

DED bead geometry and profile prediction with multimodal spatio-temporal neural networks

Wenze Zhang^{a,b,d,1}, Yichen Wang^{b,1}, Yuanzhi Chen^{a,b,d}, Xiaoke Deng^{a,d}, Yihe Wang^b, Molong Duan^a, Pengcheng Hu^b, Kai Tang^{b,c,d,*}

^a Hong Kong University of Science and Technology, Hong Kong Special Administrative Region of China

^b The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

^c HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China

^d The Hong Kong University of Science and Technology Shenzhen Research Institute, Shenzhen, China

ARTICLE INFO

Keywords:

Additive manufacturing
Directed energy deposition
Bead geometry prediction
Multimodal data fusion
Spatio-temporal neural network
Machine learning

ABSTRACT

Accurate prediction of bead geometry in multi-axis directed energy deposition (DED) remains challenging due to dynamic interactions between process parameters, positional dynamics, and rapid melt pool evolution. This study introduces MST-Net, a multimodal spatio-temporal neural network, alongside a comprehensive dataset of 1.2 million samples from a five-axis DED system, integrating co-axial melt pool images, control parameters, positional variables, and high-resolution point cloud labels. MST-Net employs a hierarchical architecture to fuse spatio-temporal features, achieving state-of-the-art prediction accuracy with a mean intersection over union (mIoU) of 0.93 for cross-sectional profiles and mean average precision (mAP) of 0.95 for bead dimensions. Ablation studies reveal melt pool images as the most critical input modality, while positional data governs peak localization. Temporal analysis shows asymmetric weighting of historical (look-back) and future (look-ahead) contexts optimizes predictions, with 201-step sequences balancing accuracy and computational efficiency at 116 FPS, which provides a foundation for real-time control given the predictive model. Transfer learning experiments highlight MST-Net's adaptability, maintaining 0.95 mAP with only 10 % of training data for new printing path geometries. By addressing data scarcity and spatio-temporal complexity, this work advances predictive modeling for multi-axis DED, offering a robust framework for real-time process control in industrial applications. The dataset and model are publicly released to foster innovation in metal additive manufacturing.

1. Introduction

Directed energy deposition (DED) is revolutionizing metal additive manufacturing by enabling the rapid production of large near-net-shape parts, high-performance materials, and repairs of damaged workpieces in aerospace, automotive, and marine industries [1–3]. In DED, a focused energy source, such as a laser beam, electron beam, or plasma arc, generates a moving melt pool on a metallic substrate. Metallic feedstock, in the form of powder or wire, is simultaneously introduced into the melt pool, where deposition and subsequent bead formation facilitate intrinsic 3D structures. Among various DED techniques, Laser Powder DED (LP-DED) is a prominent example (Fig. 2), where a laser beam forms a small melt pool on the substrate, while compressed inert gas delivers metal powder into the melt pool to form fully dense metallic

parts [1,4]. Compared to other DED methods, LP-DED offers higher dimensional accuracy, improved surface roughness, and a smaller heat-affected zone [2,3]. Additionally, the powder feedstock provides greater material versatility, enabling the combination of different powders for heterogeneous parts [1]. The integration of multi-axis motion systems, such as five-axis CNC [5] or robotic arms [6,7], further enhances LP-DED by optimizing deposition orientations on complex surfaces, reducing material waste, and minimizing the need for support structures [8–10].

However, LP-DED has several limitations due to its dynamic nature, which presents challenges for real-time geometry control, especially for large-volume and complex parts. As the laser moves, the melt pool forms, evolves, and rapidly solidifies in a spatio-temporal manner, as illustrated in the thermal images of Fig. 3. Spatially, the melt pool shape

* Corresponding author at: The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.

E-mail addresses: pengchenghu@hkust-gz.edu.cn (P. Hu), mektang@ust.hk (K. Tang).

¹ Joint first authors

changes with varying travel speed, and the bead geometry is affected by factors such as radius of curvature at corners [11,12], overlapping regions between tracks [13,14], and height variation in multi-layer deposition [15–17]. The powder collection rate, influenced by melt pool size, also leads to changes in bead geometry. Temporally, the deposition process is continuous, with the melt pool remaining quasi-stable while complex laser-powder-liquid interactions occur [3]. This process can be perturbed by fluctuations in laser power, inconsistent powder flow, elevated substrate temperature as the deposit height increases [16,18,19], or thermal deformation due to heat accumulation, which can cause shifts in powder-laser alignment [10,20]. The DED parts are prone to internal defects like porosity, lack of fusion, or cracks due to the anisotropic and nonuniform thermal properties [21]. Therefore, accurate monitoring and control of the DED process are crucial for maintaining structural integrity and preventing defects. Furthermore, in multi-axis DED with RTCP (Rotating Tool Center Point) function, combined translational and rotational motions alter the effective scan speed and beam incidence, leading to variations in melt pool symmetry. For example, when RTCP is enabled and motion is dominated by the C-axis, the melt pool can experience centrifugal forces that reduce its stability and symmetry. To address these challenges, the development of spatio-temporal predictive models is essential for interpreting the DED process and maximizing its potential for industrial applications.

Several approaches have been reported for DED process prediction, including numerical simulations and data-driven machine learning (ML) models. While numerical simulations provide insights into heat and mass transport physics, they face challenges with large-scale parts and complex geometries due to high computational costs, simplified boundary condition assumption, and limited temporal and spatial scales [22–24]. Also, physical phenomena behind AM processes are not yet fully understood [25]. Recent research is shifting toward low-fidelity or data-driven approaches to fill gaps in process-structure-property-performance predictions [25]. These data-driven approaches, emerging since 2021 [21], are well-suited for real-world applications, as they can model complex, non-linear relationships, and tolerate inherent uncertainties. Among ML algorithms, neural networks (NN) are the most widely used [26] due to the availability of large datasets, powerful computational resources, and advanced algorithms. The multi-layer perceptron (MLP) [27] is the most common NN type, typically involving linear summation and nonlinear activation functions. For example, Gühr et al. [11] used MLP to predict bead geometry, particularly at corners and across multiple layers, enhancing the understanding of the dynamic deposition process. Additionally, the convolutional neural network (CNN) excels in image processing by leveraging the spatial relationships between pixels [13], with advanced derivatives developed for image segmentation, multi-object detection, and image generation. For example, the ResNet families [28], as applied in [29], use residual learning to address the vanishing gradient problem in deep networks. Asadi et al. [30] and Yu et al. [31] demonstrated the effectiveness of YOLO-based algorithms for fast melt pool segmentation, linking process parameters to melt pool geometry. Also, Wang et al. [32] proposed a feedback control strategy based on melt pool vision, determining bead width through the EPNet segmentation network.

Recently, more studies have incorporated deep learning models with spatio-temporal capabilities for predicting geometry and thermal fields in DED processes. Techniques such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks [39], bidirectional LSTMs (Bi-LSTMs) [40], encoder-decoder architecture [40–42], attention mechanisms [42], and graph neural networks (GNNs) [43,44] have all been employed in various fields to model sequential and structured data. A brief description of them is given below:

1. **RNNs** process sequential data by maintaining a hidden state that captures temporal dependencies. For instance, Ren et al. [15] utilized RNNs to explore the relationships between laser paths and

thermal histories, while Zhou et al. [18] applied physics-informed RNNs to identify thermal fields during DED. However, RNNs face challenges like vanishing gradients in long sequences.

2. **LSTMs** overcome RNN limitations by incorporating memory cells to retain long-term dependencies, and **Bi-LSTMs** extend this by processing data in both forward and backward directions, capturing more context for improved performance. For example, Zhang et al. [19] utilized XGBoost and LSTM models for accurate melt pool temperature estimation, while Hu et al. [33] explored LSTM and CNN-BiLSTM architectures for predicting melt pool width.
3. The **encoder-decoder** architecture uses an encoder to compress an input sequence into a context vector, which is then decoded into an output sequence, making it ideal for tasks such as language translation. For example, Ren et al. [35] developed a LSTM-autoencoder for DED quality classification based on emission spectroscopy.
4. **GNNs** handle relational data by propagating information across nodes, effectively modeling spatio-temporal relationships, particularly when data is graph-structured. Graph convolutional networks (GCNs), a subtype of GNNs, focus on spatial dependencies using graph-based convolutions, making them effective for spatially correlated tasks. For instance, Mozaffar et al. [38] combined GNN with RNN to model spatio-temporal dependencies of melt pool thermal responses.
5. **Attention** mechanisms enable models to focus on key features in sequential prediction, with self-attention capturing dependencies across all positions and **multi-head attention** enhancing the handling of complex, non-linear relationships by attending to multiple subspaces. For example, Yu et al. [31] integrates self-attention for real-time monitoring and prediction of deposition height in DED processes.

These approaches, published in top-tier journals and summarized in Table 1, show that incorporating spatio-temporal data significantly enhances the prediction accuracy and model robustness.

Another approach to accommodating spatio-temporal dynamics is through multimodal data fusion [45], which integrates various types of data containing abundant intra- and cross-modality information [41, 46]. Here, "modality" refers to a signal source or encoding mechanism, typically collected at a fixed sampling rate. In DED, the collectable signals include melt pool images, emission spectra, thermal images, laser power, travel speed, layer number, motion data, point-wise temperature by pyrometers, and more (Table 1). Since each modality captures different aspects of the system dynamics, combining them can improve predictive accuracy. For example, Chen et al. [47] integrated melt pool features from high-speed camera images with plasma spectral signals using an XGBoost-based system for real-time quality prediction. Dong et al. [29] developed a CNN-based network that combines images, temperature, and process parameters to optimize process window and layer quality. Despite progress, the development of predictive models in DED is hindered by the lack of publicly available multimodal datasets [45]. Current datasets are limited to 100,000 samples, as summarized in Table 1, and, most critically, often lack detailed geometric profile labels, being primarily derived from microscope cross-sections or numerical simulations. The difficulty in capturing accurate bead geometry in real-time arises primarily from point cloud data being obstructed by deposited material and misaligned with the melt region due to variations in printing angles and directions, especially in multi-axis setups, as shown in Fig. 2d-e. These challenges emphasize the need for more comprehensive multimodal datasets that capture key process parameters, positional data, and accurate geometric profiles for more robust predictive modeling.

Given these advancements and challenges, we propose a multimodal spatio-temporal neural network (MST-Net) that utilizes a custom-designed dataset, created and curated by our team, to provide accurate and generalizable bead geometry predictions, as shown in Fig. 1. The primary contributions of this study are as follows:

Table 1

Recent Deep learning (DL) approaches for DED predictions. (MPI: Melt pool image, LP: Laser power, TS: Travel speed, PFR: Powder feed rate, LSD: Laser spot diameter, LI: Layer index, NA: Not available.).

Category	Year	Energy Source	Material Feedstock	DL Model	Model Input	Model Output	Dataset Size	Ref
Geometrical Feature Regression	2024	Laser	Wire	YOLO, CNN	MPI	Bead Geometry	3.0k	[30]
	2024	Laser	Powder	YOLO, Attention	MPI, LI, Pixel number, Scan direction, Layer height, Total height	Bead Height	5.0k	[31]
	2024	Arc	Wire	MLP	TS, LI, Material, Corner Angle, Distance to Corner	Bead Geometry	28.8k	[11]
	2023	Laser	Powder	Resnet, CNN	MPI, LP, LSD, TS, PFR, Temperature	Bead Geometry	0.1k	[29]
	2022	Laser	Powder	CNN, Bi-LSTM	LP, TS, PFR, LI, Image Index, Melt Pool Width	Melt Pool Width	48.0k	[33]
	2021	Laser	Powder	MLP	LP, TS, MPI by IR camera, Track Profile from Line Scanner	Melt Pool Depth	0.4k	[34]
Quality Classification	2021	Arc	Wire	ErfNet	MPI	Segmented MPI	0.6k	[32]
	2022	Laser	Powder	LSTM	Plasma Spectra	Surface Quality	4.3k	[35]
	2022	Laser	Powder	RandLA-Net	Point Cloud	Surface Quality	0.2k	[36]
	2021	Arc	Wire	CNN	Magneto-optical Image	Surface Quality	4.8k	[37]
Thermal Field Regression	2021	Laser	Powder	RNN, MLP	Thermal Field, Infill Toolpath	Thermal Field	0.3k	[15]
	2021	Laser	Powder	GNN, RNN	Layer Height, Distance from Laser	Thermal Field	NA	[38]
	2021	Arc	Wire	Bi-LSTM, CNN	Activation State, Thermal Field	Thermal Field	98.6k	[18]
	2021	Laser	Powder	LSTM	LP, TS, LI, Time Index, Average Height, Average Width	Melt Pool Temperature	17.5k	[19]

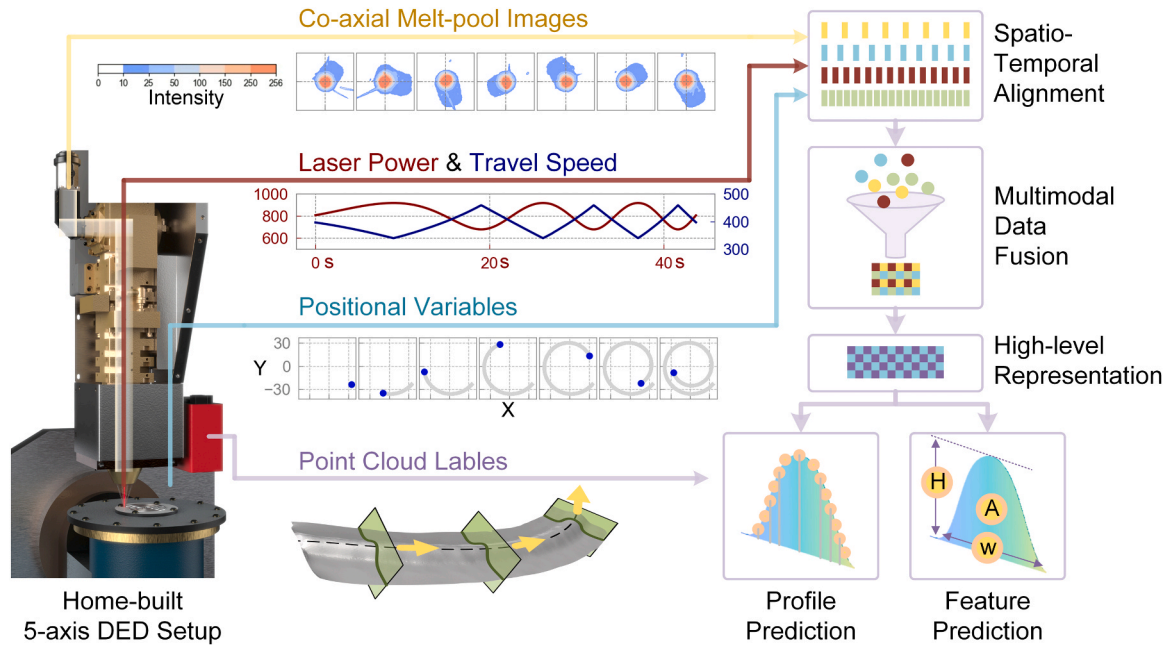


Fig. 1. Overview of the multimodal spatio-temporal neural network (MST-Net) for DED bead geometry prediction. Co-axial melt pool images, laser power, travel speed, and positional data are collected from a home-built 5-axis DED setup and input into the network. These data are spatio-temporally aligned and undergo multimodal data fusion to generate a high-level representation of melt pool dynamics, which can be used to predict DED outcomes such as geometrical profiles or other task-specific features. The point cloud of the printed bead is sliced to obtain cross-sectional geometry, which serves as the model's label.

1. **Comprehensive Multimodal Dataset:** To address the data scarcity issue, we have developed a large-scale dataset containing over 1.2 million samples, collected from a five-axis DED setup with RTCP enabled. Each sample in the dataset includes co-axial melt pool images, 5 types of control parameters (e.g., laser power, feed rate), and 15 types of positional variables (e.g., axis positions, velocities, and accelerations) as input, paired with point cloud geometries of the bead as labels. The dataset is carefully curated and temporally aligned. These publicly available resources (see [Supplementary Materials](#)) fill a crucial gap for studying the spatio-temporal dynamics in DED processes and supports data-driven approaches in additive manufacturing.

2. **Advanced Network Architecture:** We introduce MST-Net, a spatio-temporal neural network that predicts bead profiles with exceptional accuracy, achieving a mean Intersection over Union (mIoU) of 0.93 for cross-sectional profiles and a mean average precision (mAP) of 0.95 for bead width, height and area regression. These results set a new benchmark in DED predictions. MST-Net is capable of predicting irregular shapes, including those that are not symmetrical or uniform, a key feature in real-world DED applications. Its modular architecture, featuring embedding, encoder, and decoder modules, allows for task-specific adaptations, as shown by the strong transfer learning results.

