{stage2}(\mathbf{H}_{200}) \quad (11)$$

This downsampling process is repeated to reach the deepest bottleneck layer, compressing the spatial resolution to one-fourth while expanding the channels to 4D. The resulting tensor $\mathbf{F}_{100} \in \mathbb{R}^{100 \times 256}$ encapsulates the core macroscopic semantic context of the soot.

3.4.2. Feature Fusion and Progressive Transformer Decoder

To restore spatial resolution and recover high-frequency microscopic details lost during downsampling, the network employs a progressive decoding strategy with skip-connections. Deep features are upsampled via linear interpolation and concatenated with corresponding encoder features along the channel dimension ($[\cdot \parallel \cdot]$). To bridge the semantic gap between deep and shallow layers, a specialized Fuse module (comprising linear projection, Layer Normalization, and GELU) aggregates the concatenated tensor $\mathbf{C}_{200}$:

$$\mathbf{U}_{200} = \text{Interpolate}(\mathbf{F}_{100}, \text{size} = 200, \text{mode} = 'linear') \quad (12)$$

$$\mathbf{C}_{200} = [\mathbf{U}_{200} \parallel \mathbf{F}_{200}] \quad (13)$$

$$\mathbf{M}_{200} = \text{GELU}(\text{LayerNorm}(\mathbf{C}_{200} \mathbf{W}_{fuse2} + \mathbf{b}_{fuse2})) \quad (14)$$

The fused feature $\mathbf{M}_{200}$ is then refined by a decoding Transformer block. This “upsampling-fusion-decoding” cycle is repeated until the feature map returns to the original resolution of $L = 400$. The final backbone output is formulated as:

$$\mathbf{F}_{final} = \text{TransformerBlock}_{stage1}(\text{Fuse}_1([\text{Interpolate}(\mathbf{D}_{200}, \text{size} = 400) \parallel \mathbf{F}_{400}])) \quad (15)$$

This final tensor $\mathbf{F}_{final} \in \mathbb{R}^{400 \times 64}$ perfectly integrates deep macroscopic topology with shallow spatial mutations, serving as a high-fidelity memory bank for the subsequent matrix generation head.

3.5. Dynamic Low-Rank Feature Decoupling and 2D Matrix Generation

Directly mapping a 1D vector to a high-dimensional 2D matrix (e.g., 400×400) using standard dense layers inevitably triggers a “curse of dimensionality,” leading to parameter explosion and the loss of physical spatial correlations. To resolve this, Soot-MixFormer conceptualizes the 2D matrix generation as a low-rank matrix factorization process, decoupling it into the independent construction of 1D row and column feature vectors.

3.5.1. Local-Aware Row Feature Branch

The row features primarily encode structural mutations along the laser optical path. Therefore, the model generates the row weight matrix $\mathbf{W}_{row} \in \mathbb{R}^{L \times d}$ (where $L = 400$ and the latent dimension $d = 128$) by applying a direct linear projection to the high-fidelity output tensor $\mathbf{F}_{final}$ from the backbone:

$$\mathbf{W}_{row} = \mathbf{F}_{final} \mathbf{W}_r + \mathbf{b}_r \quad (16)$$

where $\mathbf{W}_r$ and $\mathbf{b}_r$ are the learnable weights and biases of the row projection layer.

3.5.2. Global-Driven Dynamic Column Query Branch

Unlike row features, generating orthogonal column features requires an awareness of the macroscopic soot morphology to ensure global structural consistency. To achieve this, an instance-aware dynamic query mechanism is introduced. First, Global Average Pooling (GAP) is applied to the deepest bottleneck feature $\mathbf{F}_{100}$ to extract a highly condensed macroscopic context vector $\mathbf{g}$. This global context is then broadcast and added to a learnable, static column prior matrix $\mathbf{P}_{col} \in \mathbb{R}^{L \times d}$ to formulate a dynamic target query $\mathbf{Q}_{dyn}$:

$$\mathbf{Q}_{dyn} = \mathbf{P}_{col} + \text{Broadcast}(\text{GAP}(\mathbf{F}_{100})) \quad (17)$$

Subsequently, a Cross-Attention mechanism is employed. The dynamic query $\mathbf{Q}_{dyn}$ serves as the Query (Q), while the detail-rich $\mathbf{F}_{final}$ acts as both the Key (K) and Value (V). This allows the model to precisely retrieve and fuse fine-grained local features under the guidance of macroscopic global intentions, yielding the column weight matrix $\mathbf{W}_{col} \in \mathbb{R}^{L \times d}$:

$$\mathbf{W}_{col} = \text{Softmax}\left(\frac{\mathbf{Q}_{dyn} \mathbf{F}_{final}^T}{\sqrt{d}}\right) \mathbf{F}_{final} \quad (18)$$

3.5.3. Low-Rank Outer Product Space Construction

With the decoupled row and column features established, the base 2D spatial feature field is constructed through a bilinear inner product (implemented via the Einstein summa-

tion convention, or Einsum). The base prediction logits $\mathbf{M}_{base} \in \mathbb{R}^{L \times L}$ are calculated as the outer product of the two matrices, scaled by the square root of the latent dimension:

$$\mathbf{M}_{base} = \frac{\mathbf{W}_{row} \mathbf{W}_{col}^T}{\sqrt{d}} \quad (19)$$

This low-rank construction fundamentally circumvents the parameter explosion of traditional MLPs, drastically reducing the computational complexity to an $O(L \cdot d)$ level. More importantly, it mathematically aligns with the inherent spatial correlations and continuous physical distributions of fluid concentration fields.

3.6. Dual-Axial Gated Refinement and Physically Constrained Layer

While the low-rank outer product generation head efficiently constructs the base 2D spatial feature field $\mathbf{M}_{base} \in \mathbb{R}^{L \times L}$, the inherent mathematical nature of rank-deficient bilinear operations inevitably introduces orthogonal grid-like artifacts (checkerboard patterns). Furthermore, purely data-driven predictions cannot naturally guarantee compliance with the underlying physical laws of laser extinction. To address these two critical issues, a refinement module and a physical hard constraint layer are sequentially deployed at the terminal stage of Soot-MixFormer.

3.6.1. Dual-Axial Gated Refinement (DAGR) Module

While the low-rank outer product generation efficiently constructs the base 2D spatial feature field $\mathbf{M}_{base} \in \mathbb{R}^{1 \times L \times L}$, the inherent mathematical nature of rank-deficient bilinear operations inevitably introduces orthogonal grid-like checkerboard artifacts. To eliminate these structural anomalies and restore high-frequency boundary details without incurring the massive computational overhead of standard large-kernel 2D convolutions ($O(K^2)$), a remarkably lightweight Dual-Axial Gated Refinement (DAGR) module is proposed.

As illustrated in Figure 5, the DAGR module first projects the single-channel base matrix into a higher-dimensional latent space $\mathbf{F}_{in} \in \mathbb{R}^{C \times L \times L}$ (where $C = 32$) via a 1×1 convolutional projection layer. The feature map is then bifurcated into two parallel branches to independently perceive contextual information along the horizontal and vertical axes. Depthwise asymmetric large-kernel convolutions ($1 \times K$ and $K \times 1$, with $K = 11$) are applied to extract direction-aware spatial structures, reducing the computational complexity to $O(2K)$:

$$\mathbf{F}_{row} = \text{Conv}_{1 \times K}(\mathbf{F}_{in}) \quad (20)$$

$$\mathbf{F}_{col} = \text{Conv}_{K \times 1}(\mathbf{F}_{in}) \quad (21)$$

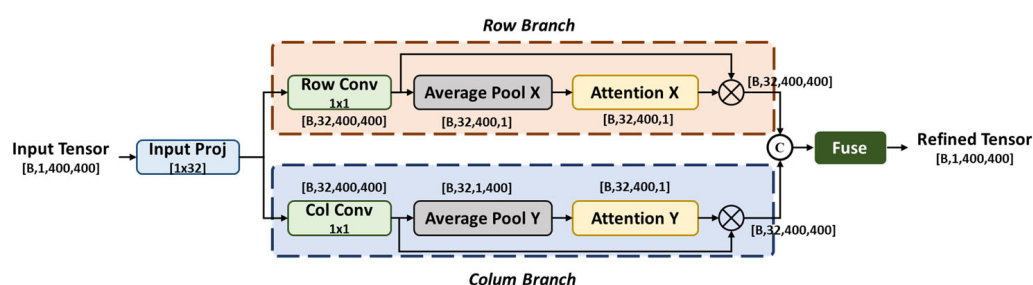


Figure 5. Illustration of Dual-Axial Gated Refinement Module.

