EasyChair Preprint

№ 11476

Inference of Melt Pool Visual Characteristics in Laser Additive Manufacturing Using Acoustic Signal Features and Robotic Motion Data

Lequn Chen, Youxiang Chew, Wenhe Feng and Seung Ki Moon

December 7, 2023

Inference of Melt Pool Visual Characteristics in Laser Additive Manufacturing using Acoustic Signal Features and Robotic Motion Data

Lequn Chen *
Advanced Remanufacturing and Technology Centre (ARTC)
Nanyang Technological University (NTU)
Singapore
chen1470@e.ntu.edu.sg

Youxiang Chew
Advanced Remanufacturing and Technology Centre (ARTC)
Agency for Science, Technology and Research (A*STAR)
Singapore
chewyx@artc.a-star.edu.sg

Wenhe Feng
Singapore Instititue of Manufacturing Technology
Agency for Science, Technology and Research (A*STAR)
Singapore
fengwh@simtech.a-star.edu.sg

Seung Ki Moon
School of Mechanical and Aerospace Engineering (MAE)
Nanyang Technological University (NTU)
Singapore
skmoon@ntu.edu.sg

***Abstract*—Laser additive manufacturing (LAM) demands robust real-time monitoring to identify defects and ensure product quality. Traditional methods, mainly reliant on coaxial cameras, lacks placement flexibility. Further complexities arise from the inherent variability in melt pool geometries and the high computational demands of image processing, hindering effective real-time monitoring. This research proposes a novel technique to directly infer melt pool visual characteristics in LAM by synergizing acoustic signal features with robotic tool-center-point (TCP) motion data. Acoustic monitoring has shown great promise in tasks typically reliant on vision sensors. In addition, the dynamics of the LAM process and defect occurrences are spatially dependent, primarily due to heat accumulation. By combining acoustic signals with spatial data from robot TCP motion, our method tracks melt pool variations with an R^2 score above 0.7. An ablation study demonstrated that the proposed method outperforms the acoustic-only models. The findings suggest that the integration of a simple microphone sensor with robot motion information emerges as a flexible, cost-effective alternative for capturing dynamic melt pool behavior. It presents new prospects for closed-loop control in the LAM process.**

Keywords—laser additive manufacturing; real-time process monitoring; Robot Operating Systems; acoustic signal processing; machine learning

I. INTRODUCTION

Laser additive manufacturing (LAM) has revolutionized the production of complex metallic components for aerospace, automotive, marine and offshore industries [1]–[4]. Laser-directed energy deposition (LDED), a category of LAM, leverages on a focused laser beam to fuse metallic powder or wire onto a substrate, building parts layer by layer. Despite its versatility, LDED faces critical challenges in ensuring consistent quality and process stability [5]–[7]. The stochastic emergence of defects such as porosities and cracks caused by factors like heat accumulation and gas entrapment can severely compromise the mechanical integrity of the printed parts. As a result, early detection and correction of such abnormalities is critical to avoiding build failures and ensuring the quality of as-built product.

This research is funded by Agency for Science, Technology and Research (A*STAR) of Singapore.

To capture LDED process dynamics, the state-of-the-art in in-situ monitoring relies heavily on vision sensors, such as a coaxial visible spectrum camera. Melt pool visual characteristics (e.g., width, length, contour and size) can be utilized to assess the process stability. Such characteristics are frequently used as feedback information to adjust process parameters in a closed-loop manner to maintain quality consistency and microstructure homogeneity [8], [9]. The melt pool visual characteristics can also be used to classify lack-of-fusion (LoF) pores using supervised machine learning (ML) models [10]. However, coaxial cameras are normally embedded in the processing head and are not easily add onto existing LDED systems. To acquire robust and reliable melt pool images, additional near infrared (NIR) optical filters are necessary for visible light spectrum cameras. Real-time acquisition and processing of high-resolution melt pool images can be challenging in practice due to the high computational demands. On the other hand, as a supplement method, acoustic emission signals captured by a simple microphone can provide useful yet low-dimensional data insight into the laser-material interactions. Acoustic monitoring has demonstrated potential in in-situ monitoring and defect identification in LAM process [11], [12]. Liu et al. [13] proposed an approach that fuses acoustic and thermal data to infer melt pool morphology in the laser powder bed fusion (LPBF) process. Our previous work introduced a multisensor fusion-based digital twin (MFDT) framework [14]–[17], which extracted key features from visual, acoustic, and thermal data for localized quality prediction. The MFDT revealed that fusing data from different sensors improves defect detection accuracy and makes in-situ monitoring more reliable.