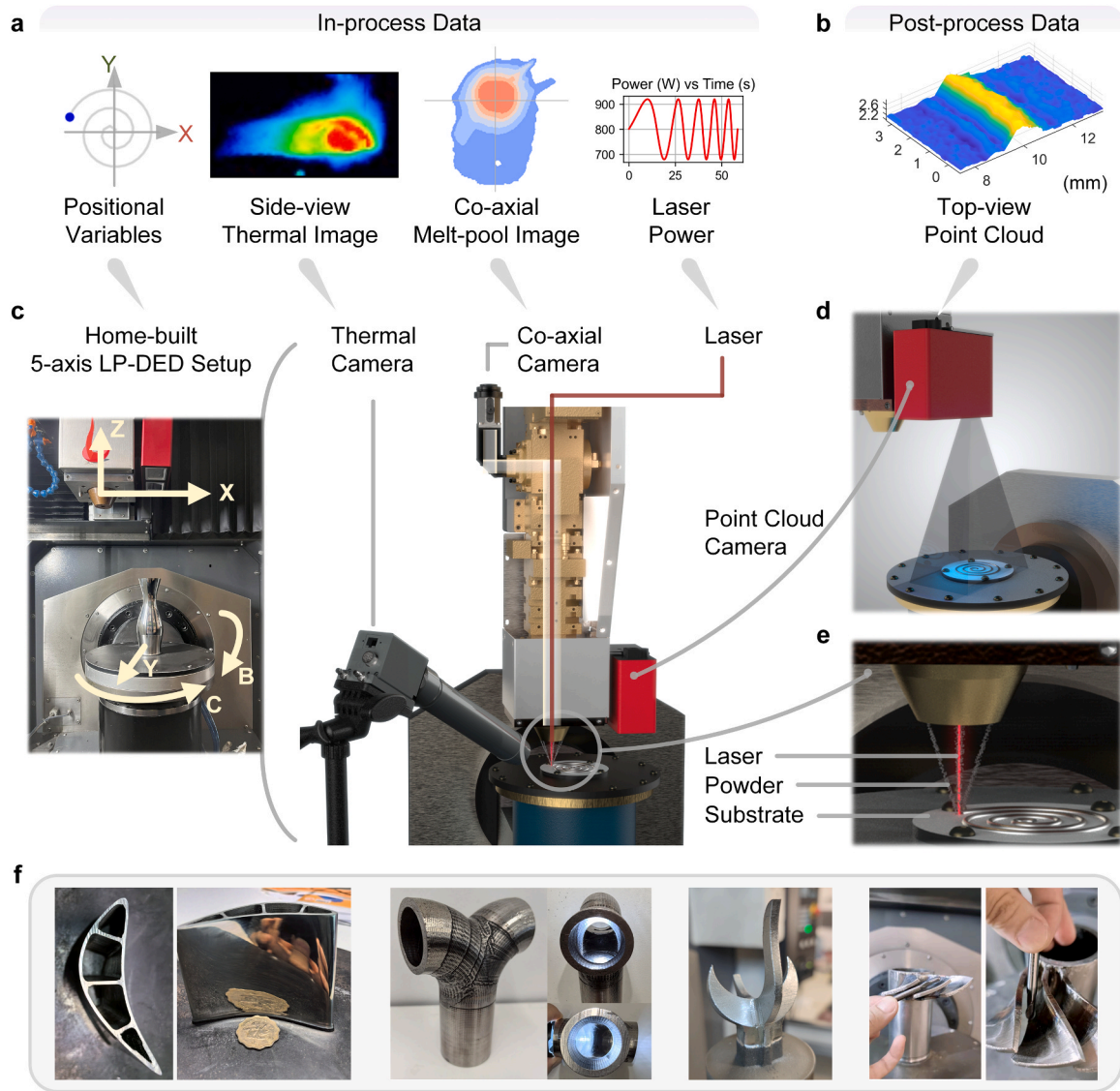


Fig. 2. The schematic diagram of the LP-DED printing and monitoring system: (a) in-process data, including positional variables, side-view thermal images, co-axial melt pool images and laser power; (b) post-process data, including the top-view point cloud data of the printed part; (c) the home-built 5-axis LP-DED setup; (d) point cloud collection process; (e) closer-up of the LP-DED process; and (f) home-printed parts produced using the 5-axis LP-DED setup.

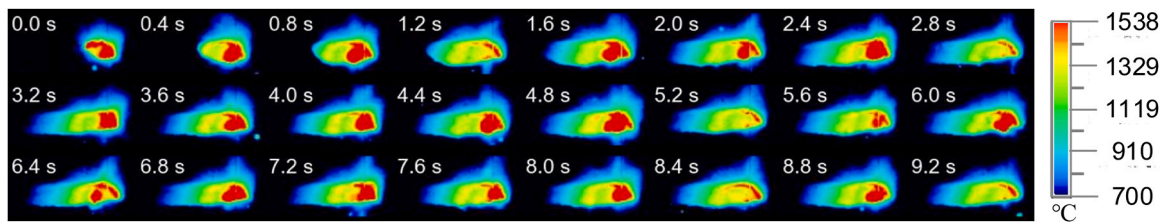


Fig. 3. Time-series thermal images of the evolving melt pool in a DED process.

3. *Temporal Dynamics Insight:* By analyzing temporal dependencies and melt pool asymmetry, we demonstrate that optimal predictions require asymmetric weighting of historical (look-back) and future (look-ahead) contexts. While look-back alone minimizes latency for real-time control, integrating look-ahead improves accuracy. Longer sequences (>300 steps) yield diminishing accuracy gains and higher computational costs, necessitating balanced sequence selection (201 steps optimal). This insight advances the understanding of temporal

dynamics in DED, enabling improved prediction accuracy while maintaining real-time feasibility.

4. *Prediction-Based Closed-Loop Control:* We demonstrate that prediction-based closed-loop control improves DED by using real-time bead geometry predictions to adjust travel speed with a PID controller. This approach reduces mean error by 45.0 % and RMSE by 48.5 % compared to open-loop control, enhancing workpiece quality.

5. **Multimodal Data Impacts:** Through ablation studies, we evaluate the impact of different input modalities on model performance. Co-axial melt pool images most significantly affect the prediction accuracy, followed by positional variables and control parameters. Positional variables, such as axis positions and velocities, mainly influence peak

location, while control parameters like laser power and travel speed are key for predicting cross-sectional geometry. These findings underscore the importance of multimodal data integration for accurate DED predictions.

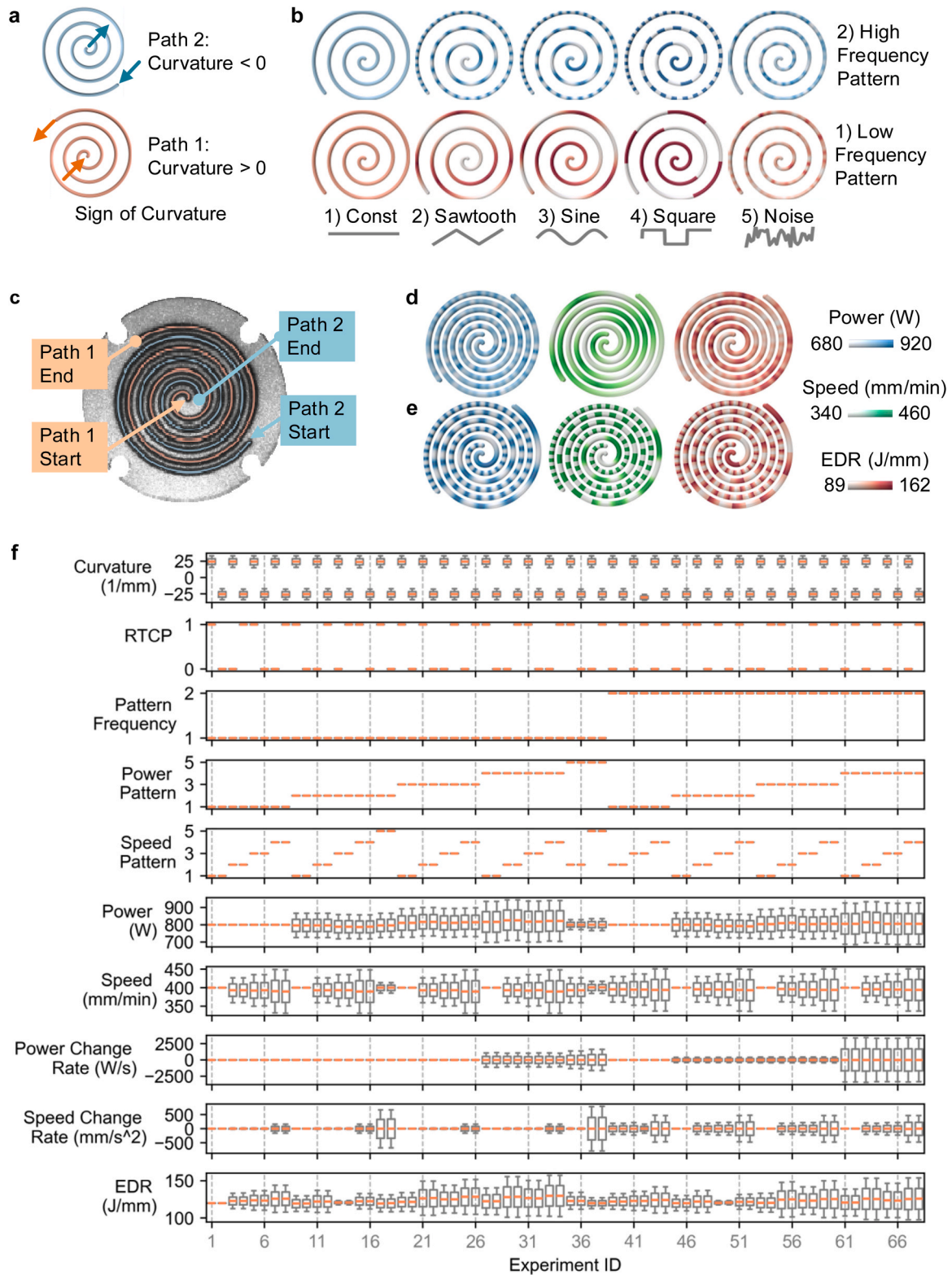


Fig. 4. Control parameter design for the dataset: (a) schematic of curvature signs with respect to the rotational direction using the right-hand rule; (b) design of periodical signal patterns for laser power and travel speed; (c) top-view image of printed part with two curves; (d-e) examples of travel speed, laser power, and energy deposition rate (EDR) superimposed on the motion path; and (f) combination of control parameters for 68 experiments.

In the rest of the paper, Section 2 describes the development and preprocessing of the dataset, including parameter design, experimental setup, data collection, feature registration, and exploratory data analysis. Section 3 details the MST-Net model, including its architecture, evaluation metrics, and training methodology. Section 4 presents the results on model prediction, deployment, and transfer learning, discussing the influence of multimodal data and temporal dynamics on the model's performance, and its application in a closed-loop control environment. Finally, Section 5 concludes the paper by summarizing key findings, limitations, and future research directions.

2. Dataset development and preprocessing

To ensure data rationality and diversity while expanding the dataset, this study focused on three key factors—printing path, motion parameters, and process parameters—that significantly influence the bead geometry, resulting in 68 distinct datasets for DED processes. During each operation, real-time melt pool images were captured using a CCD camera, and positional variables were logged by the motion controller. Post-printing, point cloud data of the printed bead were acquired using a 3D scanner. Spatial-temporal feature registration was performed at each sampling point along the DED path. The experiment design aimed to: 1) vary control parameters to induce dynamic melt pool responses, 2) expand parameter range to improve model generalizability, and 3) reduce the number of experiments to enhance efficiency in both testing and model training.

2.1. Printing path and motion parameter design

The printing path plays a crucial role in the formation of the melt pool in DED processes. This study employed an equidistant spiral path (Fig. 4a) to better reflect real-world printing scenarios, as its curvilinear nature is more realistic compared to straight-line paths. The equation for the equidistant spiral in polar coordinates is given by:

$$r = k_1 + k_2\theta \quad (1)$$

where r is the radial distance, θ is the polar angle, and k_1 and k_2 are constants. To ensure smooth motion and printing stability, the spiral path was discretized with a fixed sampling interval of $\mathcal{S} = 0.1$ mm, with a total spiral length of $\mathcal{L} = 400$ mm. The angular position θ_l and curvature $\mathcal{E}_l$ at each traveling distance l are calculated as:

$$\theta_l = \frac{\sqrt{k_1^2 + 2k_2l} - k_1}{b}, \quad \mathcal{E}_l = \frac{r^2 + 2k_2^2}{(r^2 + k_1^2)^{1.5}} \quad (2)$$

where $l = i \cdot \mathcal{S}$ and i is an integer within $[0, \mathcal{L}/\mathcal{S}]$. After discretization, the polar coordinates were converted to Cartesian coordinates for input to the motion controller.

The complex structure of the printing head (Fig. 2) introduces challenges in aligning the powder flow, laser optical path, and the machine's Z-axis, potentially causing the printing direction to affect the melt pool dynamics. To study the effects of printing direction and curvature, two representative printing paths were selected and deposited on a single substrate (Fig. 4c). Path 1 was printed counterclockwise with $(k_1, k_2) = (1, 1.7)$, while Path 2 was printed clockwise with $(k_1, k_2) = (1.7\pi + 1, 1.7)$, as depicted in Fig. 4a and the 1st row of Fig. 4f.

Additionally, the study examined the effect of the RTCP function in multi-axis setups, which coordinates translation and rotation axes for synchronized motion. A realistic multi-axis environment was simulated by randomly assigning one of the two paths to each substrate and incorporating RTCP (2nd row of Fig. 4f), introducing additional motion in the C-axis alongside X and Y axes.

2.2. Process parameter design

Laser power and travel speed were selected as the control parameters due to their rapid response (within tens of milliseconds) to control signals during printing. To investigate their effects, five distinct patterns—constant, sawtooth wave, sine wave, square wave, and random noise—were applied to both parameters (Fig. 4b). Each pattern was implemented in low and high-frequency versions to expand the dataset's diversity.

To induce time-varying melt pool responses, a fixed phase difference of $\pi/2$ was introduced between power and speed patterns along each print path (Fig. 4d-e). The laser power $\mathcal{P}$ and travel speed $\mathcal{T}$ at each traveling distance l are expressed as:

$$\mathcal{P}_{ij} = p_c \psi_j \left(\frac{2\pi l}{\mathcal{L}} (\alpha + i\beta) \right), \quad \mathcal{T}_{ij} = f_c \psi_j \left(\frac{2\pi l}{\mathcal{L}} (\alpha + i\beta) - \frac{\pi}{2} \right) \quad (3)$$

where $p_c = 800$ W and $f_c = 600$ mm/min are the centre values for power and speed, respectively; i is an integer within $[0, \mathcal{L}/\mathcal{S}]$; $j \in \{1, 2, 3, 4, 5\}$ corresponds to the constant, sawtooth wave, sine wave, square wave, or noise pattern, respectively. The function ψ_j denotes the waveform associated with each pattern. The constants α and β were set to (1, 0.001) for low frequency and (5, 0.01) for high frequency.

The combination of different signal patterns and phase shifts leads to variations in the energy deposition rate (EDR) in the unit of J/mm, which is defined as:

$$\text{EDR} = \frac{\text{Laser Power}}{\text{Travel Speed}} \quad (4)$$

In total, 2 curvatures, 2 pattern frequencies, 5 power patterns, and 5 speed patterns resulted in $2 \times 2 \times 5 \times 5 = 100$ combinations. However, since the constant and random noise waveforms are frequency-independent and excessive random noise contributes minimally to the prediction model's generalization, 68 parameter combinations were carefully selected for the datasets (Fig. 4f). The control parameters and positional variables collected in the dataset are listed in Table 2.

2.3. Experimental setup

In this study, a custom-designed DED system was utilized, equipped

Table 2

List of the model inputs, including co-axial melt image, 5 control parameters and 15 positional variables. The mean and standard deviation (Std) are calculated from the training data. (For RTCP: 1 = On, 0 = Off; for Rotary direction: 1 = Counterclockwise, -1 = Clockwise, WCS: World Coordinate System).