To dynamically suppress grid noise and enhance true physical gradients, a Coordinate Attention Gating mechanism incorporating strip pooling is introduced. By performing 1D average pooling along the complementary spatial dimensions—compressing the width (X-axis) to 1 for the row branch and the height (Y-axis) to 1 for the column branch—the

module captures global axial statistical distributions. These compressed vectors are passed through a lightweight attention network (comprising 1×1 convolutions and a Sigmoid activation σ) to generate coordinate-aware attention maps. The axial features are then element-wise modulated ($\otimes$) by these maps:

$$\hat{\mathbf{F}}_{row} = \mathbf{F}_{row} \otimes \sigma(\text{Attn}_X(\text{AvgPool}_X(\mathbf{F}_{row}))) \quad (22)$$

$$\hat{\mathbf{F}}_{col} = \mathbf{F}_{col} \otimes \sigma(\text{Attn}_Y(\text{AvgPool}_Y(\mathbf{F}_{col}))) \quad (23)$$

Finally, the cross-gated features from both branches are concatenated ($[\cdot \parallel \cdot]$) along the channel dimension. A fusion module sequentially applies a 1×1 convolution, Group Normalization, GELU, and a final 3×3 convolution to aggregate the dual-axial information, map it back to a single channel, and restore local spatial smoothness, yielding the residual tensor $\Delta\mathbf{M}$. Crucially, the final convolution weights and biases are initialized to zero. This ensures a strict identity mapping during the initial training phase, thereby preserving the structural integrity of the Transformer's original macroscopic predictions. The ultimate refined 2D logits $\mathbf{M}_{refined}$ are formulated via a residual connection:

$$\mathbf{M}_{refined} = \mathbf{M}_{base} + \Delta\mathbf{M} \quad (24)$$

3.6.2. Physical Mass Conservation Constraint

A common limitation of purely data-driven deep learning models in combustion diagnostics is their tendency to violate fundamental physical laws, leading to physically implausible predictions. Since soot distribution inversion is inherently a physics-constrained problem governed by the Beer–Lambert law, incorporating physical priors into the network is imperative. In the context of single-view laser extinction, the most fundamental physical prior is the mass conservation along the optical path: the line-of-sight integration of the 2D Soot Volume Fraction (SVF) distribution must strictly equal the 1D path-integrated measurement.

Instead of adding a penalty term to the loss function (a “soft constraint” which is prone to optimization instability and weighting trade-offs), Soot-MixFormer enforces this physical law as an absolute “hard constraint” directly within the forward propagation architecture. First, a spatial Softmax function is applied to the refined logits $\mathbf{M}_{refined}$ along the projection axis (the optical path dimension). This operation transforms the unnormalized logits into a relative spatial distribution matrix $\mathbf{S} \in \mathbb{R}^{L \times L}$, where the sum of probabilities along each specific optical path i equals 1:

$$\mathbf{S}(i, j) = \frac{\exp(\mathbf{M}_{refined}(i, j))}{\sum_{k=1}^L \exp(\mathbf{M}_{refined}(i, k))} \quad (25)$$

Subsequently, the strictly non-negative clamped physical input vector $\mathbf{x} \in \mathbb{R}^L$ is broadcast and multiplied element-wise with the relative distribution matrix $\mathbf{S}$. The final predicted 2D SVF matrix $\mathbf{M}_{final}$ is formulated as:

$$\mathbf{M}_{final}(i, j) = \mathbf{S}(i, j) \cdot \mathbf{x}(i) \quad (26)$$

By mathematically enforcing $\sum_j \mathbf{M}_{final}(i, j) \equiv \mathbf{x}(i)$, this broadcast multiplication guarantees zero physical residual between the predicted 2D field's projection and the actual 1D measurement. This pivotal step successfully transforms Soot-MixFormer from a conventional black-box mapping model into a highly interpretable, physics-guided diagnostic

paradigm, ensuring that every predicted distribution is strictly grounded in the measured physical reality.

However, it is crucial to note that the enforcement of this strict physical constraint introduces inherent trade-offs regarding noise sensitivity. Because the final prediction is directly formulated by multiplying the relative distribution matrix with the 1D input measurement, the architecture acts as a double-edged sword. On the one hand, it guarantees physical fidelity; on the other hand, it deterministically propagates any severe measurement noise, sensor calibration errors, or baseline drift present in the 1D input into the corresponding optical paths of the reconstructed 2D field. Under extreme noise conditions, this constraint could force the network to allocate “ghost” soot concentrations to maintain mass conservation, potentially biasing the local reconstruction.

Nevertheless, as empirically demonstrated in the noise robustness analysis (Section 4.6), the deep feature extraction layers within the CNN-Transformer backbone effectively act as a spatial low-pass filter. This filtering capability prevents catastrophic failure and limits error propagation to moderate localized artifacts, ensuring stable degradation even under high noise levels.

4. Experimental Results

4.1. Loss Function and Evaluation Metrics

4.1.1. Loss Function

To guarantee absolute numerical fitting accuracy while effectively constraining the spatial structural distribution of the 2D concentration matrix, this study designs a composite loss function. This function is formulated as a weighted combination of a Log-space Smooth L1 Loss and a Row-wise Kullback–Leibler (KL) Divergence loss.

For numerical regression, the model primarily employs a Smooth L1 Loss in the logarithmic domain. By applying a logarithmic transformation to both the predicted matrix ($\mathbf{M}_{pred}$) and the ground-truth matrix ($\mathbf{M}_{gt}$), the exceptionally wide dynamic range of the concentration field data is effectively compressed, mitigating the gradient interference caused by extreme values in the long-tail distribution. The reconstruction loss L_{rec} is formulated as:

$$L_{rec} = \text{Smooth}_{L1}(\ln(1 + \mathbf{M}_{pred}), \ln(1 + \mathbf{M}_{gt})) \quad (27)$$

Compared to the traditional Mean Squared Error (MSE), the Smooth L1 Loss exhibits linear behavior far from the origin, granting stronger robustness against outliers. Simultaneously, it behaves quadratically near the origin, ensuring gradient smoothness during the optimization process.

To explicitly constrain the morphological features of the base matrix, a Row-wise KL Divergence loss is introduced. This term calculates the relative entropy between the predicted normalized distribution ($\mathbf{Q}$) processed via Softmax and the row-wise target distribution ($\mathbf{P}$) derived from the ground-truth matrix. For a matrix with H rows, the distribution loss L_{dist} is calculated as:

$$L_{dist} = \frac{1}{H} \sum_{i=1}^H D_{KL}(\mathbf{P}_i \parallel \mathbf{Q}_i) = \frac{1}{H} \sum_{i=1}^H \sum_j \mathbf{P}_{i,j} \log \left(\frac{\mathbf{P}_{i,j}}{\mathbf{Q}_{i,j}} \right) \quad (28)$$

The incorporation of this constraint compels the network to deeply learn the relative structural relationships within the target flow field, successfully overcoming the structural distortion issues often induced by purely pixel-wise numerical approximation.

The ultimate total loss function (L_{total}) is defined as the weighted sum of these two components:

$$L_{total} = \lambda_1 L_{rec} + \lambda_2 L_{dist} \quad (29)$$

where λ_1 and λ_2 are the corresponding weight coefficients used to dynamically balance the dual optimization objectives of absolute numerical regression and probability distribution fitting throughout the training process.

4.1.2. Evaluation Metrics

To comprehensively assess the reconstruction fidelity of the 2D soot spatial distribution from both morphological and numerical perspectives, six standard quantitative evaluation metrics are employed. Let $\mathbf{M}_{gt}$ and $\mathbf{M}_{pred}$ denote the flattened 1D vectors of the ground-truth matrix and the predicted matrix, respectively, both with a total of N elements (where $N = H \times W$). y_i and $\hat{y}_i$ represent their respective i -th pixel values. These metrics are broadly categorized into structural similarity indicators and pixel-level numerical error measurements.

(1) Structural and Distribution Similarity Metrics

Structural Similarity Index Measure (SSIM): SSIM evaluates the perceptual and structural similarity between the predicted matrix and the ground truth by comprehensively comparing their luminance, contrast, and spatial structures.

$$\text{SSIM}(\mathbf{M}_{pred}, \mathbf{M}_{gt}) = \frac{(2\mu_{pred}\mu_{gt} + c_1)(2\sigma_{pred,gt} + c_2)}{(\mu_{pred}^2 + \mu_{gt}^2 + c_1)(\sigma_{pred}^2 + \sigma_{gt}^2 + c_2)} \quad (30)$$

where μ and σ^2 denote the mean and variance of the respective matrices, $\sigma_{pred,gt}$ is the covariance, and c_1, c_2 are constants to stabilize the division. In the context of fluid concentration fields, a higher SSIM (closer to 1) indicates superior preservation of the macroscopic physical soot topology and effective suppression of grid artifacts.