While prior research focused on multimodal sensor fusion, an inexpensive approach using a simple microphone sensor to control the process has yet to be achieved. To this end, this research introduces a novel approach using acoustic signal combined with spatial information from robot TCP motion to directly infer the key coaxial melt pool visual feature – melt pool width (MPW). MPW is the most widely used feature for closed-loop control applications in LDED. The proposed pipeline enables inference of real-time MPW with $\boldsymbol{R^2}$ score greater than 0.7 using only the acoustic signal and position, velocity data as

the input. The proposed technique outperformed the baseline model which uses only acoustic signals in an ablation study. Our work indicates the feasibility of utilizing a flexible and low-cost microphone sensor to characterize melt pool morphological dynamics. We believe it can contribute to further advancements in LAM closed-loop process control.

II. METHDOLOGY

In this section, the proposed framework for acoustic-based inference of melt pool characteristics is outlined, the data collection from robotic LAM system is introduced, and the proposed feature extraction, cross-modality correlation analysis and ML models are illustrated.

A. Proposed framework

The proposed data processing and melt pool inferencing framework is shown in Fig. 1. The proposed framework is designed to leverage both acoustic signal and robot motion data to infer the melt pool width (MPW). Capturing this synergy begins with the real-time acquisition of acoustic signals via a microphone sensor, collecting acoustic emissions from the laser-material interactions. From these signals, a comprehensive set of 83 acoustic features encompassing temporal, spectral, and cepstral data are extracted, offering a broad spectrum analysis. Simultaneously, robot tool-centre-point (TCP) motion is extracted in real-time from the KUKA robot controller through the Robot Operating System (ROS). This includes real-time coordinates (x, y, z) and velocity metrics, essential for providing a spatial context to the acoustic signals. This data enables a better interpretation that correlates to the dynamic melt pool characteristics. Ground truth validation is provided by visual monitoring of the melt pool, facilitated by a coaxial visible spectrum camera. Key visual characteristics of the melt pool, such as MPW, are extracted using image processing techniques via OpenCV. These characteristics serve as the ground truth for ML regression model training.

In the inferencing stage, the extracted acoustic features and the TCP motion data are synthesized through ML algorithms. This cross-modality fusion forms the basis for the predictive model that estimates MPW, aiming to minimize the reliance on direct visual observation.

B. System Setups and Dataset Collections

The experimental setup, illustrated in Fig.2, features a dual-robot hybrid additive-subtractive manufacturing system. For the LDED process (Fig. 2(b)), a six-axis robotic arm carried a coaxial powder-blowing nozzle to deposit material onto the substrate held by a two-axis positioner [18]. The microphone was placed near the laser nozzle to capture the acoustic signature of the LDED process sound, with noise cancellation techniques previously addressed in [11]. Detailed analyses of the LDED sound composition and feature extraction of acoustic signals are discussed in our previous work [12]. Additionally, a coaxial CCD camera equipped with a near-infrared (NIR) bandpass filter ranging from 720 to 1100 nm was used to monitor the melt pool morphologies.