	Co-axial Melt Pool Images	Mean	Std
1.	224 × 224 pixel grey-scale images	-	-
	Control Parameters	Mean	Std
1.	RTCP ON/OFF	0.46	0.50
2.	Rotary Direction	0.0	1.00
3.	Radius of Curvature (mm)	25.06	8.41
4.	Laser Power (W)	804.32	83.66
5.	Travel Speed (mm/min)	395.06	40.30
	Positional Variables	Mean	Std
1.	Print Distance (mm)	199.38	115.03
2.	Print Time (s)	30.56	17.44
3.	Axis Position X (mm)	0.42	17.85
4.	Axis Position Y (mm)	-0.66	19.57
5.	WCS Position X (mm)	4.51	19.27
6.	WCS Position Y (mm)	0.23	17.55
7.	Axis Position C (mm)	555.82	455.35
8.	Axis Velocity X (mm/s)	-0.20	7.19
9.	Axis Velocity Y (mm/s)	0.41	6.84
10.	Axis Velocity C (degree/s)	8.25	11.27
11.	WCS Theta (degree)	5.49	96.94
12.	Spiral Theta (degree)	1.94	104.14
13.	Axis Acceleration X (mm/s ²)	0.16	21.87
14.	Axis Acceleration Y (mm/s ²)	0.015	20.61
15.	Axis Acceleration C (degree/s ²)	-0.51	19.82

with state-of-the-art monitoring technologies, such as a co-axial CCD camera, a side-view structured light camera, and an infrared (IR) thermal camera, as depicted in Fig. 2c. The DED system is responsible for depositing metal beads, while a motion controller tracked spatial and temporal coordinates at a 1000 Hz frequency. The powder material used is 316 L stainless steel, with the particle size between 53 and 150 μm , while the substrate was also composed of 316 L stainless steel. A Raycus RFL-C2000 laser (6 kW max power) was utilized, paired with a Precitec YC52 printing head featuring a 2 mm laser spot diameter.

The co-axial CCD camera (Daheng MER2-202-60GM/C) captured top-down images of the melt pool at a 200 Hz frequency and a resolution of 640×480 pixels. To assess the bead geometry, a side-view structured light camera (Deep Innovation M05G016-230) was installed, providing measurements with an accuracy of 50 μm . After deposition, the camera was repositioned above the bead to capture its point cloud data, as shown in Fig. 2d. The Xiris VXIR-3000 shortwave infrared camera, placed outside the printing chamber, monitored the thermal map surrounding the melt pool. It remained stationary while the printing head moves and measures temperatures from 350 to 3000 $^{\circ}\text{C}$, with a resolution of 640×512 pixels (15 $\mu\text{m}/\text{pixel}$) at a 100 Hz frequency.

2.4. Data collection and feature registration

Feature registration to the same ST coordinates is crucial for integrating data collected at different sampling rates and timestamps. This process requires resampling and temporal alignment to ensure the creation of a coherent dataset. Since the melt pool images cannot be

interpolated directly, the control parameters and positional variables are first aligned to the temporal stamps of the co-axial images. Subsequently, the point cloud data is aligned using the positional timestamps.

During the DED process, the co-axial camera captures top-view images of the melt pool at a fixed time interval τ (Fig. 5a). These melt pool images, captured along the entire path, can be represented as $X_0 \in \mathbb{R}^{N_0 \times H_0 \times W_0}$, where N_0 is the number of images, and H_0 and W_0 are the height and width of the images in pixels. The timestamps corresponding to X_0 are denoted as:

$$T_0 \in \mathbb{R}^{N_0} \quad (5)$$

The five control parameters and fifteen positional variables sampled at various points along the printing path (as detailed in Table 2) are represented as $Q_c \in \mathbb{R}^{N_c \times 5}$, and $S_c \in \mathbb{R}^{N_c \times 15}$, where $N_c = \mathcal{L}/\mathcal{I} + 1$ is the number of signals gathered, with the subscript c denoting the DED controller. These variables include parameters such as laser power, positions of the X, Y, and C axes, as well as their velocities and accelerations. The associated timestamps T_c are given by:

$$T_c \in \mathbb{R}^{N_c} \quad (6)$$

Since T_0 and T_c share identical starting and ending points on the temporal axis, spatial interpolation of the recorded toolpath was used to determine the control parameters q_i and positional variables s_i for each corresponding melt pool image x_i from X_0 . By collecting s_i for all images, we obtain the spatial stamps of the system, represented as:

$$S_0 \in \mathbb{R}^{N_0 \times 15} \quad (7)$$

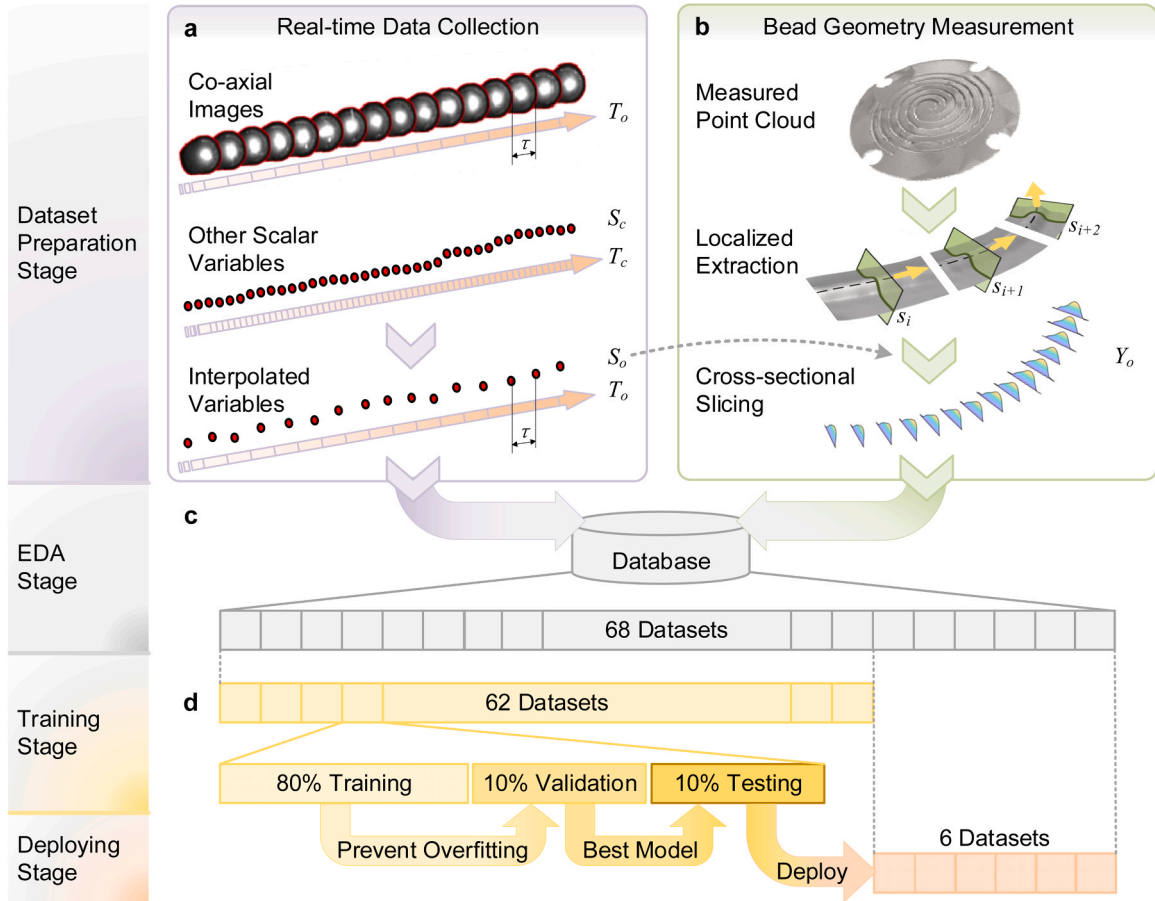


Fig. 5. Dataset preparation and utilization workflow, including stages of dataset preparation, exploratory data analysis (EDA), model training, and model deploying: (a) real-time signal collection and alignment to positional stamps; (b) bead geometry measurement from point cloud collection to cross-section profile sectioning along the motion direction; (c) spatio-temporal registration of pre-processed data into the database; and (d) data splitting into training, validation, test, and deploying datasets (with 6 out of 68 datasets for deployment).

After printing, bead geometry was recorded as point clouds using a structured light scanner. For each melt pool image x_i , point cloud data within a $6 \text{ mm} \times 4 \text{ mm}$ rectangular window centered on its corresponding s_i and oriented with its 4 mm edge parallel to the print direction were extracted to fully capture the bead geometry while maintaining computational efficiency. These points were converted into a continuous surface mesh using natural neighbor interpolation [48], ensuring C^1 gradient continuity, preserving smoothness, and reducing noise in sparse regions. This method was preferred for its robustness to irregular point distributions and high-fidelity curvature representation.

The mesh was sliced orthogonally to the printing direction at S_i to obtain the bead's cross-sectional profile, aligned with the melt pool image, and uniformly sampled to 40 points to ensure sufficient spatial resolution (see more details in Fig. S1). This procedure is repeated for all spatial stamps S_0 to generate consistent profiles along the printed bead as:

$$Y_0 \in \mathbb{R}^{N_0 \times 40} \quad (8)$$

as shown in Fig. 5b.

To further study the width, height, and area of the cross-sectional

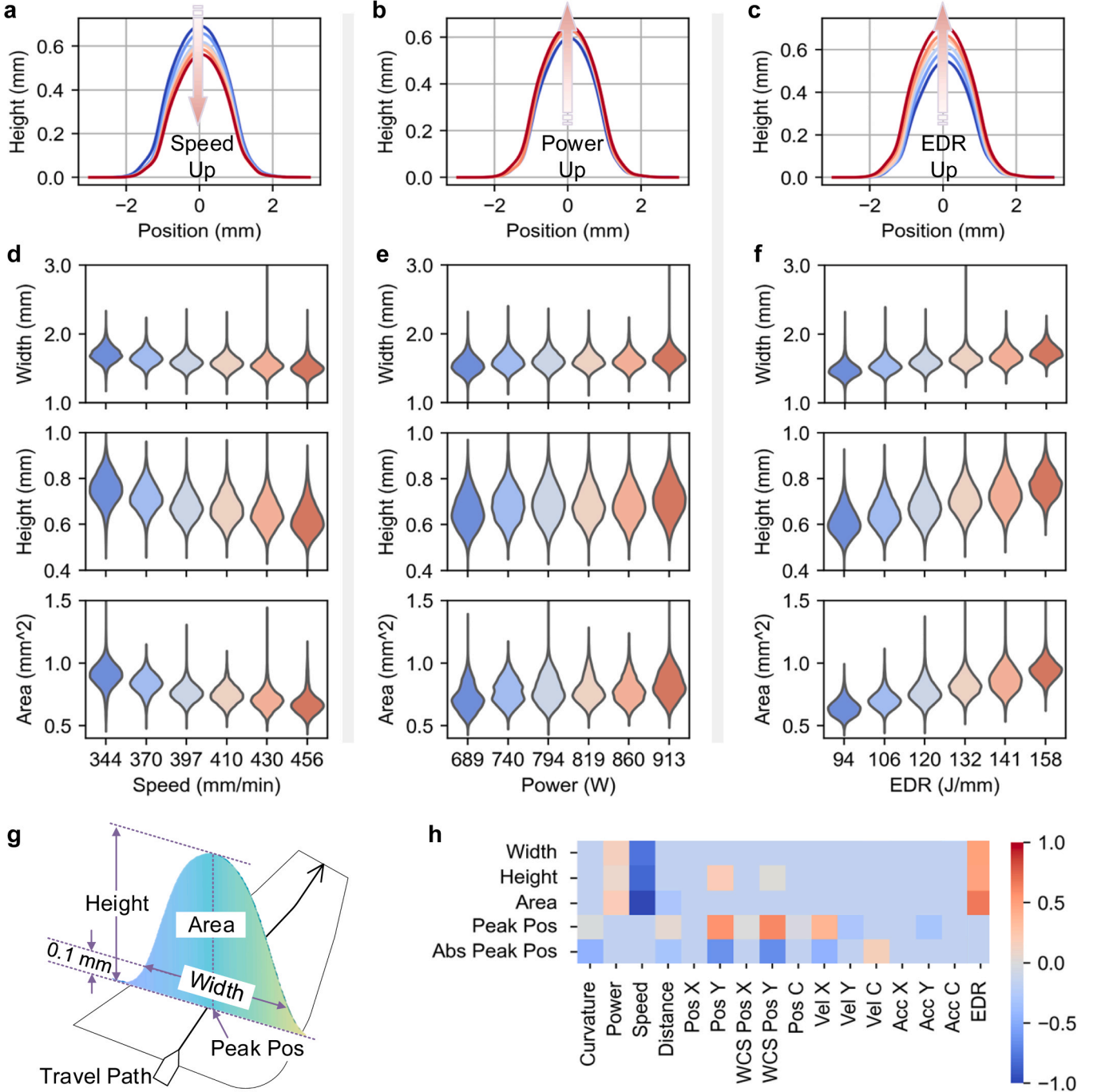


Fig. 6. Exploratory data analysis: (a-c) effects of travel speed, laser power, and energy deposition rate (EDR) on cross-sectional profiles; (d-f) effects of travel speed, laser power, and EDR on bead geometry features (width, height, and area), with the width of the violin plots representing data distribution; the filling colors of the violin plots in (d), (e), and (f) correspond to the lines in (a), (b), and (c), respectively; (g) schematic representation of cross-sectional width, height, and area; note that the travel path may not perfectly align with the peak position; and (h) correlation heatmap of control parameters, positional variables and geometry features, with values ranging from 1 (perfect positive correlation) to -1 (perfect negative correlation), and 0 indicating no correlation.

profiles (Fig. 6g), a non-mirrored sigmoid model σ' is utilized to fit the actual profile and compute the relevant parameters, which is defined as:

$$\sigma'(x) = \begin{cases} s \cdot \frac{1}{1 + e^{t_1(x-\omega+d_1)}} & (t_1, d_1 \leq 0) \quad x \leq \omega \\ s \cdot \frac{1}{1 + e^{t_2(x-\omega+d_2)}} & (t_2, d_2 \leq 0) \quad x > \omega \end{cases} \quad (9)$$

where s is the scaling parameter, t_1 and t_2 are the stretching parameters, ω is the central peak position, and d_1, d_2 are the shifting parameters. The width is defined as the distance across the profile at a height of 0.1 mm, the height corresponds to the maximum value of the function (Fig. 6g), and the area represents the region enclosed by the function and the horizontal axis within a specified range.

In conclusion, the collection of ST coordinates for the dataset is defined as the concatenation of the co-axial melt pool images, control parameters, positional variables, and the cross-sectional geometrical labels. This integrated dataset forms the basis for further analysis of the DED process and the printed bead geometry.

2.5. Exploratory data analysis

Exploratory data analysis (EDA) was conducted to investigate the influence of key control parameters—travel speed, laser power, and EDA—on the cross-sectional geometry of the deposited beads. The results revealed that while laser power and speed affect bead dimensions (width, height, area), the peak location is primarily influenced by the positional dynamics.

The violin plots in Fig. 6d-f summarize global trends across the dataset of 1.2 million samples, where the ‘violin’ shape represents data distribution, with wider sections indicating a higher density of data

points. The results indicate that the bead width, height, and area decrease significantly as speed increases, demonstrating a strong negative correlation. In contrast, laser power shows a weak positive correlation with geometry dimensions, leading to slight increases in width, height, and area. EDR, defined as the ratio of laser power over speed, exhibits a strong positive correlation, with all three geometric features increasing substantially as EDR rises. These trends are further validated by the average cross-sectional profiles in Fig. 6a-c, f where higher speeds result in narrower, smaller profiles, and higher EDR produces larger, more pronounced profiles. While larger laser power marginally increases the width and height, its effect is less pronounced. The correlation heatmap (Fig. 6h) reinforces these findings, confirming the strong negative correlation between speed and bead geometry, and the positive correlation with EDR. These results underscore the complex interplay of control parameters in shaping bead geometry, with EDR playing a dominant role in the DED process.

Additionally, the correlation heatmap (Fig. 6g) reveals the influence of positional variables, such as axis position, speed, and acceleration, on bead geometry, particularly affecting the lateral shift of the peak location during the printing process. This shift may be attributed to centrifugal force or the tilting angle of the laser head. Unlike width, height, and area, which describe the profile shape under ideal conditions, the peak location is not directly influenced by laser power or EDR. These findings highlight the role of positional factors in shaping bead geometry, particularly in non-ideal printing scenarios.