Pearson Correlation Coefficient (PCC): PCC quantifies the linear correlation between the predicted and actual concentration fields. It rigorously assesses whether the global distribution trend and the relative gradients of the reconstructed soot match the ground truth.

$$\text{PCC} = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \quad (31)$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the mean values of $\mathbf{M}_{gt}$ and $\mathbf{M}_{pred}$, respectively.

Cosine Similarity: This metric computes the cosine of the angle between the two high-dimensional feature vectors. It evaluates the directional alignment and proportional consistency of the spatial distributions, independent of the absolute magnitude.

$$\text{Cosine Similarity} = \frac{\mathbf{M}_{pred} \cdot \mathbf{M}_{gt}}{\|\mathbf{M}_{pred}\|_2 \|\mathbf{M}_{gt}\|_2} = \frac{\sum_{i=1}^N \hat{y}_i y_i}{\sqrt{\sum_{i=1}^N \hat{y}_i^2} \sqrt{\sum_{i=1}^N y_i^2}} \quad (32)$$

(2) Pixel-Level Numerical Error Metrics

Mean Squared Error (MSE) and Root Mean Square Error (RMSE): MSE calculates the average squared differences between the predicted and actual Soot Volume Fraction (SVF) values across all spatial grids, while RMSE is its square root.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (33)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (34)$$

Because the squared penalty heavily magnifies large discrepancies, RMSE is particularly sensitive to prediction deviations in the high-concentration reaction zones of the flame.

Mean Absolute Error (MAE): MAE computes the average absolute pixel-wise difference. Compared to RMSE, it is less sensitive to extreme outliers and provides a robust, linear assessment of the overall absolute numerical reconstruction accuracy.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (35)$$

By jointly analyzing these metrics, the structural integrity (SSIM $\uparrow$, PCC $\uparrow$, Cosine Sim $\uparrow$) and the absolute numerical precision (MSE $\downarrow$, RMSE $\downarrow$, MAE $\downarrow$) of the proposed Soot-MixFormer can be objectively quantified.

While the mathematical definitions of these metrics are standard, their physical implications are highly specific in the context of combustion diagnostics. Structural metrics (SSIM, PCC, and Cosine Similarity) directly reflect the model's capability to capture the macroscopic flame topology, such as the continuity of the reaction zones and the precise localization of high-concentration soot pockets. High structural fidelity is critical for identifying flame asymmetry and burner maldistribution. Conversely, absolute numerical error metrics (MSE, RMSE, and MAE) quantify the exact deviation in the Soot Volume Fraction (SVF). Because SVF directly governs thermal radiation, these numerical deviations strictly dictate the accuracy of subsequent radiative heat transfer and pollutant emission calculations.

Regarding practical industrial benchmarks, traditional optical tomographic algorithms (e.g., Abel inversion or algebraic reconstruction techniques) applied in turbulent or noisy environments frequently exhibit relative errors ranging from 15% to 30%, often failing to preserve structural integrity. Therefore, in practical industrial monitoring scenarios—such as utility boilers or gas turbine combustors—an SSIM exceeding 0.70, combined with an RMSE that translates to an overall relative error below 10%, is generally considered a highly reliable threshold. Meeting these benchmarks ensures that the reconstructed 2D distributions are sufficiently accurate for real-time, actionable closed-loop control and early anomaly detection.

4.2. Baseline Models

To comprehensively evaluate the reconstruction performance of the proposed architecture, four representative baseline models were implemented for comparison. These models encompass pure Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and hybrid architectures, ensuring a rigorous validation of our structural design. The schematic diagrams of these model architectures are shown in Figure 6.

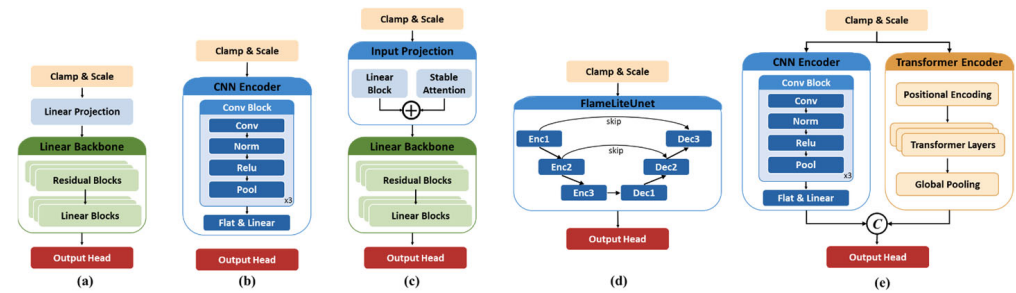


Figure 6. Schematic diagrams of the baseline model architectures implemented for performance comparison. The figure illustrates the detailed structural designs and feature mapping strategies of four comparative models: (a) MLP, a pure MLP-based network; (b) CNN, a purely 1D-CNN-based architecture; (c) EnhancedMLP, an attention-augmented MLP featuring asymmetric generation; (d) FlameLiteUnet, a 1D Lite-UNet backbone featuring symmetric encoder–decoder stages with skip connections to extract multi-scale features from 1D laser attenuation data; and (e) CNNTransformer, a hybrid architecture combining local convolutional inductive bias with global Transformer sequence modeling.

SmoothedKNN (Traditional Baseline): A data-driven, non-deep-learning algorithmic baseline implemented to evaluate traditional reconstruction capabilities. It first establishes an initial 2D matrix estimation by retrieving the k -nearest neighbors from the training dataset based on the input 1D attenuation vector. To mitigate noise and structural discontinuities in the retrieved templates, the initial prediction is iteratively refined using 2D Laplacian smoothing. Crucially, during each smoothing iteration, the matrix is strictly projected to enforce non-negativity and to ensure that the row sums perfectly reconstruct the physical mass conservation constraints dictated by the original 1D input sequence.

MLP Baseline: A pure MLP architecture designed to capture global correlations within the input vector. It consists of stacked linear layers accompanied by ReLU activations and Layer Normalization to mitigate gradient vanishing. The network constructs the 2D matrix via a low-rank decomposition strategy ($\text{rank} = 64$), utilizing parallel linear projections to generate row and column feature matrices.

CNN Baseline: A 1D-CNN architecture aimed at extracting local correlation patterns through constrained receptive fields. The encoder features four hierarchical convolutional blocks (Conv1D, BatchNorm, ReLU, Max Pooling) to sequentially downsample the input sequence and extract translation-invariant high-order features. The flattened features are then mapped to the output space using the same low-rank decomposition head as the MLP baseline.

EnhancedMLP: An advanced MLP variant augmented with statistical feature injection and a StableAttention module. By concatenating explicitly computed statistical properties (e.g., mean and variance) and leveraging multi-head self-attention, the model dynamically captures positional dependencies. Notably, it introduces an asymmetric generation strategy: row features are directly encoded, whereas column features are derived by fusing a learnable column query with global context features.

FlameLiteUnet: A robust deep learning architecture built upon a 1D Lite-U-Net backbone, designed to capture both fine-grained local details and broad contextual features. It features a symmetric encoder–decoder structure: the encoder sequentially downsamples the 1D input utilizing double convolutional blocks and max pooling to extract hierarchical representations, while the decoder upsamples these features and integrates them with corresponding encoder outputs via skip connections. This multi-scale feature fusion effectively mitigates spatial information loss during downsampling. The aggregated sequence features are then flattened and processed through a low-rank decomposition head.

(rank = 64) to independently project row and column matrices, ultimately reconstructing the 2D distribution field under strict input-scaled constraints.

CNNTransformer (Hybrid Baseline): A hybrid network integrating the local inductive bias of CNNs with the global sequence modeling capabilities of Transformers. It first employs a U-Net style convolutional backbone to perform multi-scale downsampling and upsampling ($400 \rightarrow 100 \rightarrow 400$). The extracted features are subsequently processed by a Transformer encoder equipped with positional encoding to capture long-range dependencies, maintaining a learnable query mechanism for structural column generation.

4.3. Implementation Details

For network optimization, the AdamW optimizer was universally adopted for parameter updating across all models. By integrating an adaptive momentum mechanism with explicitly decoupled weight decay, this optimizer effectively mitigates overfitting while accelerating model convergence. Specifically, the initial learning rate was set to 1×10^{-4} , the weight decay coefficient was specified as 1×10^{-4} , and the batch size was uniformly fixed at 32. To dynamically adjust the optimization process and prevent the network from trapping in local optima, a ReduceLROnPlateau learning rate scheduler was introduced. Whenever the validation loss plateaued for 2 consecutive epochs without a significant reduction, the learning rate was automatically decayed by a factor of 0.5. Furthermore, an early stopping mechanism was implemented to prevent unnecessary computational expenditure and to secure the model weights with the optimal generalization capability. Training was automatically terminated if the improvement in validation loss remained below a predefined minimal threshold for 5 consecutive epochs, with the maximum number of training epochs strictly capped at 200. Finally, to guarantee the absolute reproducibility of the experimental results, the global random seed was fixed at 42 across all hardware and software configurations.