Synchronization of visual and acoustic data streams was achieved via an in-house developed ROS-based software platform. The details of multi-sensor data synchronization and the subsequent feature extraction are elaborated in our preliminary works [15], [16]. The collected multimodal dataset includes time-aligned acoustic emissions, robot TCP motion data (comprising coordinates and velocity), and the coaxial melt pool images. This multimodal dataset was created by fabricating different single bead walls with MS C300 powder and varying process parameters, as seen in Table 1. Further details on experimental procedures were reported in previous work [15], [16].

TABLE I. LDED PROCESS PARAMETERS USED FOR DATA COLLECTION

Process Parameters	Values
Laser power (kW)	2.3 – 2.53
Scanning speed (mm/s)	25 – 27.5
Dwell time between layers (seconds)	[0, 5, 10]
Powder feeding rate (g/min)	12

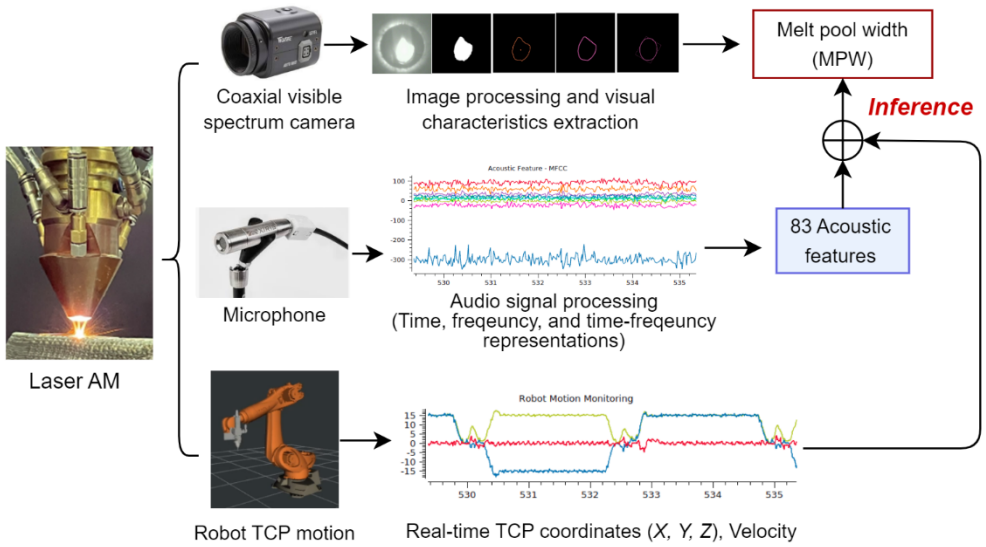

Fig. 1. Proposed framework for inferencing melt pool width using audio signal and spatio-temporal robot motion data.

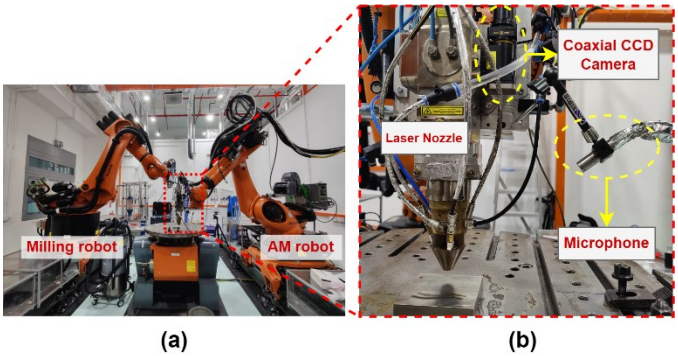

Fig. 2. Photos of experiment and sensor setup.

C. Feature extraction and signal visualization

This study utilizes Essentia library [19] for audio feature extraction, distilling 83 acoustic signatures in time, frequency, and time-frequency domains. For capturing coaxial visual characteristics, the melt pool is approximated as an elliptical shape using the OpenCV library [20]. The minor axis of this ellipse is used to represent the MPW.