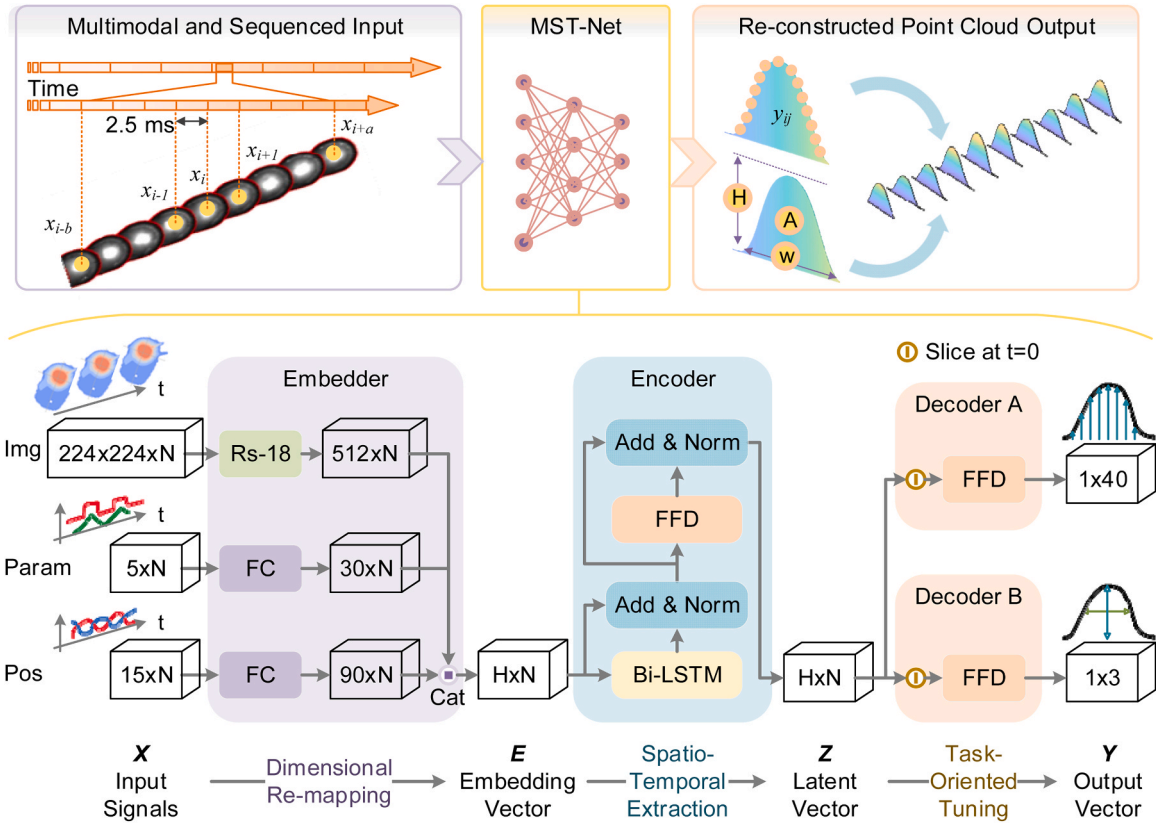


Fig. 7. The multimodal spatio-temporal neural network (MST-Net) architecture. (Img: Co-axial images, Param: Control parameters, Pos: Positional variables, Rs-18: Resnet-18 without the final linear layer, FC: Fully connected, Cat: Concatenate, Add: Residual addition, Norm: layer normalization, FFD: Feed-forward module with two linear layers and dropout, Bi-LSTM: Bidirectional long short-term memory, Decoder A: For cross-sectional geometric profile prediction, Decoder B: For cross-sectional width, height and area prediction).

3. MST-Net design, evaluation, and training

3.1. Neural network design

The neural network for cross-sectional geometric prediction first generates a joint representation of multimodal data (co-axial images, control parameters, and positional variables) [41] and then uses an encoder-decoder backbone with Bi-LSTM layers [40] to capture both spatial and temporal responses (Fig. 7). Here, the input is a sequence of hardware signals $\mathbf{X} = (x_{i-a}, \dots, x_i, \dots, x_{i+b})$ within a fixed time window centered at time instance i , where a and b represent number of elements after and before the current time instance i , respectively; and the output profile is $\mathbf{y} = (y_{i1}, \dots, y_{im}) \in \mathbb{R}^m$, where m is the number of task-specific features. The overall architecture, inspired by [42,49,50], consists of three modules: the embedding layer, the encoder, and the decoder.

3.1.1. Embedding layer

The embedding layer maps the input sequence $\mathbf{X}$ into a lower-dimensional matrix $\mathbf{E} = (b_{i-a}, \dots, b_i, \dots, b_{i+b}) \in \mathbb{R}^{N \times H}$, where $N = a + b + 1$ is the sequence length and H is the hidden dimension. Here, $H = 1 \cdot d^{img} + n^{prm} \cdot d^{prm} + n^{pos} \cdot d^{pos}$, where n is the number of signal channels, d is the feature dimension in each channel, and the superscripts *img*, *prm*, and *pos* represent the image, control parameters, and positional variables, respectively.

To incorporate both spatial and sequential information into the input, the elements in $\mathbf{E}$ concatenate features from all hardware channels:

$$e_i = \text{Concat}(e_i^{img}, e_i^{prm}, e_i^{pos}) \quad (10)$$

Following established multimodal approaches [41,51], the image embedding transforms the raw signals into low-dimensional vectors that capture essential spatial features while leveraging computational efficiency. Each image is mapped into $e_i^{img} \in \mathbb{R}^{d^{img}}$ using a pre-trained ResNet-18 model [28] as the feature extractor. During training, ResNet-18 parameters are frozen to preserve learned features and reduce computational overhead. The parameter and positional embeddings match the image embedding dimension, balancing input sizes to prevent the image dimension from overwhelming the model and hindering the learning. Each input parameter and positional signal is processed by a 1×1 convolutional layer, resulting in parametric embedding $e_i^{prm} \in \mathbb{R}^{n^{prm} \cdot d^{prm}}$ and positional embedding $e_i^{pos} \in \mathbb{R}^{n^{pos} \cdot d^{pos}}$. In this study, $d^{img} = 512$, and $d^{prm} = d^{pos} = 6$.

3.1.2. Encoder

The encoder captures both spatial and temporal dependencies in the data, transforming the input embedding $\mathbf{B}$ to a latent vector $\mathbf{Z} = (z_{i-a}, \dots, z_i, \dots, z_{i+b}) \in \mathbb{R}^{N \times H}$. The encoder comprises two sub-layers: a bi-directional long short-term memory (Bi-LSTM) layer for temporal extraction, and a fully-connected feed-forward (FFD) layer for feature transformation. To ensure training stability, residual connections and layer normalization are applied to both sub-layers.

The Bi-LSTM architecture [33,40,49] extracts spatio-temporal features. A Bi-LSTM processes sequential data bidirectionally by combining two LSTM layers: one processes the sequence backward, and the other processes it forward. This bidirectional design enables the model to integrate both past and future information, capturing dependencies throughout the entire sequence. This capability is critical for tasks requiring temporal continuity, such as analyzing thermal history and geometrical profiles. The Bi-LSTM in MST-Net uses a single layer with an input size of 632, which is chosen to match the output size of the preceding embedding layer. This configuration results in a hidden size of 316 units per direction, giving a bidirectional output dimension of 632.

The FFD layer [42] performs a temporal-wise transformation, applying the same and separate operation across different temporal

steps. It consists of two linear transformations with a non-linear activation in between:

$$x \leftarrow \max(0, x\mathbf{W}_1 + b_1)\mathbf{W}_2 + b_2 \quad (11)$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are weight matrices, b_1 and b_2 are bias vectors, and $\max(0, \cdot)$ represents the ReLU activation function. The non-linearity introduced by ReLU enables the model to capture complex relationships and enhance the features extracted by the Bi-LSTM.

To improve gradient flow and stabilize training, residual connections [52,53] are incorporated. These connections add the output of each sub-layer to its input, followed by the layer normalization [54], expressed as:

$$x \leftarrow \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (12)$$

This approach ensures dimensional consistency across the encoder, maintaining the same embedding size H . Residual connections address the vanishing gradient problem, while layer normalization helps standardize activations, facilitating stable and efficient learning.

3.1.3. Decoder

The decoder in our network architecture is modular, offering flexibility in generating various types of outputs. The latent vector, which encodes the learned representations of the input data, is used to produce the cross-sectional geometrical profile in the form of $y_i = (y_{i1}, \dots, y_{im}) \in \mathbb{R}^{m \times 1}$, where m is the number of points in each profile.

There are two decoder configurations, as shown in Table 3: the first directly generates cross-sectional profiles as slices of the 3D shape, providing a direct representation of the geometry. The second configuration generates geometrical features (e.g., width, height, and area) that can be fitted to a parametric function, such as a Gaussian shape or the non-mirrored sigmoid model as in Eq. (9), facilitating 3D shape reconstruction. This modular design also enables varying levels of output accuracy by changing m in the output profile.

This structure enables fine-tuning of the decoder without retraining the encoder or embedding layers, which comprise the majority of the model's parameters. As shown in Table 3, the decoder accounts for only 1.66–1.69 % of the total parameters. This facilitates efficient transfer learning, where the encoder remains fixed while the lightweight decoder adapts to new tasks with minimal computational cost.

3.2. Metrics for evaluation

The mean squared error (MSE) between the true and the predicted ones is used to evaluate the model performance:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of observations. To quantitatively compare the prediction accuracy of height, width, and area, the mean average precision (mAP) is used, as

Table 3

Model parameter sizes for two types of decoder configurations.

	Decoder A		Decoder B	
	Parameter Number	Parameter Pct (%)	Parameter Number	Parameter Pct (%)
Embedder	11,689,752	77.16	11,689,752	77.19
Resnet-18	11,689,512	77.16	11,689,512	77.18
Others	240	0.002	240	0.002
Encoder	3204,240	21.15	3204,240	21.16
Bi-LSTM	2401,600	15.85	2401,600	15.86
Others	802,640	5.30	802,640	5.30
Decoder	256,474	1.69	251,068	1.66
Total	15,150,466	100.00	15,145,060	100.00

given by:

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{y}_i}{y_i} \quad (14)$$

However, for profile prediction, mAP becomes unreliable due to the prevalence of data points close to zero, which can destabilize the evaluation. To overcome this, the Intersection over Union (IoU) metric is utilized for assessing the profile fitting accuracy, focusing on the overlap between the predicted and ground truth curves. IoU measures the degree of overlap between two regions (the predicted and ground truth curves), defined as the ratio of their intersection to their union:

$$\text{IoU}_i = \frac{\sum_{j=1}^n \min(f_i(j), g_i(j))}{\sum_{j=1}^n \max(f_i(j), g_i(j))} \quad (15)$$

where j ranges from 1 to n , representing each discrete point on the curve, and $f_i(j)$ and $g_i(j)$ denote the values at the j -th point of the predicted curve and the ground truth curve for the i -th pair, respectively. The mean Intersection of Union (mIoU) for a total of m curve pairs is calculated as:

$$\text{mIoU} = \frac{1}{m} \sum_{i=1}^m \text{IoU}_i \quad (16)$$

The mIoU metric is particularly well-suited for profile prediction due to its robustness against the non-uniformity of bead geometry, such as asymmetry, skewness, and changes in width at different heights. Additionally, it avoids instability caused by small denominators in percentage-based metrics, ensuring smoother training. Fig. 8 illustrates how IoU improves with the training progression for 4 samples. Each column represents the same sample at different epochs (1, 11, and 100), showing an increase in IoU from approximately 0.85 to 0.95. Notably, the cross-sectional profiles exhibit significant asymmetry, often skewing

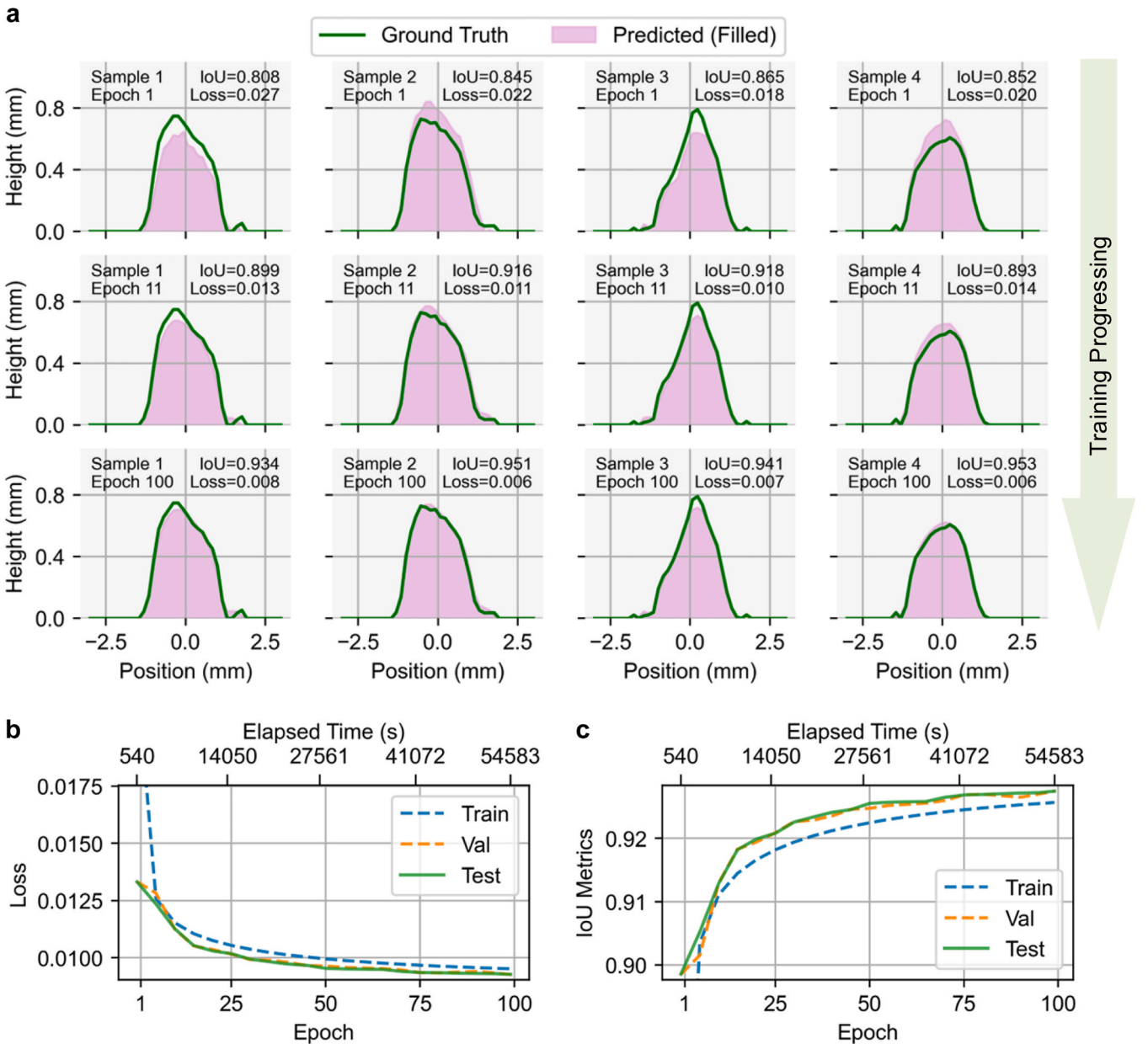


Fig. 8. Improvement of loss and metrics during training, validation, and test processes: (a) examples of improvement of the loss and the IoU (Intersection over Union) metric during training; (b) loss of train-val-test (training-validation-test) processes vs. epoch; and (c) IoU metrics of train-val-test processes vs. epoch.

to the left or right, which complicates the prediction and requires the model to be adaptable. Early in training, the predicted profiles (shaded regions) poorly match the ground truth (lines), resulting in lower IoU scores. As the training advances, the predictions align more closely with the ground truth, with higher IoU values indicating better accuracy. In general, IoU values around 0.85 are considered of moderate accuracy, while values around 0.9 are regarded as satisfactory, and values near 0.95 indicate highly precise predictions, showcasing the model's ability to converge effectively.

3.3. Model training

The training process for MST-Net followed a structured three-stage workflow: dataset preprocessing, train-val-test (training, validation, and testing), and deploying. Out of 68 datasets, 62 were used for train-val-test, while the remaining 6 datasets were reserved for deployment, as illustrated in Fig. 5d. This ensures the model's generalizability across different stages.

Datasets were pre-processed with techniques such as size cropping, feature standardization, and outlier removal to integrate images, control parameters, and positional variables into a cohesive feature space. Image data were cropped to 224×224 pixels and processed through ResNet-18 (excluding its final linear layer), serving as a feature extractor with its parameters frozen. This resulted in the extraction of 512 features from each image. Control parameters and positional variables were standardized using mean and standard calculated from the training data (Table 2) and then embedded into vectors of dimension 6 using two consecutive linear layers. Time-sequence data were then prepared, centered on the current instance, with additional time steps for both look-ahead and look-back, enabling the model to capture temporal variations.