4.4. Quantitative Performance Evaluation

4.4.1. Quantitative Metric Comparison

To comprehensively evaluate the reconstruction performance of the proposed Soot-MixFormer, its predictive accuracy and structural fidelity are compared against six baseline models on the test set. The quantitative results are grouped into structural similarity metrics, pixel-level error metrics, and localized/frequency-domain metrics as detailed in Tables 2–4 respectively.

Table 2. Quantitative comparison of structural similarity and distribution correlation metrics among different models.

Model	SSIM	PCC	Cosine Similarity
Smoothed KNN	0.741 ± 0.084	0.877 ± 0.153	0.933 ± 0.072
MLP	0.696 ± 0.085	0.688 ± 0.146	0.803 ± 0.101
CNN	0.760 ± 0.099	0.880 ± 0.102	0.920 ± 0.050
Enhanced MLP	0.757 ± 0.098	0.872 ± 0.103	0.925 ± 0.052
CNNTransformer	0.755 ± 0.101	0.878 ± 0.102	0.919 ± 0.051
FlameLiteUnet	0.787 ± 0.097	0.912 ± 0.129	0.950 ± 0.059
Soot-MixFormer	0.800 ± 0.078	0.915 ± 0.127	0.957 ± 0.043

Table 3. Quantitative comparison of pixel-level numerical reconstruction errors among different models.

Model	RMSE	MAE	MSE
SmoothedKNN	35.20 ± 35.62	17.79 ± 16.75	2507.06 ± 5886.22
MLP	78.624 ± 78.477	45.685 ± 11.452	12,335.99 ± 21,096.08
CNN	28.966 ± 26.235	14.334 ± 5.308	1526.828 ± 2836.419
Enhanced MLP	33.985 ± 35.993	16.879 ± 5.759	2449.535 ± 6288.259
CNNTransformer	29.577 ± 26.545	14.556 ± 5.392	1578.915 ± 2823.721
FlameLiteUnet	30.342 ± 28.751	15.062 ± 13.033	1746.481 ± 3368.997
Soot-MixFormer	27.666 ± 21.111	14.030 ± 6.474	1210.676 ± 1790.497

Table 4. Quantitative comparison of localized regional errors and frequency-domain correlation.

Model	Core MAE	Background MAE	Spectrum PCC
SmoothedKNN	66.44 ± 63.92	3.82 ± 3.71	0.68 ± 0.13
MLP	118.89 ± 131.69	23.83 ± 24.71	0.51 ± 0.16
EnhancedMLP	55.22 ± 65.98	3.57 ± 3.46	0.53 ± 0.18
CNN	45.84 ± 46.04	2.90 ± 2.47	0.64 ± 0.19
CNNTransformer	46.91 ± 47.04	2.92 ± 2.45	0.62 ± 0.14
FlameLiteUnet	47.53 ± 49.47	3.09 ± 2.65	0.61 ± 0.09
Soot-MixFormer	41.31 ± 38.87	2.50 ± 2.13	0.78 ± 0.09

As indicated in Table 2, the purely linear MLP baseline exhibits the poorest performance across all structural metrics (e.g., SSIM of 0.696), revealing its inherent inability to capture complex spatial topologies from 1D inputs. The traditional SmoothedKNN algorithm achieves an SSIM of 0.741, which surpasses the basic MLP but falls short of advanced deep learning approaches due to its limited non-linear mapping capacity. The CNN, Enhanced MLP, CNNTransformer and FlameLiteUnet models demonstrate comparable structural reconstruction capabilities, achieving SSIM scores of approximately 0.75 to 0.78 and Cosine Similarities around 0.92. In contrast, the proposed Soot-MixFormer significantly outperforms all baseline architectures, achieving the highest SSIM (0.800), PCC (0.915), and Cosine Similarity (0.957). Furthermore, the lowest standard deviations in SSIM (0.078) and Cosine Similarity (0.043) demonstrate superior stability and robustness in preserving the macroscopic physical topology and vector directionality of the soot flow field. It is worth noting that while its standard deviation for PCC (0.127) is slightly higher than that of the heavily smoothed CNN-based models (e.g., 0.102), it remains lower than those of the traditional algorithms and FlameLiteUnet. This slight variance in PCC is a typical trade-off, as Soot-MixFormer generates much sharper, high-fidelity localized soot textures rather than the overly smoothed, blurry outputs typical of basic CNNs, thereby sacrificing a marginal amount of linear correlation stability for significantly higher structural accuracy.

Table 3 presents the absolute numerical errors of the reconstructed 2D concentration matrices. The MLP model suffers from severe numerical deviations and extreme variance (MSE exceeding 12,000), indicating severe overfitting and failure in generalizing to unseen spatial distributions. Similarly, although the traditional SmoothedKNN avoids severe neural network overfitting, it still yields a high MSE of 2507.06 with significant variance, reflecting its instability across diverse flame morphologies. Models incorporating convolutional inductive biases (CNN and CNNTransformer) significantly reduce the reconstruction errors, validating the necessity of local receptive fields in extracting smooth concentration gradients. Interestingly, while FlameLiteUnet shows strong structural similarity, its numerical errors (RMSE of 30.342, MSE of 1746.481) are slightly higher than those of standard CNNs. This implies that heavily parameterized generic architectures might still suffer from

minor overfitting or physical inconsistency when processing low-dimensional 1D inputs without explicit physical constraints.

Remarkably, Soot-MixFormer achieves the lowest error margins across all criteria, recording an RMSE of 27.666, an MAE of 14.030, and an MSE of 1210.676. Notably, the variance of the MSE in Soot-MixFormer (± 1790.497) is drastically smaller than that of the CNN model (± 2836.419) and the Enhanced MLP (± 6288.259). This substantial reduction in both absolute error and variance empirically validates the effectiveness of the Dual-Axial Gated Refinement (DAGR) module and the physical hard constraint layer. By strictly enforcing mass conservation along the optical path, the network effectively eliminates severe local outliers, thereby ensuring highly accurate and physically consistent pixel-level numerical reconstructions.

To further dissect the reconstruction accuracy across different topological zones and frequency domains, Table 4 introduces localized boundary metrics (Core MAE and Background MAE) and the Spectrum Pearson Correlation Coefficient (Spectrum PCC). The high-concentration core region of the soot field inherently exhibits steeper gradients and complex topologies, making it notoriously difficult to reconstruct. Consequently, models lacking spatial awareness, such as MLP and SmoothedKNN, show substantial Core_MAEs of 118.89 and 66.44, respectively. Even the advanced FlameLiteUnet struggles slightly in the core zone (Core MAE of 47.53). Conversely, Soot-MixFormer attains the lowest Core MAE (41.31) and Background MAE (2.50), demonstrating highly precise pixel-level regression in both dense flame cores and sparse background boundaries.

Furthermore, in the frequency domain, Soot-MixFormer achieves an outstanding Spectrum PCC of 0.78, vastly outperforming all other models. This critical metric indicates that the proposed network not only reconstructs accurate spatial intensities but also faithfully recovers the high-frequency structural textures and low-frequency global components of the soot distributions, preventing the over-smoothed artifacts typical of conventional methods.

4.4.2. Statistical Distribution Analysis

To intuitively dissect the prediction stability and morphological fidelity of each model when confronted with diverse, unseen samples in the test set, a detailed statistical distribution analysis was conducted. Relying solely on mean metrics can obscure localized failures or extreme outliers; thus, visualizing the distribution of error and similarity metrics provides a more comprehensive evaluation of model robustness.

Figure 7 illustrates the probability density distribution of the Structural Similarity (SSIM) index for the seven evaluated models using violin plots. The purely linear MLP exhibits a heavily bottom-heavy distribution with the lowest median, indicating widespread failure in preserving spatial topologies. Similarly, the traditional SmoothedKNN algorithm displays a relatively wide distribution with a lower median than the advanced deep learning models, highlighting its limited capability in capturing complex spatial variations. The CNN, Enhanced MLP, and CNNTransformer show relatively similar, moderately dispersed distributions centered around the 0.75 mark, with FlameLiteUnet showing a slight structural advantage over the basic CNNs.

In contrast, the SSIM distribution for Soot-MixFormer is distinctly shifted toward the upper echelon (closer to 1.0). The prominent “bulge” at the top and the elevated quartiles (dashed lines) statistically demonstrate that the proposed model consistently achieves high structural fidelity across the vast majority of test samples, minimizing instances of severe structural distortion.