Fig. 3 presents synchronized audio-visual signals during the LDED process under various conditions, including defect-free and defective scenarios. It demonstrates a correlation between anomalies in audio and visual signals with physical defects. For example, the occurrence of cracks is signified by a noticeable increase in acoustic amplitude accompanied by changes in the melt pool brightness, as shown in Fig. 3(a). In contrast, a defect-free process exhibits stable and uniform audio-visual signals. The keyhole pore printing mode, as evident in Fig. 3(b), is characterized by irregularly enlarged melt pools and elevated acoustic signal amplitudes. In a 'laser-off' state, the absence of the melt pool is clearly visible, while the acoustic signal captures only ambient sounds.

Fig. 4 displays optical microscope (OM) imaging of defects in 3D printed parts, correlating with anomalies in the multimodal signals. This correlation confirmed that print quality was spatially dependent. For instance, during the printing of the single bead wall structure, a shift from defect-free to defective modes was observed, primarily due to heat accumulation. This transition led to keyhole porosity in the upper layers. Furthermore, variations in audio-visual signal features were linked to the height of the parts. Robot TCP movements, particularly acceleration and deceleration at the beginning and end of deposition tracks, contributed to melt pool fluctuations. In idle states, when the robot remained stationary (velocity = 0), the laser was typically off. As the part height increased, defects became more prevalent due to heat accumulation, causing destabilization and enlargement of the melt pool. These observations suggested that integrating real-time robot TCP coordinates (x, y, z) and velocities with acoustic signals could enhance the accuracy of MPW prediction.

The discernible relationship between acoustic signals and spatial variations in melt pool characteristics underscores the potential of acoustic analysis combined with spatial data for MPW prediction. The subsequent section will provide a detailed quantitative analysis of these correlations, further elucidating the audio-visual cross-modal connections.

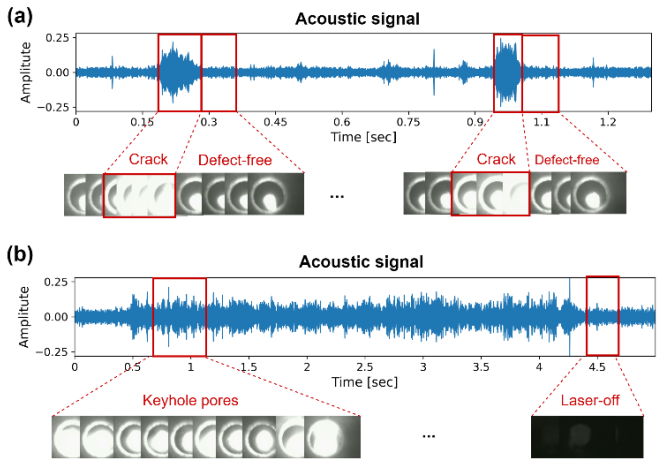

Fig. 3. Audio-visual signal temporal synchronization and visualization.

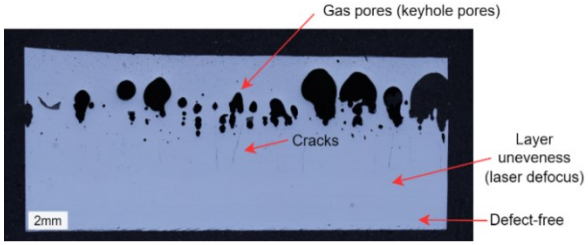

Fig. 4. Defects in the 3D printed parts observed from optical microscope (OM) image.

D. Cross-modality feature correlation analysis

Fig. 5 presents the Spearman correlation heatmap that elucidates the relationship between acoustic features and melt pool visual characteristics (MPW). Spearman's rank correlation coefficient ρ is a non-parametric measure of correlation, representing statistical dependence between of two variables. It can be calculated as:

$$\rho = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)} \quad (1)$$

where d_i is the difference between the ranks of corresponding variables, and n is the number of data points. The heatmap reveals that, among the top 12 acoustic features, "*spectral_flux_mean*" and "*spectral_valley_2_mean*" exhibit a moderately strong correlation (> 0.5) with the target MPW.