In the train-val-test stage, each dataset was split into 80 % training, 10 % validation, and 10 % test sets to mitigate overfitting. Dropout regularization (with a rate of 0.3), residual connections, and layer normalization were applied in the encoder to stabilize training. The model was optimized using the Adam optimizer with an exponentially adaptive learning rate, starting at 0.4×10^{-5} and decaying by 0.985 over 100 epochs. A batch size of 64 was used, and a hybrid loss function combining MSE with weight of 0.88 and mIoU with weight of 0.12 balanced the prediction accuracy and profile fitting. As shown in Fig. 8b-c, the loss steadily decreased over epochs, while the IoU metric increased consistently across training, validation, and test sets, indicating enhanced generalization and stability as training progressed. Training proceeded for the full 100 epochs, and the best-performing model based on validation-set performance was saved and subsequently used in the test stage. In the deployment stage, the trained MST-Net model was evaluated on 6 datasets featuring previously unseen control parameter patterns (Fig. 5d). This three-stage process highlights the critical role of thorough data preprocessing, regularization techniques, and rigorous validation in optimizing the model's predictive capabilities, ensuring robustness and generalization to unseen data and diverse real-world conditions.

Model training and deploying were performed using PyTorch (version 2.3.0 +cu121) on a workstation with an AMD Ryzen Threadripper PRO 5955WX 16-core CPU, two NVIDIA GeForce RTX 4090 GPUs (24 GB of memory each), and Ubuntu 22.04.5 LTS (64-bit) as the operating system. All benchmarking architectures - including MST-Net and the competing baselines - were trained under the exact same hardware and software environment, using identical configuration of optimizer, learning rate, batch size, and number of epochs to ensure fairness. No additional regularization techniques such as weight decay, dropout, or early stopping were applied in any of the benchmark training runs (see Table S1 for details).

4. Results and discussion

4.1. Prediction performance

The prediction results in Fig. 9a provide a comparative analysis of model performance across 68 datasets, with experiment IDs 1, 2, 37, 38, 55, and 56 designated as deploying datasets and the remaining datasets used for training. Performance is evaluated using the mIoU metric for profile prediction, with the model achieving a maximum mIoU of 0.927 and an average mIoU of 0.92 on the training datasets, based on the evaluation using test partition. On the deploying datasets, the model yields a slightly reduced average mIoU from 0.89 to 0.91, indicating a modest drop in performance when generalizing to unknown data distributions. This 2–3 % reduction highlights the challenges of adapting the model to deployment scenarios, underscoring the importance of further tuning for robust real-world performance. In these tests, the control parameters featured a noise pattern that was absent from the training dataset, challenging the model to predict under new and varied conditions. A comparison of the actual printed bead geometry and the predicted geometry (Fig. 9b-d) reveals high similarity, with a root mean square error (RMSE) of 0.117 mm. Point cloud visualizations of a local area (Fig. 9e-g) confirm the fidelity of the spatial predictions, with the predicted point cloud profile being noticeably smoother and less noisy than the measured point cloud. These results affirm the model's robustness and its ability to generalize to new variations in control parameters.

Table 4 compares the performance of MST-Net with complete encoder configurations, considering both prediction accuracy (measured in mIoU) and computational efficiency (measured in frame per second, or FPS). Our analysis highlights the encoder's critical role as an ST extractor (Fig. 7), with fixed embedding and task-specific decoder layers to isolate encoder performance. The 1-layer Bi-LSTM (0.926 mIoU, 119 FPS) produced the closest performance to our MST-Net. Other temporal-aware models like the 1-layer RNN (0.916 mIoU, 143 FPS) marginally outperformed the MLP baseline (0.912 mIoU, 170 FPS) in accuracy, though deeper variants (2-layer Bi-LSTM: 0.873 mIoU, 92 FPS) suffered significant accuracy degradation despite increased parameters (e.g., 5.06 M for 2-layer Bi-LSTM vs. 2.4 M for 1-layer), likely due to overfitting. Self-attention and its MSHA variants achieved subpar accuracy despite high efficiency (~163 FPS). The 2-layer GNN matched the MLP's accuracy with 75 % fewer parameters (800 K) and higher efficiency (149 FPS). To conclude, MST-Net achieves the best accuracy (0.927 mIoU) with moderate computational overhead (116 FPS), demonstrating a favourable accuracy-efficiency trade-off for real-time DED control, where high precision is prioritized over marginal latency gains.

4.2. Temporal effects

The model input sequence length includes look-back, current instance, and look-ahead length (Fig. 10b), with temporal effects being crucial for performance. Fig. 10a's thermal map shows the melt pool's asymmetric shape, where the tailing length exceeds the heading length due to the travel speed of the melt pool. This asymmetry indicates the necessity for the model to weigh historical and future data differently. "Look-back" reflects previously deposited material's influence, while "look-ahead" anticipates process parameter changes but can introduce lags in real-time control. Results (Fig. 10d-e) indicate the best mIoU (0.927) occurs with total sequence lengths between 100 and 400, peaking at 200, corresponding to 5 mm in bead deposition at 10 mm/s. Optimal look-ahead lengths are 100–200, and look-back lengths 25–200. In Fig. 10c, we compare the distance between melt pool thermal maps for different sequence input lengths. As sequence length increases, the temporal separation between frames grows, weakening the spatio-temporal correlation between them. This increased distance reduces the model's ability to extract relevant features, leading to a performance

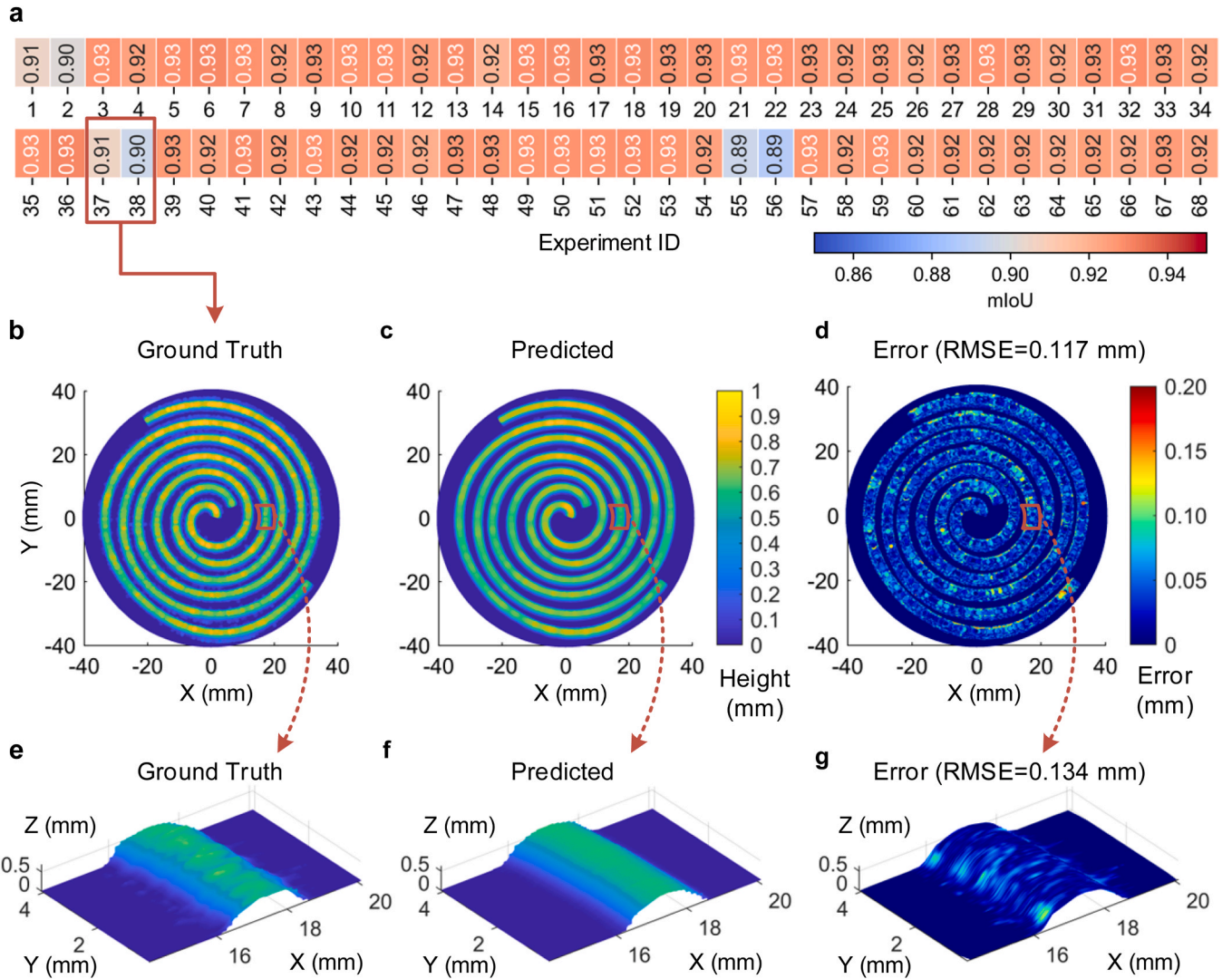


Fig. 9. Prediction results: (a) training and deploying results of 68 datasets, with experiment 1,2,37,38,55,56 used for deploying, and the remaining used for training; (b-d) deployment results for two complete paths (experiment 37 and 38) on the same substrate: (b) ground truth point cloud; (c) reconstructed bead geometry from the predicted profiles; and (d) prediction error map; (e-g) deployment results for a local region in experiment 38: (e) ground truth point cloud; (f) reconstructed bead geometry from the predicted profiles; and (g) prediction error map.

Table 4

Prediction performance of competitive encoder options. (The MLP, consisting of two fully-connected (FC) layers and one ReLU layer, uses only the current time instance. Other encoders have the target NN structure plus one ReLU layer, and process temporal sequences of length 201, with equal look-ahead and look-back lengths. The accuracy is measured on the test dataset. The efficiency, measured in frame per second, or FPS, is tested for the whole model including the encoders and other layers.).

Encoder Option	Encoder Parameter Size	Accuracy (mIoU)	Efficiency (FPS)
MLP (2 FC layers)	800,112	0.912	170
RNN (1 layer)	800,112	0.916	143
Bi-LSTM (1 layer)	2401,600	0.926	119
Bi-LSTM (2 layers)	5059,914	0.873	92
Self-Attention	1600,224	0.856	164
MHSA (4 heads)	1600,224	0.877	163
MHSA (8 heads)	1600,224	0.863	164
GNN (1 GCN layer)	400,056	0.895	156
GNN (2 GCN layers)	800,112	0.907	149
MST-Net (Ours)	3204,240	0.927	116

plateau for sequence lengths beyond 200. Sequence lengths longer than 300, exhibit diminishing returns in accuracy and higher computational costs (Fig. 10e-g). These findings highlight the need for a balanced approach that integrates sufficient historical data with future anticipation.

Aside from accuracy, latency is another key concern in real-time monitoring and control processes [45], primarily caused by three factors: 1) data processing, 2) control looping, and 3) signal communication. This discussion focuses on the first aspect. Univariate analysis (Fig. 10e) shows that using "look-ahead-only" achieves a mIoU score of 0.925, while "look-back-only" reaches 0.923. The "look-back-only" approach eliminates temporal lags by using only the current and past instances. While prediction accuracy is slightly lower, it emphasizes the model's real-time control capabilities. Fig. 10f-g shows that the computational time for MST-Net increases from 6.4 to 23.8 ms with the total sequence length increasing from 1 to 601. The increase is mainly due to the encoder, which scales linearly with sequence length, while the decoder and embedding layer's computation is unaffected, as embeddings are stored and reused across time steps. For an ideal sequence length of 201 that yields the best accuracy, the loop time for a "look-back-only" model is approximately 8.6 ms (~116 FPS), comparable to

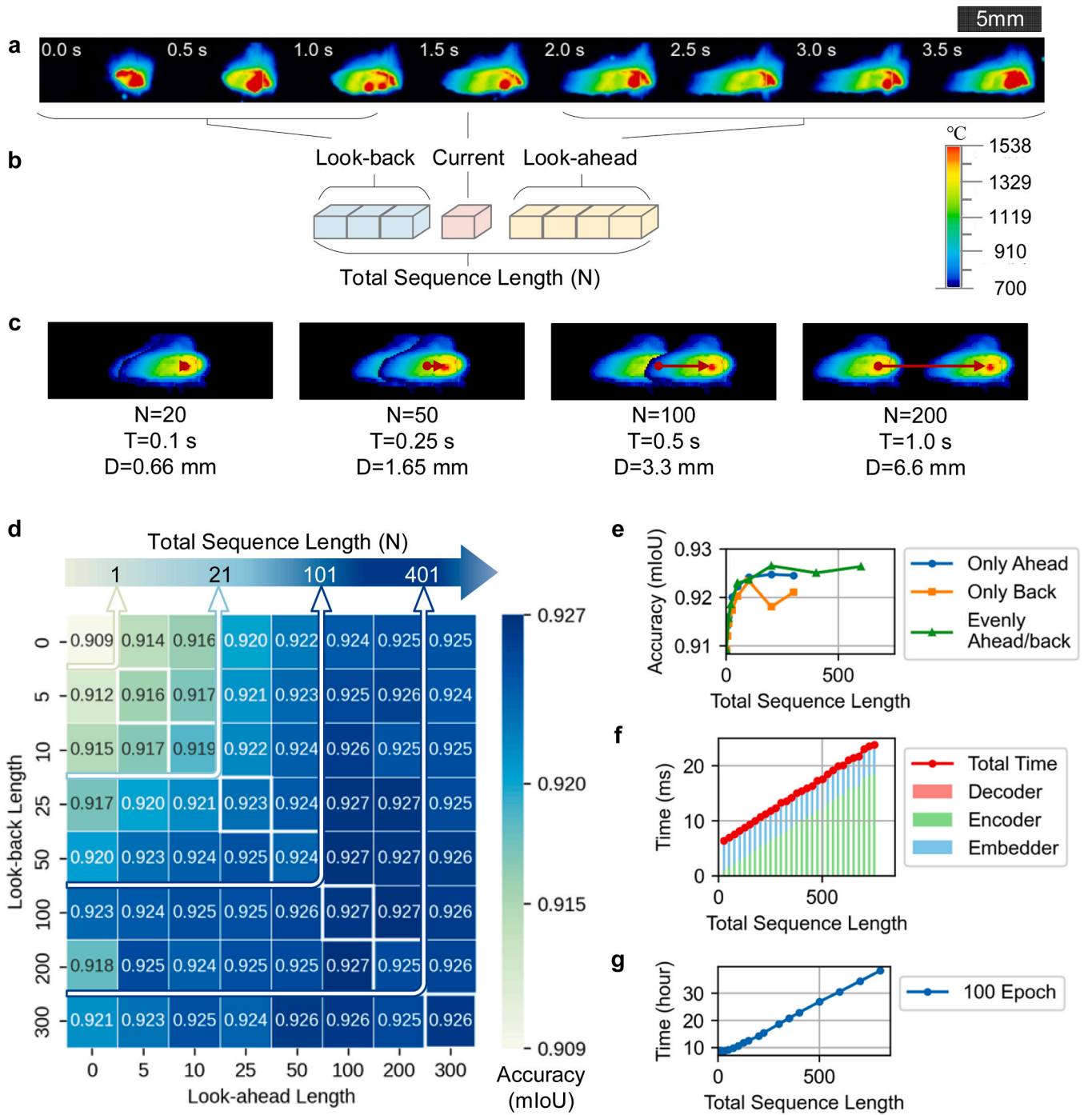


Fig. 10. Effects of look-ahead and look-back lengths on mIoU: (a) side-view thermal map of the evolving melt pool; (b) sequence composition of the model input; (c) comparison of distance between melt pool thermal maps for different sequence input lengths; (d) heatmap of mIoU with varying look-ahead and look-back lengths; (e) univariate analysis showing the cases of "only ahead", "only back", and "evenly ahead/back"; (f) computation time per iteration vs. total sequence length with 'only back' setup; and (g) training time vs. total sequence length, with batch size equalling 64.

state-of-the-art methods like [30], which uses 60 FPS for hardware-bead geometry synchronization. Of this, 5.5 ms (64 % of the time) is spent on embedding, as ResNet-18 accounts for over 77 % of the parameters (Table 3), 2.7 ms for the encoder, and 0.1 ms for the decoder. This processing speed is sufficiently fast for DED control, where control points are just 0.08 mm apart at a 10 mm/s travel speed. In conclusion, choosing a proper total sequence length for high accuracy, and using "look-back-only" for low latency is crucial for real-time DED control processes.