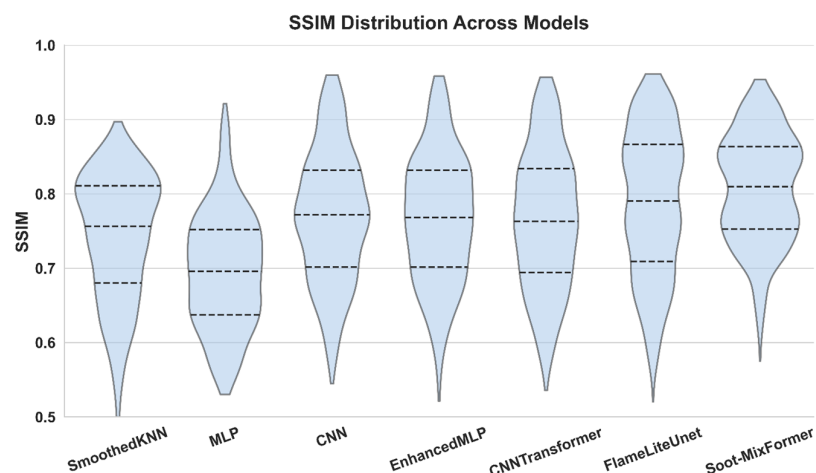


Figure 7. Violin plots illustrating the probability density distribution of the Structural Similarity (SSIM) index on the test set.

Figure 8 presents a boxplot analysis of the Root Mean Square Error (RMSE), highlighting the median, interquartile range (IQR), and the spread of outliers across the models.

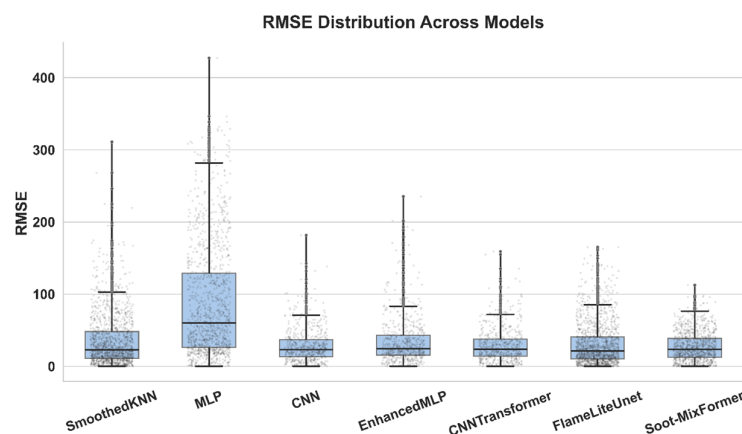


Figure 8. Boxplot analysis of the Root Mean Square Error (RMSE) distributions across the evaluated models.

The MLP model suffers from extreme variance and an excessive number of severe outliers, demonstrating its unreliability. The SmoothedKNN model, while possessing a lower median error than MLP, still exhibits a substantial spread of extreme outliers extending beyond an RMSE of 300, indicating instability across diverse samples. While models with convolutional or attention mechanisms (CNN, Enhanced MLP, CNNTransformer and FlameLiteUnet) significantly compress the overall error scale, they still exhibit pronounced “long tails” (scattered points extending upwards), indicating occasional vulnerability to complex or asymmetric soot structures.

Soot-MixFormer achieves the lowest median RMSE and the most compact IQR among all competitors. More importantly, the upper boundary of its whiskers and the density of extreme outliers are drastically suppressed. This restricted variance is a direct empirical validation of the physical hard constraint layer, which effectively bounds the prediction matrix and prevents physically impossible pixel-level anomalies.

Figure 9 displays the Mean Absolute Error (MAE) distribution using histograms overlaid with Kernel Density Estimation (KDE) curves. A highly performant model should ideally exhibit a steep, narrow distribution concentrated near zero. The MLP’s distribution is flat and heavily dispersed, confirming widespread prediction inaccuracies. The

SmoothedKNN distribution, though an improvement over MLP, is visibly wider than those of the neural network baselines. The other baseline models show improved left-skewness but still possess noticeable rightward tails.

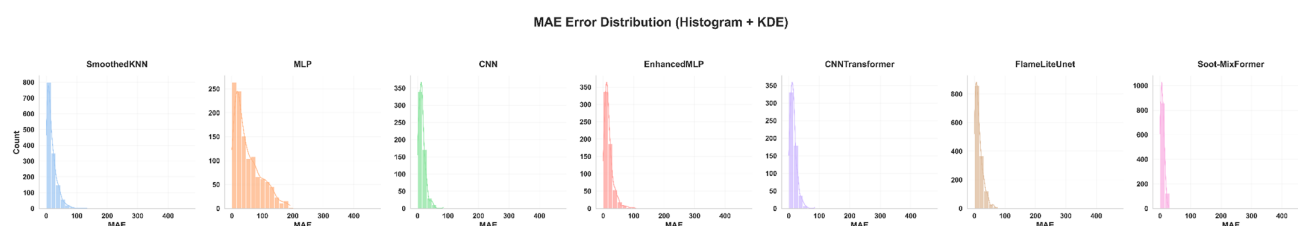


Figure 9. Histograms overlaid with Kernel Density Estimation (KDE) curves depicting the distribution of Mean Absolute Error (MAE).

The MAE distribution for Soot-MixFormer (the rightmost subplot) is exceptionally sharp, highly leptokurtic, and tightly concentrated in the extremely low-error region (near 0–25). The rapid decay of its KDE curve indicates that instances of moderate-to-high errors are exceedingly rare. This confirms that Soot-MixFormer not only performs well on average but does so with remarkable consistency across virtually the entire test dataset.

4.4.3. Model Complexity and Inference Time

To comprehensively evaluate the deployment feasibility of the models in real-world industrial applications, a quantitative comparative analysis was conducted on the number of parameters (Params), floating-point operations (FLOPs), and the processing throughput measured in Frames Per Second (FPS). The detailed computational metrics for all evaluated architectures are summarized in Table 5.

Table 5. Comparison of computational complexity and inference efficiency across evaluated models.

Model Name	Params (M)	FLOPs (G)	FPS
SmoothedKNN	–	–	15
MLP	13.3942	0.0133	1195
CNN	33.0364	0.0474	628
Enhanced MLP	2.091	0.7703	193
CNNTransformer	0.2396	0.0925	317
FlameLiteUnet	27.9475	0.1321	201
Soot-MixFormer	0.9733	0.3493	246

As demonstrated in Table 5, while the traditional purely linear MLP and CNN baseline models require extremely low computational operations (FLOPs < 0.05 G), this comes at the severe cost of heavily bloated parameter counts, reaching 13.39 M and 33.04 M, respectively. Such massive parameter scales not only exacerbate memory consumption on edge devices but also fail to achieve high-precision spatial reconstruction, as proven in previous quantitative experiments. In stark contrast, the proposed Soot-MixFormer exhibits exceptional parameter efficiency. By strictly restraining its total parameter count to merely 0.97 M, which is significantly lower than the 27.94 M of the FlameLiteUnet, it qualifies as an ultra-lightweight network, significantly alleviating storage constraints for practical hardware deployment.

Regarding inference time and computational throughput, although the integration of sequence modeling and physical constraint mechanisms slightly increases the FLOPs (0.349 G) of Soot-MixFormer compared to naive baselines, its overall computational cost remains substantially lower than that of the Enhanced MLP (0.770 G). Most crucially, Soot-MixFormer achieves a remarkable inference throughput of approximately 246 FPS

(equivalent to an ultra-low inference time of ~ 4.06 ms per frame). This is drastically faster than the traditional SmoothedKNN baseline (15 FPS) and highly competitive with classical CNN architectures. This extraordinary processing speed far exceeds the typical sampling frequencies of standard optical sensors (which generally operate at 30 to 60 FPS). Consequently, the proposed architecture perfectly satisfies the rigorous demands for the continuous, real-time, and high-fidelity dynamic monitoring of 2D concentration distributions in actual industrial combustion environments.

4.5. Qualitative Visual Comparison

To qualitatively validate the structural fidelity and local error distribution of the models, a visual comparison of the predicted 2D concentration fields and their corresponding pixel-wise absolute error maps is presented in Figure 10. Three representative test samples featuring diverse and complex soot topologies were selected for this demonstration.