4.3. Prediction-based closed-loop control for printing height

Accurate prediction of bead geometry in DED is essential for effective process control, as it directly impacts the quality of the final workpiece. The 'look-back-only' model, with its real-time processing speed of 116 FPS, provides notable advantages for closed-loop control. To maintain temporal compatibility with this high-speed prediction capability, synchronous data acquisition is required. Before each deposition, control parameters—including RTCP state, rotary direction, laser power, travel speed, and curvature—are predefined with the toolpath, sampled at

0.1 mm intervals, and stored in memory. During printing, these parameters are retrieved in real time via a fast kd-tree search and combined with positional variables from the motion controller sampled at 1 kHz. Curvature and other parameters are continuously updated through lightweight interpolation between adjacent toolpath nodes, ensuring precise alignment with live axis coordinates. For a 10,000-point toolpath, the kd-tree query and interpolation complete in under 10 μ s, which is negligible within the 10 ms cycle time of a 100 Hz control loop. The synchronized control and positional data are then aligned with melt pool images captured at 200 Hz (as in Fig. 5a), and all modalities are pushed into a first-in-first-out (FIFO) queue of length 201—the optimal sequence length determined in Section 4.2. This streamlined data flow delivers temporally consistent, high-fidelity multimodal inputs at each control step, enabling stable and responsive closed-loop operation.

To assess the model's performance, open-loop and prediction-based closed-loop control printing experiments were conducted under identical conditions. As shown in Fig. 11a-b, a star-shaped conic pattern was selected for both experiments, following the same printing path trajectory. The designed structure comprised 20 deposited layers, yielding a target height of 10 mm. In the open-loop control setup, process parameters remained fixed throughout the printing process. In contrast, the prediction-based closed-loop control employed real-time multimodal data acquisition via sensors, which were fed into MST-Net to generate predictive bead geometry profiles (Fig. 11c). For simplicity, a proportional-integral-derivative (PID) controller was used to adjust the travel speed based on deviations between the predicted and target bead heights, enabling continuous compensation for height inaccuracies. The actual control frequency is 100 Hz, ensuring efficient and timely adjustments.

When comparing the open-loop and closed-loop control strategies, the open-loop system exhibited significant surface undulations, likely due to heterogeneous heat accumulation along the toolpath, particularly at the corners where path overlap occurs. This is a challenge that static process parameters cannot mitigate, as illustrated in Fig. 12a. On the other hand, the closed-loop control system produced a much smoother surface by dynamically adjusting the travel speed to compensate for layer height deviations, as shown in Fig. 12b. Top surface topography characterization was performed using a structured-light camera, and dimensional fidelity was evaluated against the designed model, as shown in Fig. 12c-d. The open-loop specimens demonstrated a mean error of 0.45 mm and a RMSE of 0.58 mm, whereas the prediction-based closed-loop control achieved higher accuracy with a mean error of 0.23 mm (45.0 % reduction) and RMSE of 0.30 mm (48.5 % reduction). These results underscore the efficacy of bead geometry prediction in optimizing printing process control, leading to a substantial

improvement in workpiece quality.

While our initial prediction-control framework has shown promising performance, we acknowledge two main limitations. First, it considers only single-layer parameters, which risks error propagation across layers. To address this, we are developing a laser-head-integrated height sensor for real-time measurement of deposited height, allowing predictive outputs to be corrected with ground-truth feedback. Second, the framework was trained on single-layer data but applied to multi-layer builds, where cumulative thermal history, melt pool dynamics, and support geometry effects become significant. To overcome this, we are expanding the dataset with in-situ multi-layer measurements, including synchronized point cloud-based height mapping (to be available later in a subsequent paper), enabling retraining of MST-Net to capture the physics of extended builds more accurately.

4.4. Transfer learning with variant tasks

Given the modular design of the decoder (Section 3.1.3), this study applies inductive transfer learning [55], where the input remains constant across both the source and target domains, but the output task differs, as shown in Fig. 13. By transferring parameters extracted by the encoder as high-level representations, the approach leverages knowledge from the source domain to enhance learning in the target domain, despite differences in task specifics.

We evaluate two setups in this study:

- In Setup I ($A \rightarrow B$), the ST encoder, pre-trained with Decoder A for profile prediction, is reused with Decoder B for WHA prediction.
- In Setup II ($B \rightarrow A$), the ST encoder, pre-trained with Decoder B, is reused with Decoder A for profile prediction.

As shown in Table 5, Setup I with 100 % of training data performs similarly to training without transfer learning (Decoder B), demonstrating effective knowledge transfer. However, Setup II with 100 % of training data experiences slight performance degradation, because WHA information is embedded in geometric profiles, but not vice versa. Notably, the transfer learning approach remains highly effective even with reduced target data. For Setup I, the model maintains strong performance even with only 10 % of the data, with mPA remaining stable at approximately 0.950. Setup II shows greater degradation as data size decreases, with mIoU dropping to 0.883 at 5 % data, reflecting the challenge of transferring knowledge from WHA to geometric profile prediction. As a result, the encoder in Setup I captures richer ST dynamics into the latent vector Z , making it the preferred base model for future task-specific transfers. With the decoder parameters comprising less than 1.7 % of the entire model (Table 3), this approach minimizes

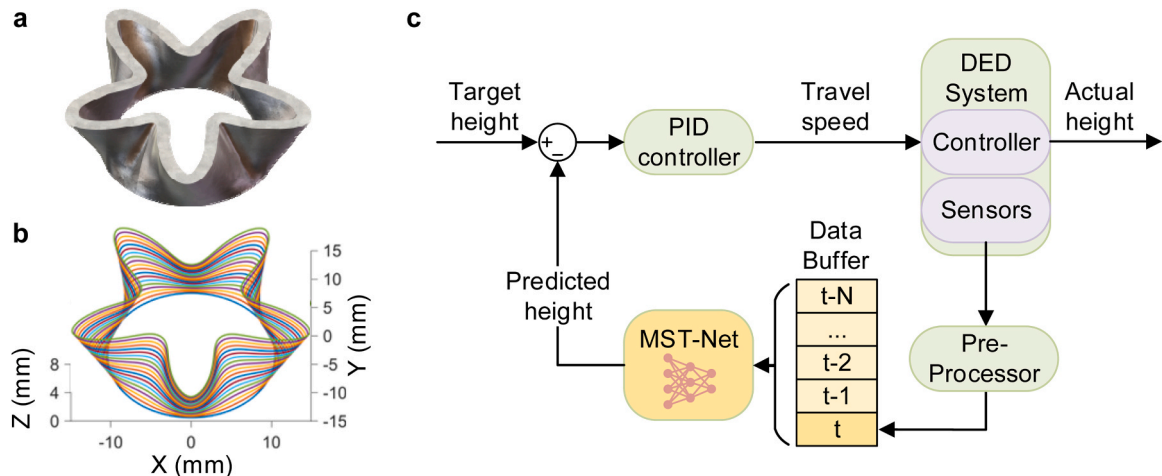


Fig. 11. Design of the closed-loop control experiment: (a) designed model; (b) printing trajectory; (c) closed-loop predictive control system integrating the MST-Net.

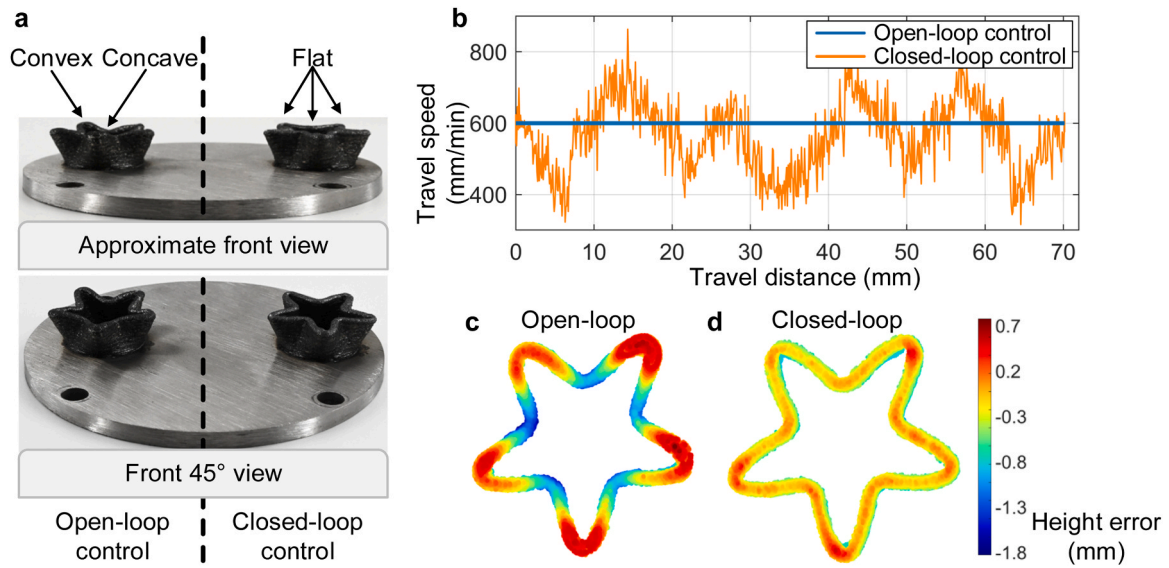


Fig. 12. Comparison between the open-loop and closed-loop control: (a) printed specimens; (b) travel speed; (c) height error heatmap comparing the printed top surface and the designed model.

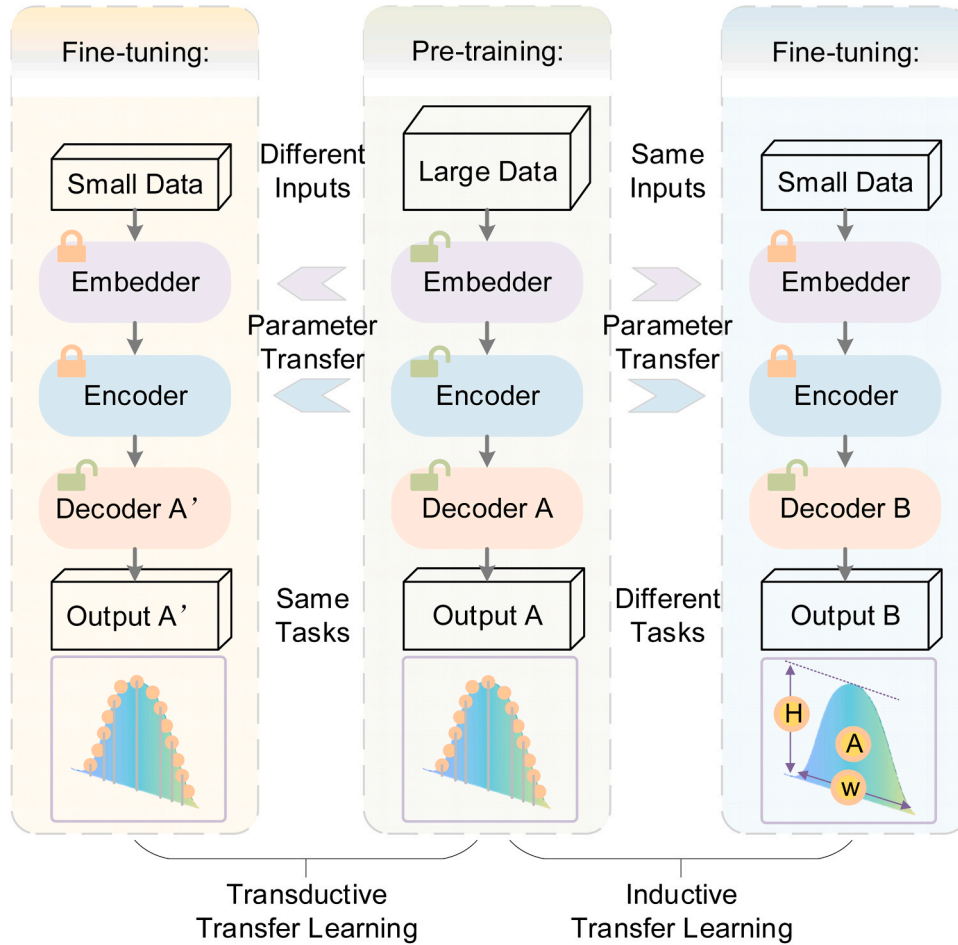


Fig. 13. Schematics of inductive and transductive transfer learning (Model parameters in the embedder and encoder are copied from the pre-trained base model, with only the decoder parameters being updated during parameter transfer).

Table 5

Comparison between direct training and transfer learning. The transfer learning has two setups: Setup I ($A \rightarrow B$), where the ST encoder pre-trained on profile prediction is used for WHA (Width, Height, and Area) prediction, and Setup II ($B \rightarrow A$), where the encoder pre-trained for WHA prediction is used for geometric profile prediction.

Direct Training	Decoder B (WHA)		Decoder A (Profile)	
	Loss	mPA	Loss	mIoU
100 % data	0.0095	0.951	0.0094	0.927
Transfer Learning	Setup I ($A \rightarrow B$)		Setup II ($B \rightarrow A$)	
	Loss	mPA	Loss	mIoU
100 % data	0.0096	0.952	0.0121	0.906
50 % data	0.0096	0.953	0.0126	0.904
25 % data	0.0102	0.953	0.0137	0.896
10 % data	0.0112	0.950	0.0152	0.886
5 % data	0.0164	0.933	0.0156	0.883

retraining needs and adapts efficiently to tasks with limited data, demonstrating both its robustness and adaptability.

4.5. Transfer learning with variant inputs

Additional transfer learning experiments were conducted to verify the model's adaptability to variant inputs, such as different patterns in control signals or printing paths. A typical example is the use of varying the path geometries as input. In addition to the equidistant spiral path

used in the pre-trained model, three new geometries (hexagon-like, decagon-like, and four-leaf clover-like) were employed in the transfer learning process (Fig. 14). In this case, transductive transfer learning [52] was applied, where the tasks remain the same, but the input data distribution differs, as shown in Fig. 10. The resultant error, evaluated by RMSE, ranged from 0.132 to 0.141 mm for these new paths, demonstrating that the MST-Net model successfully adapts to new geometries. In terms of computational performance, the training dataset size for the transfer learning experiments was only 0.017 million samples compared to the original 1.15 million. Owing to the much reduced size of the smaller dataset, its training time was under 20 min, as compared to the 8.5 h required for training on the full dataset. These results not only highlight the model's ability to generalize to new geometries but also emphasize its efficiency in terms of both training time and computational resources, further validating its robustness for real-world applications.

4.6. Multimodal data effects

An ablation study was conducted on the test dataset to systematically remove or alter specific inputs to assess their impact on the model performance, where a significant degradation in metrics indicates that the excluded input has a large effect. The results (Table 6) highlight the effects of various multimodal inputs (co-axial images, control parameters, and positional variables) on the model performance.

For decoder A (geometrical profile prediction), all input modalities are crucial, where the model achieves the highest mIoU (0.927) and the lowest loss (0.0094) in row #1 of Table 6. Excluding positional variables

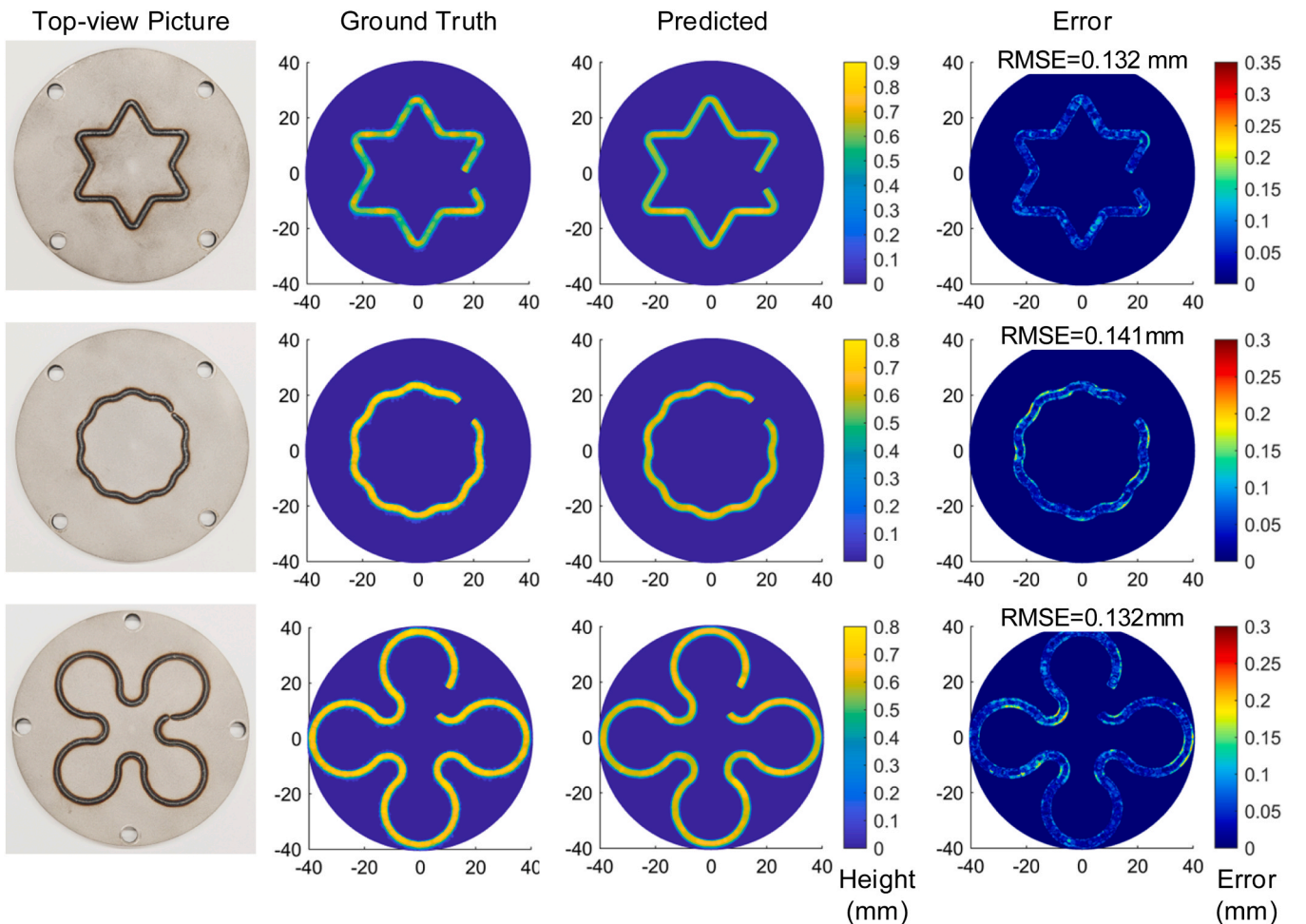


Fig. 14. Transfer learning results for new geometries: hexagon-like, decagon-like, and four-leaf clover-like shapes (top to bottom).

Table 6

Ablation study results with two decoder configurations: Decoder A for geometrical profile prediction and Decoder B for WHA prediction (width, height, and area). The study evaluates the impact of ablated modalities, including co-axial images, control parameters, and positional variables.

#	Co-axial Images	Control Parameters	Positional Variables	Decoder A (Profile)		Decoder B (WHA)	
				Loss	mIoU	Loss	mAP
1	Yes	Yes	Yes	0.0094	0.927	0.0095	0.951
2	No	Yes	Yes	0.0138	0.896	0.0573	0.866
3	Yes	No	Yes	0.0113	0.913	0.0167	0.933
4	Yes	Yes	No	0.0142	0.897	0.0131	0.944
5	Yes	No	No	0.0137	0.897	0.0152	0.939
6	No	Yes	No	0.0470	0.735	0.0244	0.918
7	No	No	Yes	0.0180	0.871	0.9024	0.766

results in a significant performance drop (row #4 of Table 6), with mIoU dropping to 0.897, as the peak location of the cross-sectional profile could shift due to positional and speed-wise variation of the CNC axes. Control parameters, while contributing to performance, appear to be largely embedded within the co-axial images, as evidenced by a smaller drop in performance when control parameters are ablated, but a large drop when co-axial images are ablated (row #2–3 of Table 6).

For decoder B (width, height, and area, or WHA prediction), the model's performance is primarily influenced by co-axial images, with minimal impact from positional variables. When both co-axial images and control parameters are included, the model achieves the highest mAP (0.951) and the lowest loss (0.0095). Excluding positional variables results in a small drop in performance, with the mAP decreasing to 0.944. Since width, height, and area measure overall shape rather than peak position, positional changes due to CNC axis movement have little effect on the prediction.

To illustrate why melt pool imagery emerges as the most influential modality in our ablation study, we analyzed three sets of key features: melt pool images (area, mean intensity, and center of mass (CoM) in X/Y), bead cross-sectional geometries (width, height, and area), and printing parameters (laser power, scan speed, and decomposed axial velocities (Vel X/Y)). Here, melt pool area is defined as the fraction of pixels above a very small normalized intensity threshold, mean intensity is the average normalized pixel value, and CoM represents the intensity-weighted centroid coordinates, indicating the position of the highest concentration of intensity. As shown in Fig. 15b and d, higher laser power and lower scan speed increase the EDR, which correlates positively with the melt pool mean intensity. In Fig. 15a, increasing EDR from 1.5 and 1.7–2.1 raises melt pool mean intensity from 0.073 and 0.081–0.093, accompanied by a larger bead cross-sectional area. For bead geometries, melt pool mean intensity exhibits the strongest correlation with bead area, followed by height and width, reflecting its stronger link to deposition volume, whereas the melt pool area primarily reflects the thermal boundary extent.

The melt pool morphology is also closely linked to the RTCP function: because the melt pool camera is mounted on the laser head (Fig. 2c), and the RTCP function alters the decomposition of axial velocities, positional dynamics directly affects the captured images in the camera frame. When RTCP is enabled, C-axis rotation dominates with minimal X-axis motion, and the melt pool tail consistently points downward regardless of the travel direction; when disabled, X–Y translation drives melt pool movement, and the tail direction shows a strong inverse relationship to motion direction. The correlation map in Fig. 15d shows strong negative correlations between melt pool CoM X and Vel Y, and a positive Vel Y–CoM Y correlation due to the opposite orientation of the motion coordinate system and the image Y-axis. These results confirm that melt pool images embed both physical and process-specific information that control parameters alone cannot provide, making them essential for high-accuracy geometry prediction in DED.

To visualize where the model focuses when interpreting melt pool images, we applied Gradient-weighted Class Activation Mapping (Grad-

CAM) [56] to a ResNet-18 backbone modified for single-channel melt pool images, as inspired by [57,58]. Grad-CAM was computed from the layer2 feature maps, producing activation maps that preserve fine spatial details while capturing higher-level semantic patterns. We note that the backbone was not fine-tuned on our dataset; therefore, we interpret Grad-CAM as a model-agnostic diagnostic rather than task-specific evidence. As shown in Fig. 15, the model's attention is primarily concentrated at the melt pool core, with the secondary focus along the edges and leading front, while the wake/tail region receives consistently lower attention. Spatter regions show localized but less dominant responses; for example, in the third row of Fig. 15a, spatter slightly increases local gradient values without shifting focus away from the core. These visualizations complement the ablation experiments by confirming that the model attends to physically meaningful regions (core, edges, leading front) and that this attention is influenced by variations in multi-axis motions.

More specifically, control parameters such as travel speed, laser power, and EDR have distinct impacts on the prediction accuracy, as shown in Fig. 16. In Fig. 16a, lower IoU values are associated with irregular cross-sectional profiles (blue curves), while smoother, Gaussian-like profiles (red curves) correspond to higher IoU values, indicating a better prediction accuracy. Correlation analyses in Fig. 16b reveal that IoU is negatively correlated with travel speed, indicating that higher speeds reduce the prediction accuracy. Conversely, IoU is positively correlated with the laser power and strongly influenced by EDR, with EDR emerging as the most significant control parameter, as demonstrated by the violin statistic plots in Fig. 16c–e. This is because EDR directly affects the cross-sectional area, aligning with the previous EDA analysis (Section 2.5). Besides, the peak location is primarily affected by the positional variables, such as axis position, speed, and acceleration. Statistical violin plots in Fig. 16f–h show that mIoU decreases with increasing axis C velocity, higher Y axis position, and greater absolute peak position. These findings emphasize the importance of fusing multimodal data in the prediction process.

5. Conclusion

In this study, we tackle data scarcity and dynamic geometry control in DED by proposing a multimodal spatio-temporal neural network (MST-Net) and a comprehensive, publicly available dataset. The dataset, comprising over 1.2 million multimodal samples, bridges a critical gap in studying the spatio-temporal dynamics of DED and supports data-driven approaches in additive manufacturing. We introduce MST-Net, which sets new benchmarks in DED predictions, successfully predicting irregular bead shapes and demonstrating strong transfer learning capabilities. Its modular architecture—embedding, encoder, and decoder modules—ensures robust performance and task-specific adaptations. Temporal analysis identifies an optimal 201-step sequence for accuracy at 116 FPS, while "look-back-only" minimizes latency, crucial for real-time DED control. Ablation studies emphasize the importance of multimodal data, with co-axial melt pool images being the most

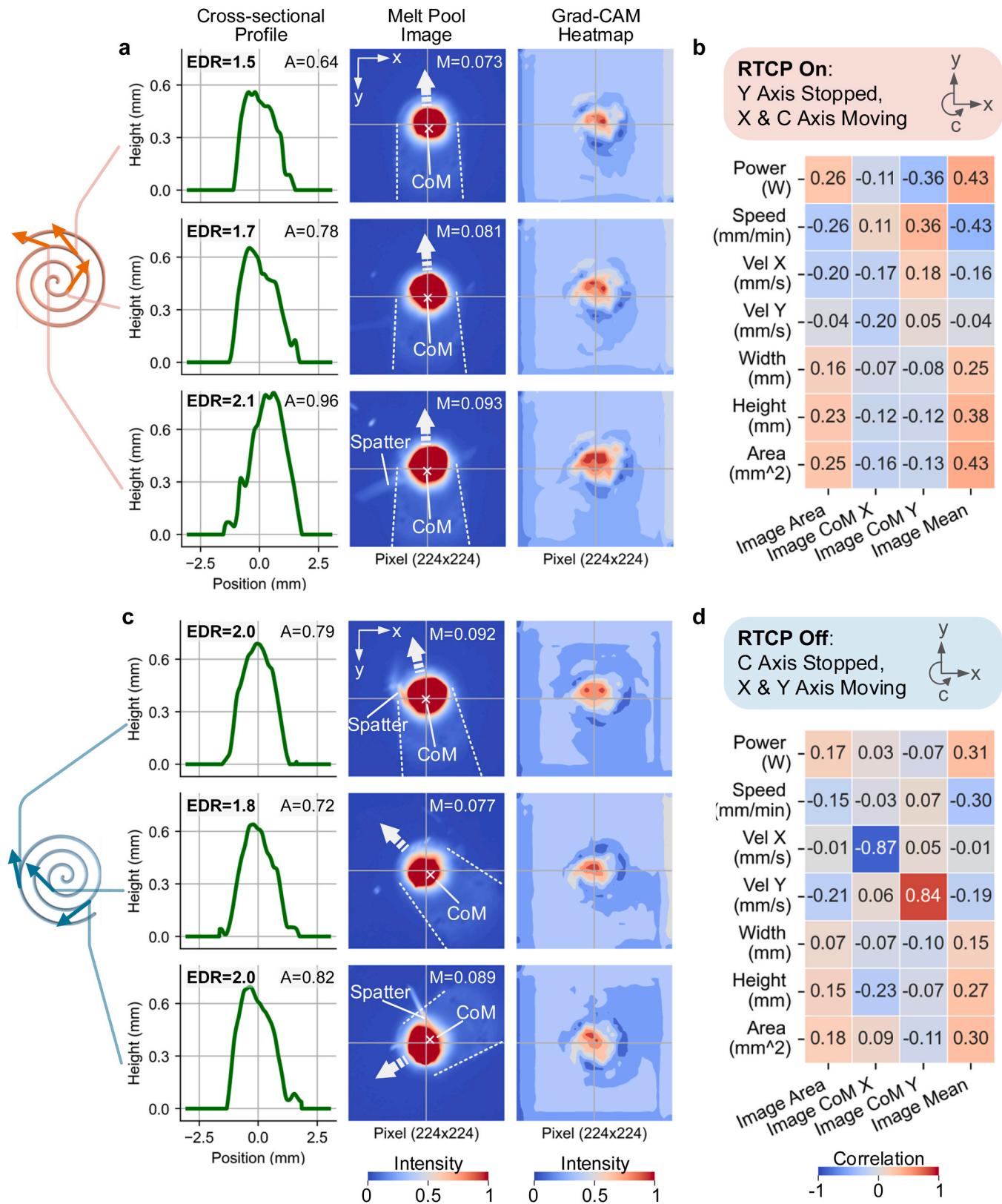


Fig. 15. Effects of melt pool images on prediction results with respect to positional and control parameters: (a) and (c) comparison of cross-sectional profiles, melt pool images, and Grad-CAM heatmaps with the RTCP function on and off, respectively (M: mean intensity, CoM: center of mass); (b) and (d) comparison of the correlations between positional/control parameters and statistical features of melt pool images, with the RTCP function on and off, respectively.

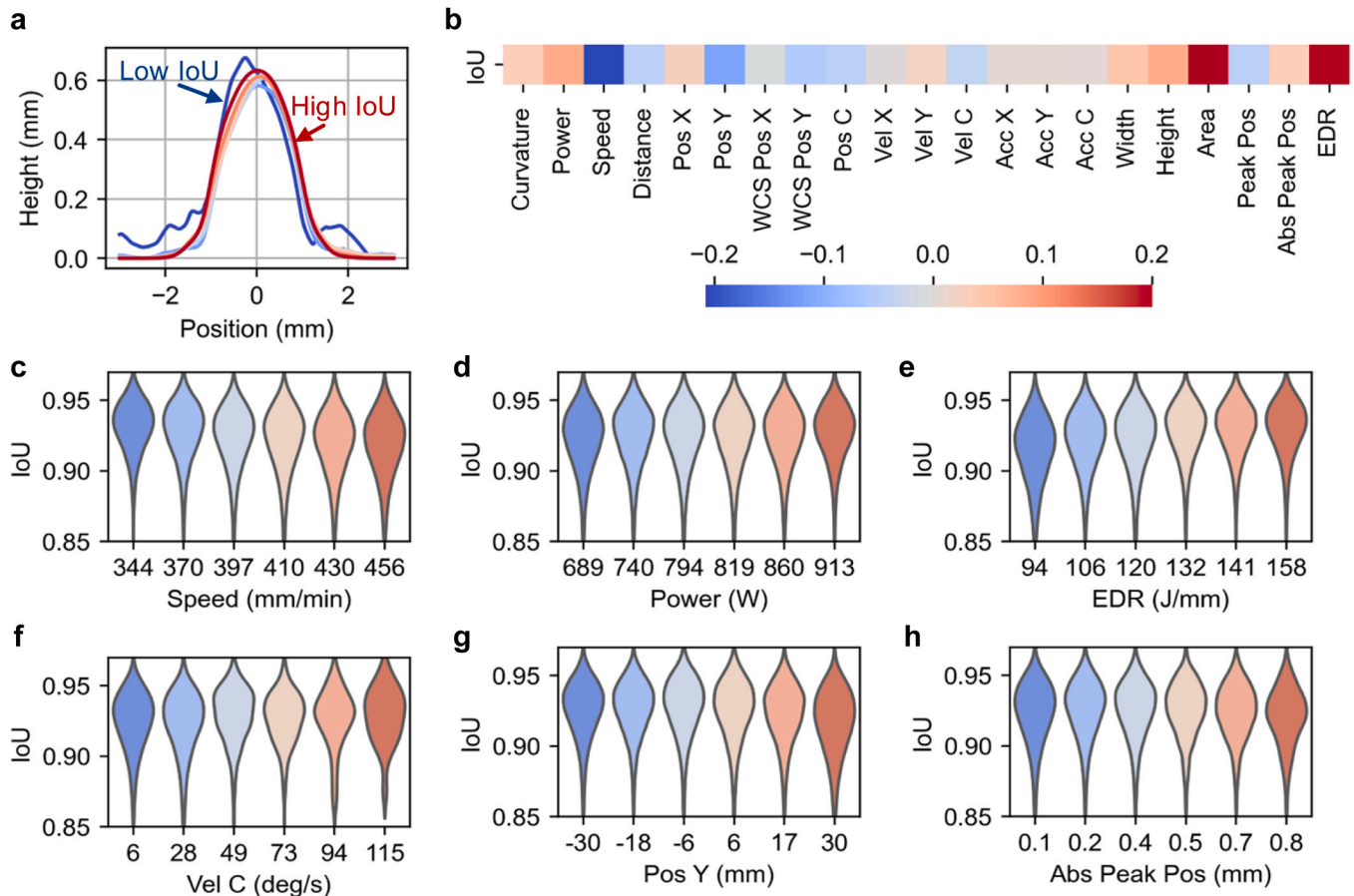


Fig. 16. Effect of control parameters and positional variables on prediction results: (a) mean cross-sectional profiles at different IoU levels; (b) correlation heatmap of input features and IoU metrics, with values ranging from 1 (perfect positive correlation) to -1 (perfect negative correlation), and 0 indicating no correlation; (c-e) effects of travel speed, laser power, and energy deposition rate (EDR) on the IoU metric; and (f-h) effects of C-axis speed, Y-axis position, and the absolute peak position on the prediction accuracy (IoU), with the width of the violin plots representing data distribution.

impactful. These advancements enhance DED modeling and control, providing a valuable tool for future industrial applications.