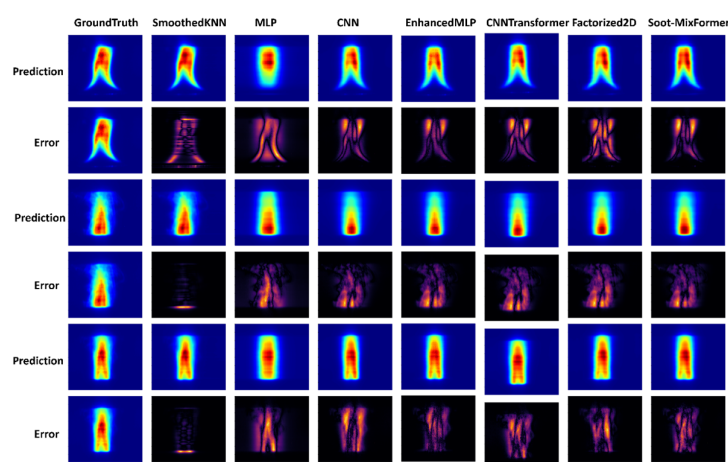


Figure 10. Qualitative visual comparison of the reconstructed 2D soot concentration fields and their absolute error distributions.

As is visually evident, the purely linear MLP model (second column) completely fails to capture complex spatial topologies. It generates overly smoothed, homogenized spatial distributions, entirely missing intricate local features such as the bifurcated soot tips observed in the first sample.

While models incorporating local convolutions or attention mechanisms (CNN, Enhanced MLP, CNNTransformer) successfully recover the macroscopic outlines of the soot, they struggle with high-frequency details. Their corresponding absolute error maps (the dark alternating rows) reveal pronounced bright regions along the soot boundaries and internal concentration gradients, indicating significant localized prediction deviations.

In stark contrast, the proposed Soot-MixFormer (rightmost column) demonstrates overwhelming visual superiority. It consistently reconstructs highly accurate spatial morphologies that are virtually indistinguishable from the Ground Truth. Notably, it perfectly resolves fine-grained structural nuances, including sharp soot edges and complex internal concentration variations. Furthermore, the absolute error maps for Soot-MixFormer are visibly the darkest and most uniform across all samples. This confirms that the model effectively suppresses extreme localized errors, achieving the highest pixel-level fidelity and reinforcing the quantitative findings presented in the previous sections.

4.6. Robustness to Measurement Noise

To evaluate noise robustness, zero-mean Gaussian noise with standard deviations ranging from 1% to 10% of the maximum signal amplitude was added to the 1D test

inputs. As summarized in Table 6, reconstruction performance degrades gradually with increasing noise intensity. Under noise-free conditions, the model achieves an SSIM of 0.800 and a PCC of 0.915. Even at 1% noise, the cosine similarity remains 0.900 and the SSIM stays above 0.71, indicating good preservation of the overall soot structure. Under more severe noise contamination, although the RMSE increases to 116.56 at 10% noise, the cosine similarity remains above 0.72, suggesting stable degradation rather than catastrophic failure. As shown in Figure 11, high noise levels mainly introduce localized artifacts and moderate peak deviations, while the major geometric structure and high-soot regions are still accurately recovered. These results demonstrate the strong noise tolerance and practical robustness of the proposed model.

Table 6. Quantitative evaluation metrics of the reconstruction model under varying levels of simulated measurement noise.

Noise_Std (%)	SSIM	PCC	Cosine	MSE	RMSE	MAE
0.00	0.800	0.915	0.957	1210.676	27.666	14.030
1.00	0.712	0.839	0.900	1586.96	32.70085	24.62321
2.00	0.597	0.618	0.737	4148.85	48.3472	36.80631
5.00	0.505	0.605	0.738	5177.13	58.8557	45.98187
10.00	0.424	0.559	0.722	5738.81	66.5669	54.0132

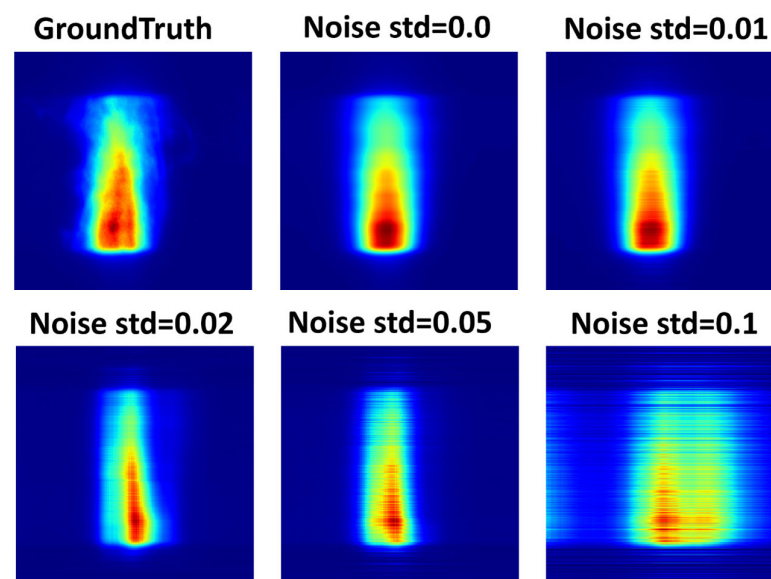


Figure 11. Qualitative comparison of the 2D soot spatial distribution reconstructions and absolute error maps under different noise levels.

4.7. Ablation Study

To systematically evaluate the individual contributions and synergistic effects of the core components within the proposed Soot-MixFormer, an extensive ablation study was conducted. Seven model variants were evaluated by removing specific modules or their combinations: Positional Encoding (w/o PosEnc), the Pyramid structural hierarchy (w/o Pyramid), the Dynamic Query mechanism (w/o DynQuery), the Dual-Axial Gated Refinement module (w/o DAGR), and the Physical Hard Constraint layer (w/o Hard Const.). The quantitative results for these variants on the test set are summarized in Table 7.

Table 7. Quantitative results of the ablation study on the test set. (↑ indicates that the changes made have caused the corresponding parameter value to increase, while ↓ signifies that the changes made have led to a decrease in the corresponding parameter value).

Model Variants	SSIM ↑	PCC ↑	Cosine Sim ↑	RMSE ↓	MAE ↓	MSE ↓
w/o PosEnc	0.692	0.821	0.884	39.421	21.854	2364.815
w/o Pyramid	0.745	0.864	0.912	34.115	18.230	1805.340
w/o DynQuery	0.768	0.875	0.931	31.842	16.540	1550.412
w/o DAGR	0.751	0.889	0.942	29.560	15.320	1358.105
w/o Hard Const.	0.785	0.901	0.951	32.410	16.890	1630.750
w/o DAGR & Hard Const	0.712	0.843	0.898	38.372	19.251	2131.243
w/o DAGR & DynQuery	0.725	0.832	0.882	36.224	18.903	1938.836
Soot-MixFormer	0.800	0.915	0.957	27.666	14.030	1210.676

Impact of Positional Encoding (w/o PosEnc): Removing positional encoding results in the most catastrophic performance drop across all metrics, with SSIM plummeting from 0.800 to 0.692 and MSE nearly doubling (from 1210.676 to 2364.815). This indicates that without explicit spatial coordinate awareness, the Transformer completely loses its ability to organize the 1D sequential tokens into a meaningful 2D topological structure.

Impact of the Pyramid Architecture (w/o Pyramid): The absence of the hierarchical multi-scale feature extraction leads to a severe structural degradation (SSIM drops to 0.745). This confirms that multi-scale representations are indispensable for capturing soot features across different receptive fields, especially for accurately predicting macroscopic outlines.

Impact of Dynamic Query Mapping (w/o DynQuery): Disabling the dynamic query generation severely limits the model's cross-dimensional mapping capability. The increase in RMSE (from 27.666 to 31.842) demonstrates that dynamically decoupling the 1D sensory input into a 2D feature space is vital for feature alignment.

Impact of Dual-Axial Gated Refinement (w/o DAGR): The removal of the DAGR module primarily harms the structural similarity (SSIM drops to 0.751). This empirical evidence proves that applying orthogonal refinement along the horizontal and vertical axes is crucial for restoring high-frequency local details and sharp concentration gradients.

Impact of the Physical Hard Constraint (w/o Hard Const.): Interestingly, removing the physical hard constraint only marginally affects the SSIM (0.785), but it drastically compromises the pixel-level absolute accuracy, causing the MSE to surge to 1630.750. This perfectly aligns with our theoretical hypothesis: while the base network can learn the general shape of the soot, the hard constraint is absolutely mandatory for strictly enforcing mass conservation along the optical path and suppressing extreme localized numerical anomalies.