Despite these advancements, the study has certain limitations. It primarily focused on single-track deposition of a thin-shell structure, whereas multi-track applications could introduce more complex dynamics that were not addressed in this work. Furthermore, although the dataset was generated using a five-axis printing setup, the variation in the laser beam angle with respect to the deposition surface, which is inherent in five-axis printing, was not considered. In practice, additional factors such as slight ellipticity of the laser/powder spot and workpiece tilt under gravity can further affect melt pool shape. This work focuses on a simplified case without B-axis tilt or spot asymmetry, and these effects will be investigated in future studies. Finally, the controller model employed in this study, a proportional-integral-derivative (PID) controller, is relatively simple and could be enhanced with more advanced control models. Addressing these limitations in future research will be essential to improving the model's accuracy and robustness.

CRedit authorship contribution statement

Kai Tang: Writing – review & editing, Supervision, Funding acquisition. **Pengcheng Hu:** Writing – review & editing, Project administration. **Wenze Zhang:** Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Yuanzhi Chen:** Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Yichen Wang:** Writing – original draft, Visualization, Software, Methodology, Investigation,

Conceptualization. **Xiaoke Deng:** Investigation, Data curation. **Molong Duan:** Writing – review & editing, Supervision. **Yihe Wang:** Visualization, Formal analysis.

Declaration of Generative AI and AI-assisted technologies in the writing process

The writing process only uses Grammarly AI to perform grammar checks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported in part by (1) National Natural Science Foundation of China (grant No. 52375518); (2) Foshan HKUST Projects (Project ID: FSUST20-SRI09E).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.addma.2025.104952](https://doi.org/10.1016/j.addma.2025.104952).

Data Availability

Data will be made available on request.

References

- [1] J. Mazumder, Laser-aided direct metal deposition of metals and alloys. *Laser Addit. Manuf.* Elsevier, 2017, pp. 21–53, <https://doi.org/10.1016/B978-0-08-100433-3.00001-4>.
- [2] T. DebRoy, H.L. Wei, J.S. Zuback, T. Mukherjee, J.W. Elmer, J.O. Milewski, et al., Additive manufacturing of metallic components – process, structure and properties, *Prog. Mater. Sci.* 92 (2018) 112–224, <https://doi.org/10.1016/j.pmatsci.2017.10.001>.
- [3] D. Svetlizky, M. Das, B. Zheng, A.L. Vyatskikh, S. Bose, A. Bandyopadhyay, et al., Directed energy deposition (DED) additive manufacturing: physical characteristics, defects, challenges and applications, *Mater. Today* 49 (2021) 271–295, <https://doi.org/10.1016/j.mattod.2021.03.020>.
- [4] N. Shamsaei, A. Yadollahi, L. Bian, S.M. Thompson, An overview of direct laser deposition for additive manufacturing; part II: mechanical behavior, process parameter optimization and control, *Addit. Manuf.* 8 (2015) 12–35, <https://doi.org/10.1016/j.addma.2015.07.002>.
- [5] D. He, X. Deng, Y. Chen, W. Zhang, J. Hao, P. Hu, et al., Feedrate-preserved motion planning for five-axis directed energy deposition of freeform metal parts, *Addit. Manuf.* 89 (2024) 104302, <https://doi.org/10.1016/j.addma.2024.104302>.
- [6] D. Bi, M. Duan, T.Y. Lau, F. Xie, K. Tang, Strength-enhanced volume decomposition for multi-directional additive manufacturing, *Addit. Manuf.* 69 (2023) 103529, <https://doi.org/ARTN/2520103529/252010.1016/j.addma.2023.103529>.
- [7] F. Xie, L. Chen, Z. Li, K. Tang, Path smoothing and feed rate planning for robotic curved layer additive manufacturing, *Robot Comput. Integr. Manuf.* 65 (2020) 101967, <https://doi.org/10.1016/j.rcim.2020.101967>.
- [8] U.M. Dilberoglu, B. Gharehpapagh, U. Yaman, M. Dolen, Current trends and research opportunities in hybrid additive manufacturing, *Int. J. Adv. Manuf. Technol.* 113 (2021) 623–648, <https://doi.org/10.1007/s00170-021-06688-1>.
- [9] L. Zhu, S. Wang, H. Lu, D. Qi, D. Wang, Z. Yang, et al., Investigation on synergism between additive and subtractive manufacturing for curved thin-walled structure, *Virtual Phys. Prototyp.* 17 (2022) 220–238, <https://doi.org/10.1080/17452759.2022.2029009>.
- [10] T. Feldhausen, L. Heinrich, K. Saleeby, A. Burl, B. Post, E. MacDonald, et al., Review of Computer-Aided manufacturing (CAM) strategies for hybrid directed energy deposition, *Addit. Manuf.* 56 (2022) 102900, <https://doi.org/10.1016/j.addma.2022.102900>.
- [11] M. Gühr, A. Rashid, S.N. Melkote, Bead geometry prediction and optimization for corner structures in directed energy deposition using machine learning, *Addit. Manuf.* 84 (2024) 104080, <https://doi.org/10.1016/j.addma.2024.104080>.
- [12] J. Jiang, Y. Ma, Path planning strategies to optimize accuracy, quality, build time and material use in additive manufacturing: a review, *Micromachines* 11 (2020) 633, <https://doi.org/10.3390/mi11070633>.
- [13] D. Zhao, W. Guo, Shape and performance controlled advanced design for additive manufacturing: a review of slicing and path planning, *J. Manuf. Sci. Eng.* 142 (2020) 010801, <https://doi.org/10.1115/1.4045055>.
- [14] J. Xiong, G. Zhang, H. Gao, L. Wu, Modeling of bead section profile and overlapping beads with experimental validation for robotic GMAW-based rapid manufacturing, *Robot Comput. Integr. Manuf.* 29 (2013) 417–423, <https://doi.org/10.1016/j.rcim.2012.09.011>.
- [15] K. Ren, Y. Chew, N. Liu, Y.F. Zhang, J.Y.H. Fuh, G.J. Bi, Integrated numerical modelling and deep learning for multi-layer cube deposition planning in laser aided additive manufacturing, *Virtual Phys. Prototyp.* 16 (2021) 318–332, <https://doi.org/10.1080/17452759.2021.1922714>.
- [16] F.Q. Liu, L. Wei, S.Q. Shi, H.L. Wei, On the varieties of build features during multi-layer laser directed energy deposition, *Addit. Manuf.* 36 (2020) 101491, <https://doi.org/10.1016/j.addma.2020.101491>.
- [17] C. Wacker, M. Köhler, M. David, F. Aschersleben, F. Gabriel, J. Hensel, et al., Geometry and distortion prediction of multiple layers for wire arc additive manufacturing with artificial neural networks, *Appl. Sci.* 11 (2021) 4694, <https://doi.org/10.3390/app11104694>.
- [18] Z. Zhou, H. Shen, B. Liu, W. Du, J. Jin, Thermal field prediction for welding paths in multi-layer gas metal arc welding-based additive manufacturing: a machine learning approach, *J. Manuf. Process* 64 (2021) 960–971, <https://doi.org/10.1016/j.jmapro.2021.02.033>.
- [19] Z. Zhang, Z. Liu, D. Wu, Prediction of melt pool temperature in directed energy deposition using machine learning, *Addit. Manuf.* 37 (2021) 101692, <https://doi.org/10.1016/j.addma.2020.101692>.
- [20] E.W. Reutzel, A.R. Nassar, A survey of sensing and control systems for machine and process monitoring of directed energy, metal-based additive manufacturing, *Rapid Prototyp. J.* 21 (2015) 159–167, <https://doi.org/10.1108/RPJ-12-2014-0177>.
- [21] I.Z. Era, M.A. Farahani, T. Wuest, Z. Liu, Machine learning in directed energy deposition (DED) additive manufacturing: a state-of-the-art review, *Manuf. Lett.* 35 (2023) 689–700, <https://doi.org/10.1016/j.mfglet.2023.08.079>.
- [22] S. Hartmann, O. Murua, J.I. Arrizubieta, A. Lamikiz, P. Mayr, Digital twin of the laser-DED process based on a multiscale approach, *Simul. Model. Pr. Theory* 132 (2024) 102881, <https://doi.org/10.1016/j.simpat.2023.102881>.
- [23] Z. Sun, W. Guo, L. Li, Numerical modelling of heat transfer, mass transport and microstructure formation in a high deposition rate laser directed energy deposition process, *Addit. Manuf.* 33 (2020) 101175, <https://doi.org/10.1016/j.addma.2020.101175>.
- [24] C. Vundru, R. Singh, W. Yan, S. Karagadde, A comprehensive analytical-computational model of laser directed energy deposition to predict deposition geometry and integrity for sustainable repair, *Int. J. Mech. Sci.* 211 (2021) 106790.
- [25] S.M. Hashemi, S. Parvizi, H. Baghbanjavid, A.T.L. Tan, M. Nematollahi, A. Ramazani, et al., Computational modelling of process-structure-property-performance relationships in metal additive manufacturing: a review, *Int. Mater. Rev.* 67 (2022) 1–46, <https://doi.org/10.1080/09506608.2020.1868889>.
- [26] X. Qi, G. Chen, Y. Li, X. Cheng, C. Li, Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives, *Engineering* 5 (2019) 721–729, <https://doi.org/10.1016/j.eng.2019.04.012>.
- [27] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, *Atmos. Environ.* 32 (1998) 2627–2636, [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0).
- [28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR (2015) 770–778.
- [29] F. Dong, L. Kong, H. Wang, Y. Chen, X. Liang, Cross-section geometry prediction for laser metal deposition layer-based on multi-mode convolutional neural network and multi-sensor data fusion, *J. Manuf. Process* 108 (2023) 791–803, <https://doi.org/10.1016/j.jmapro.2023.11.036>.
- [30] R. Asadi, A. Queguineur, O. Wiikinkoski, H. Mokhtarian, T. Aihkialo, A. Revuelta, et al., Process monitoring by deep neural networks in directed energy deposition: CNN-based detection, segmentation, and statistical analysis of melt pools, *Robot Comput. Integr. Manuf.* 87 (2024) 102710, <https://doi.org/10.1016/j.rcim.2023.102710>.
- [31] M. Yu, L. Zhu, J. Ning, Z. Yang, Z. Jiang, L. Xu, et al., Machine vision and novel attention mechanism TCN for enhanced prediction of future deposition height in directed energy deposition, *Mech. Syst. Signal Process* 216 (2024) 111492, <https://doi.org/10.1016/j.ymssp.2024.111492>.
- [32] Y. Wang, J. Lu, Z. Zhao, W. Deng, J. Han, L. Bai, et al., Active disturbance rejection control of layer width in wire arc additive manufacturing based on deep learning, *J. Manuf. Process* 67 (2021) 364–375, <https://doi.org/10.1016/j.jmapro.2021.05.005>.
- [33] K. Hu, Y. Wang, W. Li, L. Wang, CNN-BiLSTM enabled prediction on molten pool width for thin-walled part fabrication using laser directed energy deposition, *J. Manuf. Process* 78 (2022) 32–45, <https://doi.org/10.1016/j.jmapro.2022.04.010>.
- [34] I. Jeon, L. Yang, K. Ryu, H. Sohn, Online melt pool depth estimation during directed energy deposition using coaxial infrared camera, laser line scanner, and artificial neural network, *Addit. Manuf.* 47 (2021) 102295, <https://doi.org/10.1016/j.addma.2021.102295>.
- [35] W. Ren, G. Wen, Z. Zhang, J. Mazumder, Quality monitoring in additive manufacturing using emission spectroscopy and unsupervised deep learning, *Mater. Manuf. Process* 37 (2022) 1339–1346, <https://doi.org/10.1080/10426914.2021.1906891>.
- [36] F. Kaji, H. Nguyen-Huu, A. Budhwani, J.A. Narayanan, M. Zimny, E. Toyserkani, A deep-learning-based in-situ surface anomaly detection methodology for laser directed energy deposition via powder feeding, *J. Manuf. Process* 81 (2022) 624–637, <https://doi.org/10.1016/j.jmapro.2022.06.046>.
- [37] X. He, T. Wang, K. Wu, H. Liu, Automatic defects detection and classification of low carbon steel WAAM products using improved remanence/magneto-optical imaging and cost-sensitive convolutional neural network, *Measurement* 173 (2021) 108633, <https://doi.org/10.1016/j.measurement.2020.108633>.
- [38] M. Mozaffar, S. Liao, H. Lin, K. Ehmman, J. Cao, Geometry-agnostic data-driven thermal modeling of additive manufacturing processes using graph neural networks, *Addit. Manuf.* 48 (2021) 102449, <https://doi.org/10.1016/j.addma.2021.102449>.
- [39] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997) 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [40] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, *IEEE Trans. Signal Process* 45 (1997) 2673–2681, <https://doi.org/10.1109/78.650093>.
- [41] W. Guo, J. Wang, S. Wang, Deep multimodal representation learning: a survey, *IEEE Access* 7 (2019) 63373–63394, <https://doi.org/10.1109/ACCESS.2019.2916887>.
- [42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention is all you need. *Proc. 31st Int. Conf. Neural Inf. Process. Syst. Curran Associates Inc, Red Hook, NY, USA, 2017*, pp. 6000–6010.
- [43] F. Scarselli, M. Gori, Chung Ah, Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, *IEEE Trans. Neural Netw.* 20 (2009) 61–80, <https://doi.org/10.1109/TNN.2008.2005605>.
- [44] Bruna J., Zaremba W., Szlam A., LeCun Y. Spectral Networks and Locally Connected Networks on Graphs. *CoRR* 2013;abs/1312.6203. <https://doi.org/10.48550/arXiv.1312.6203>.
- [45] F.Y. He, L. Yuan, H.C. Mu, M. Ros, D.H. Ding, Z.X. Pan, et al., Research and application of artificial intelligence techniques for wire arc additive manufacturing: a state-of-the-art review, *Robot Comput. Integr. Manuf.* 82 (2023) 102525, <https://doi.org/ARTN/2520102525/252010.1016/j.rcim.2023.102525>.
- [46] J. Gao, P. Li, Z. Chen, J. Zhang, A survey on deep learning for multimodal data fusion, *Neural Comput.* 32 (2020) 829–864, https://doi.org/10.1162/neco_a.01273.
- [47] X. Chen, M. Yao, F. Kong, Y. Fu, J. Wu, H. Zhang, In-situ quality monitoring of laser hot wire cladding process based on multi-sensing diagnosis and machine learning

- model, *J. Manuf. Process* 87 (2023) 183–198, <https://doi.org/10.1016/j.jmapro.2023.01.031>.
- [48] J. Xia, T. Ju, Variational surface reconstruction using natural neighbors, *ACM Trans. Graph* 44 (2025) 1–19, <https://doi.org/10.1145/3731191>.
- [49] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, et al., Long-Term recurrent convolutional networks for visual recognition and description, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 677–691, <https://doi.org/10.1109/TPAMI.2016.2599174>.
- [50] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks (2014), <https://doi.org/10.48550/arXiv.1409.3215>.
- [51] H.R. Vaezi Joze, A. Shaban, M.L. Iuzzolino, K. Koishida, MMTM: multimodal transfer module for CNN fusion. 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR, Seattle, WA, USA, IEEE, 2020, pp. 13286–13296, <https://doi.org/10.1109/CVPR42600.2020.01330>.
- [52] J.G. Zilly, R.K. Srivastava, J. Koutník, J. Schmidhuber, Recurrent highway networks (2017).
- [53] Zhang H., Goodfellow I., Metaxas D., Odena A. Self-attention generative adversarial networks 2019. <https://doi.org/10.48550/arXiv.1805.08318>.
- [54] Ba J.L., Kiros J.R., Hinton G.E. Layer normalization 2016.
- [55] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359, <https://doi.org/10.1109/Tkde.2009.191>.
- [56] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, *Int. J. Comput. Vis.* 128 (2020) 336–359, <https://doi.org/10.1007/s11263-019-01228-7>.
- [57] J. Yuan, H. Liu, W. Liu, F. Wang, S. Peng, A method for melt pool state monitoring in laser-based direct energy deposition based on DenseNet, *Measurement* 195 (2022) 111146, <https://doi.org/10.1016/j.measurement.2022.111146>.
- [58] E. Westphal, H. Seitz, A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks, *Addit. Manuf.* 41 (2021) 101965, <https://doi.org/10.1016/j.addma.2021.101965>.