Synergistic Effects of Module Combinations: To deepen the insights into the architecture's design, combined ablation scenarios were analyzed. Removing both DAGR and the Hard Constraint (w/o DAGR & Hard Const.) triggers a cascading performance failure, with SSIM dropping to 0.712 and MSE surging to 2131.243. This demonstrates a strong orthogonal complementarity: the Hard Constraint enforces global mass conservation, while DAGR secures localized high-frequency precision. Removing either one limits the other, but removing both cripples the network's spatial and physical fidelity. Similarly, removing DAGR alongside Dynamic Query (w/o DAGR & DynQuery) yields severe degradation (SSIM of 0.725). This proves that without DynQuery establishing an accurate global spatial mapping prior, DAGR lacks a reliable foundation for its residual spatial refinement.

Sensitivity Analysis on Rank Size (d): To justify the hyperparameter selection for the low-rank approximation in the generation head, we conducted a sensitivity analysis on the rank size d, as detailed in Table 8.

Table 8. Sensitivity analysis of the rank size (d) on the test set. (↑ indicates that the changes made have caused the corresponding parameter value to increase, while ↓ signifies that the changes made have led to a decrease in the corresponding parameter value).

Rank Size (d)	SSIM ↑	PCC ↑	Cosine Sim ↑	RMSE ↓	MAE ↓	RMSE ↓
d = 32	0.764	0.871	0.925	32.105	16.85	32.105
d = 64	0.786	0.894	0.941	29.84	15.21	29.84
d = 128 (Ours)	0.8	0.915	0.957	27.666	14.03	27.666
d = 256	0.802	0.916	0.958	27.51	13.95	27.51

The rank size dictates the representational capacity of the outer product generation. Setting a small rank ($d = 32$) imposes a severe information bottleneck, preventing the network from effectively decoding complex spatio-temporal features. While $d = 64$ shows clear improvement, increasing the rank to $d = 128$ achieves a significant performance breakthrough. Further expanding the rank to $d = 256$ yields extreme diminishing returns, but exponentially increases the computational complexity and memory footprint of the $\mathcal{O}(L^2 \cdot d)$ outer product operation. Therefore, $d = 128$ was selected as the optimal configuration, representing the ideal rank that maximizes reconstruction fidelity without incurring unnecessary computational overhead.

In summary, the full Soot-MixFormer achieves optimal performance by harmoniously integrating these modules, proving that both advanced sequence modeling, intelligent structural decoupling, and rigorous physical constraints are essential for high-fidelity tomographic reconstruction.

5. Discussion

The experimental results and visual comparisons presented in Section 4 demonstrate that the Soot-MixFormer provides a robust and efficient solution for the 2D reconstruction of soot concentration fields from 1D integrated data. This section discusses the inherent advantages of the proposed architecture and its performance characteristics under challenging morphological conditions.

The superior performance of Soot-MixFormer (SSIM of 0.800 and PCC of 0.915) over traditional CNNs and MLPs can be attributed to its unique hybrid design. By combining convolutional layers with Transformer blocks, the model simultaneously captures local soot gradients and the global symmetry inherent in soot structures. Unlike purely data-driven models, the integration of a physical hard constraint layer ensures that the reconstructed 2D field remains consistent with the original 1D line-of-sight measurements, effectively eliminating non-physical artifacts. Achieving a throughput of ~246 FPS with only 0.97 M parameters makes the model highly suitable for real-time monitoring in industrial settings where computational resources may be limited.

Despite its structural accuracy, a critical analysis of complex soot samples reveals specific limitations in texture rendering. As illustrated in Figure 12, the model successfully predicts the macroscopic envelope and the primary concentration zones. However, a discrepancy exists between the predicted internal textures and the Ground Truth. While the global topology is preserved, the fine-grained, stochastic textures and micro-scale fluctuations are often “smoothed” by the network. This is an inherent challenge in the 1D-to-2D inversion task; since 1D extinction data involves spatial integration, much of the high-frequency information regarding localized turbulence is lost or aliased.

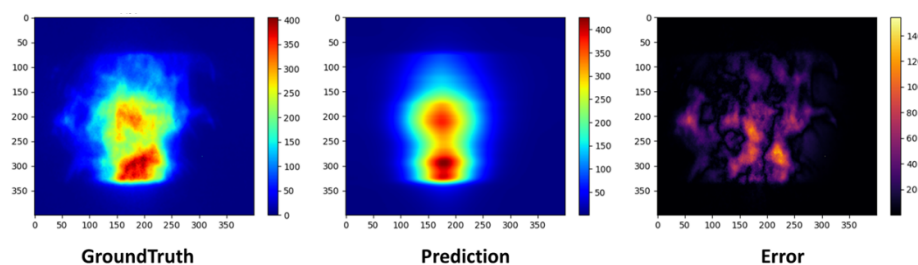


Figure 12. Detailed visualization of the reconstruction discrepancy in complex soot topologies.

The model tends to produce a “physically plausible” average of the internal structure rather than capturing the exact pixel-wise stochasticity. This suggests that while Soot-MixFormer is excellent for capturing the soot’s evolution and total soot yield, further research into generative adversarial loss or higher-order physical constraints may be required to recover sub-grid scale textures. This is an unavoidable limitation stemming from the severe information loss inherent in 1D line-of-sight integrated measurements. Because multiple highly textured 2D distributions can mathematically yield the same 1D projection vector, deterministic regression models inevitably tend to output a slightly smoothed, averaged representation to minimize global statistical errors, thereby losing high-frequency localized details.

To fundamentally overcome this ill-posed bottleneck, several advanced strategies must be explored in future research. First, transitioning from single-view to multi-view measurement systems, or utilizing limited-angle tomography, would provide additional spatial constraints to recover complex asymmetric textures. Second, multi-modal data fusion—such as combining 1D laser extinction signals with 2D emission imaging or laser-induced incandescence (LII)—could cross-validate and enrich the spatial features. Finally, rather than relying solely on deterministic point estimates, introducing probabilistic reconstruction approaches (e.g., Bayesian Neural Networks or Conditional Diffusion Models) would be highly beneficial. Such probabilistic frameworks could not only generate multiple plausible fine-texture candidates that satisfy the 1D physical constraints but also mathematically quantify the epistemic uncertainty of the reconstructed fields.

Future iterations could integrate a Discriminator network or specialized perceptual loss functions to mitigate the “averaging” effect. By introducing adversarial training, the model may better recover high-frequency stochastic textures and sub-grid scale turbulent features that are currently smoothed out.

While the current model focuses on 1D-to-2D inversion, the architecture can be extended to process multi-angle extinction data or be coupled with chemical kinetic information. Leveraging the Transformer’s ability to handle long-range dependencies, future work will explore the reconstruction of 3D volumetric soot distributions, providing a more comprehensive diagnostic tool for asymmetric and highly transient combustion processes.

6. Conclusions

This study successfully developed and validated Soot-MixFormer, a hybrid deep learning architecture designed to reconstruct 2D soot concentration distributions from 1D integrated extinction data. By synergistically combining the local feature extraction of CNNs with the global dependency modeling of Transformers, the model effectively overcomes the ill-posed nature of the 1D-to-2D inversion task. The integration of a Dual-Axial Gated Refinement (DAGR) module and a physical hard constraint layer ensures that the reconstructed fields not only achieve high structural fidelity but also strictly adhere to mass conservation laws, thereby eliminating non-physical artifacts common in purely data-driven approaches.

Experimental evaluations demonstrate that Soot-MixFormer delivers superior performance, achieving a Structural Similarity Index (SSIM) of 0.800 and a Pearson Correlation Coefficient (PCC) of 0.915, significantly outperforming traditional MLP and CNN baselines. With a compact parameter size of 0.97 M and an exceptional inference throughput of approximately 246 FPS, the model meets the rigorous demands for real-time, high-accuracy monitoring in industrial combustion environments. While the current architecture exhibits a slight smoothing effect on stochastic sub-grid scale textures due to the inherent information loss in 1D spatial integration, it establishes a robust and efficient diagnostic framework for soot field analysis. Future research will focus on incorporating generative adversarial training to enhance micro-scale details and extending the model to 3D volumetric soot reconstruction for more complex, transient soot.

Author Contributions: Conceptualization, Z.H.; methodology, Z.H.; software, Z.H.; validation, Z.H., W.Y. and X.F.; formal analysis, Z.H.; investigation, X.F.; resources, Z.H.; data curation, W.Y. and X.F.; writing—original draft preparation, Z.H.; writing—review and editing, W.Y.; visualization, Z.H. and W.Y.; supervision, S.L. and W.Y.; project administration, S.L.; funding acquisition, W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China [No. 52506170].

Data Availability Statement: The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Franzelli, B.; Roussillo, M.; Scoufflaire, P.; Bonnet, J.; Jalain, R.; Dormieux, T.; Candel, S.; Legros, G. Multi-diagnostic soot measurements in a laminar diffusion flame to assess the ISF database consistency. *Proc. Combust. Inst.* **2019**, *37*, 1355–1363. [\[CrossRef\]](#)
2. Bodor, A.L. Numerical Modelling of Soot Formation and Evolution in Laminar Flames with Detailed Kinetics. Ph.D. Thesis, Université Paris Saclay (COMUE), Gif-sur-Yvette, France, Politecnico di Milano, Milano, Italy, 2019.
3. Khanezhari, A.; Jadidi, M.; Zimmer, L.; Dworkin, S.B. Application of machine learning for the low-cost prediction of soot concentration in a turbulent flame. *Environ. Sci. Pollut. Res.* **2023**, *30*, 27103–27112. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Di Liddo, L.; Saldinger, J.C.; Jadidi, M.; Elvati, P.; Violi, A.; Dworkin, S.B. Exploring soot inception rate with stochastic modelling and machine learning. *Combust. Flame* **2023**, *258*, 112375. [\[CrossRef\]](#)
5. Wu, Y.; Zhang, Z.; Chai, W.S.; Zhou, L. A comprehensive review of optical systems for soot volume fraction measurements in Co-flow laminar flames: Laser-induced incandescence (lii) and laser extinction method (lem). *Combust. Sci. Technol.* **2025**, *197*, 874–906. [\[CrossRef\]](#)
6. Hofmann, M.; Bessler, W.G.; Schulz, C.; Jander, H. Laser-induced incandescence for soot diagnostics at high pressures. *Appl. Opt.* **2003**, *42*, 2052–2062. [\[CrossRef\]](#)
7. Huang, Q.; Wang, F.; Liu, D.; Ma, Z.; Yan, J.; Chi, Y.; Cen, K. Reconstruction of soot temperature and volume fraction profiles of an asymmetric flame using stereoscopic Tomography. *Combust. Flame* **2009**, *156*, 565–573. [\[CrossRef\]](#)
8. Verdugo, I.; Cruz, J.J.; Álvarez, E.; Reszka, P. Candle flame soot sizing by planar time-resolved laser-induced incandescence. *Sci. Rep.* **2020**, *10*, 11364. [\[CrossRef\]](#)
9. Snelling, D.R.; Thomson, K.A.; Smallwood, G.J.; Gülder, Ö.L. Spectrally resolved measurement of flame radiation to determine soot temperature and concentration. *AIAA J.* **2002**, *40*, 1789–1795. [\[CrossRef\]](#)
10. Lee, S.M.; Hwang, J.Y.; Chung, S.H. Tomographic reconstruction of asymmetric soot structure from multi-angular scanning. *J. Vis.* **2004**, *7*, 159–166. [\[CrossRef\]](#)
11. Li, Y.; Ren, N. Wind turbine wake reconstruction based on physical information guided neural network (PINN). In Proceedings of the 2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE), Changchun, China, 29–31 December 2024.
12. Zheng, S.; Zhang, M.; Yu, Y.; Na, M.; Liu, D. Simultaneous reconstruction of temperature and concentration of H₂O and soot based on infrared emission spectrum by inverse analysis. *Int. J. Therm. Sci.* **2024**, *195*, 108641. [\[CrossRef\]](#)

13. Dai, M.; Zhou, B.; Zhang, J.; Cheng, Y.; Wang, J.; Dong, S. 3-D soot temperature and volume fraction reconstruction of afterburner flame via deep learning algorithms. *Combust. Flame* **2023**, *252*, 112743. [\[CrossRef\]](#)
14. Müller, M.N.; Wang, Q.; Cai, W.; Huber, F.J.T.; Will, S. Tomographic single-shot time-resolved laser-induced incandescence for soot characterization in turbulent flames. *Proc. Combust. Inst.* **2024**, *40*, 105262. [\[CrossRef\]](#)
15. Chen, M.; Chen, Y.; Li, T.; Liu, D. Numerical and Experimental Reconstruction of Temperature Distribution and Soot Concentration Field for Thin-Slice Flames Using a Charge Coupled Device Camera. *Fire* **2023**, *6*, 392. [\[CrossRef\]](#)
16. Liu, G.; Liu, D. Direct simultaneous reconstruction for temperature and concentration profiles of soot and metal-oxide nanoparticles in nanofluid fuel flames by a CCD camera. *Int. J. Heat Mass Transf.* **2018**, *124*, 564–575. [\[CrossRef\]](#)
17. Liu, G.; Liu, D. Reconstruction model for temperature and concentration profiles of soot and metal-oxide nanoparticles in a nanofluid fuel flame by using a CCD camera. *Chin. Phys. B* **2018**, *27*, 054401. [\[CrossRef\]](#)
18. Liu, G.; Liu, D. Inverse radiation analysis for simultaneous reconstruction of temperature and volume fraction fields of soot and metal-oxide nanoparticles in a nanofluid fuel sooting flame. *Int. J. Heat Mass Transf.* **2018**, *118*, 1080–1089. [\[CrossRef\]](#)
19. Yao, W. Study on Soot Distribution Laws and Radiative Characteristics of Binary Hydrocarbon Buoyant Diffusion Flames. Ph.D. Thesis, University of Science and Technology of China, Hefei, China, 2023. (In Chinese) [\[CrossRef\]](#)
20. de Francqueville, L.; Bruneaux, G.; Thirouard, B. soot volume fraction measurements in a gasoline direct injection engine by combined laser induced incandescence and laser extinction method. *SAE Int. J. Engines* **2010**, *3*, 163–182. [\[CrossRef\]](#)
21. Kim, B.; Lee, J. A Video-Based Fire Detection Using Deep Learning Models. *Appl. Sci.* **2019**, *9*, 2862. [\[CrossRef\]](#)
22. Rezaeian, A.; Davoodi, M.; Jafari, M.K.; Bagheri, M.; Asgari, A.; Kafshgarkolaei, H.J. Optimization of earth dam cross-sections using the max-min ant system and artificial neural networks with real case studies. *Buildings* **2026**, *16*, 501. [\[CrossRef\]](#)
23. Akbarzadeh, M.R.; Naeim, B.; Asgari, A.; Estekanchi, H.E. Framework for multi-hazard parameterized fragility based uncertainty quantification and sensitivity analysis of offshore wind turbines. *Soil Dyn. Earthq. Eng.* **2026**, *201*, 109963. [\[CrossRef\]](#)
24. Dewangan, A.; Pande, Y.; Braun, H.-W.; Vernon, F.; Perez, I.; Altintas, I.; Cottrell, G.W.; Nguyen, M.H. FigLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection. *Remote Sens.* **2022**, *14*, 1007. [\[CrossRef\]](#)
25. Xie, Y.; Zhu, J.; Cao, Y.; Zhang, Y.; Feng, D.; Zhang, Y.; Chen, M. Efficient Video Fire Detection Exploiting Motion-Flicker-Based Dynamic Features and Deep Static Features. *IEEE Access* **2020**, *8*, 81904–81917. [\[CrossRef\]](#)
26. Cheng, X.; Ren, F.; Gao, Z.; Wang, L.; Zhu, L.; Huang, Z. Predicting 3D distribution of soot particle from luminosity of turbulent flame based on conditional-generative adversarial networks. *Combust. Flame* **2023**, *247*, 112489. [\[CrossRef\]](#)
27. Leite, V.C.; Merzari, E.; Ponciroli, R.; Ibarra, L. Convolutional neural network-aided temperature field reconstruction: An innovative method for advanced reactor monitoring. *Nucl. Technol.* **2023**, *209*, 5645–5666. [\[CrossRef\]](#)
28. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
29. Chen, M.; Zheng, R.; Zhao, X.; Liu, D. 2D flame temperature and soot concentration reconstruction from partial discrete data via machine learning: A case study. *Case Stud. Therm. Eng.* **2025**, *69*, 106005. [\[CrossRef\]](#)
30. Wang, Q.; Gong, M.; Matynia, A.; Zhang, L.; Qian, Y.; Dang, C. Soot temperature and volume fraction field predictions via line-of-sight soot integral radiation equation informed neural networks in laminar sooting flames. *Phys. Fluids* **2024**, *36*, 121703. [\[CrossRef\]](#)
31. Li, H.; Ren, T.; Zhao, C. A physics-informed neural network for non-linear laser absorption tomography. *J. Quant. Spectrosc. Radiat. Transf.* **2025**, *330*, 109229. [\[CrossRef\]](#)
32. Molnar, J.P.; Xia, J.; Zhang, R.; Grauer, S.J.; Liu, C. Unsupervised neural-implicit laser absorption tomography for quantitative imaging of unsteady flames. *Combust. Flame* **2025**, *279*, 114298. [\[CrossRef\]](#)
33. Widmann, J.; Yang, J.; Smith, T.; Manzello, S.; Mulholland, G. Measurement of the optical extinction coefficients of post-flame soot in the infrared. *Combust. Flame* **2003**, *134*, 119–129. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.