Spectral flux is a measure of the variability of the spectrum over time. It can reflect changes in the acoustic signal's frequency content. A higher spectral flux indicates significant variations in these signals. In the context of LDED, this implies that fluctuations in acoustic emissions are directly linked to changes in melt pool dimensions. Similarly, the spectral valley represents specific points in the spectrum where the energy is at a minimum, and its mean value over time provides insights into the consistency of these low-energy occurrences.

The positive correlation of these features with MPW suggests that both the variability captured by spectral flux and the consistency of low-energy points indicated by spectral valleys are indicative of the melt pool's behavior. Particularly, in a defective regime, high fluctuations in acoustic signals correspond to unstable or irregular MPW. This relationship

underscores the potential of acoustic monitoring as a diagnostic tool in detecting anomalies or variations in the melt pool, thereby offering a pathway for control and quality assurance in LAM processes.

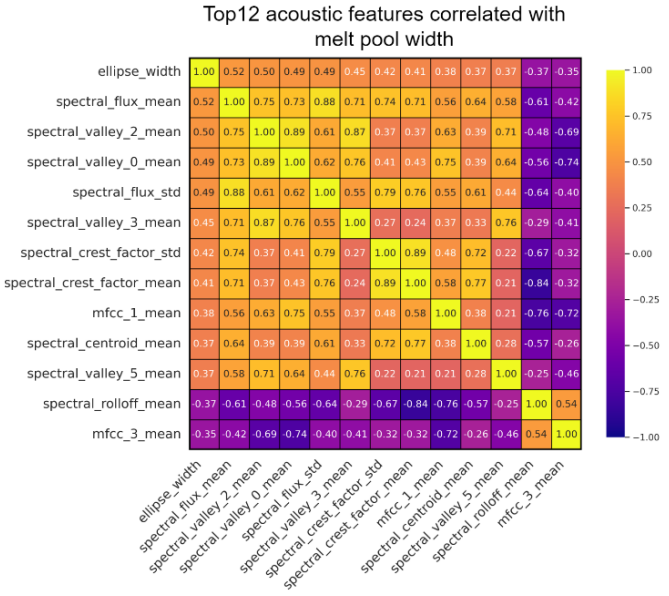

Fig. 5. Top 12 acoustic features correlated with melt pool visual characteristics (i.e., melt pool width approximated by an ellipse shape).

E. Machine learning models for melt pool width prediction

Following the qualitative and quantitative analysis of the correlation between acoustic signals, robot TCP position data, and melt pool width (MPW), various ML models were explored for MPW prediction. Selected for their ability to handle complex, non-linear relationships, these models leverage the complex relationships identified in the data to provide robust MPW regression prediction. The employed machine learning algorithms include:

- *Random Forest Regressor (RF)*: The Random Forest Regressor [21] aims to minimize the mean squared error by aggregating predictions from individual decision trees. The objective function is as follows:

$$\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (2)$$

where n is number of data, $\hat{y}_i$ is the predicted value and y_i is ground truth. This averaging approach helps to improve accuracy and control over-fitting.

- *XGBoost Regressor (XGBRegressor)*: XGBoost stands for eXtreme Gradient Boosting [22], a scalable and accurate implementation of gradient boosting machines. It builds the model in a stage-wise manner, where each new model incrementally reduces the loss function (usually a mean squared error (MSE)) using the gradient descent algorithm. It minimizes the following objective function:

$$\min \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (3)$$

where L is a loss function (MSE), $\hat{y}_i$ is the predicted value, y_i is the ground truth, K is number of trees, and Ω is a regularization term.

- *LightGBM Regressor (LGBMRegressor)*: LightGBM is a gradient boosting framework that uses tree-based learning algorithms [23]. It is characterized by its efficiency in handling large datasets and high speed. LightGBM grows tree leaf-wise (vertically) while other algorithms grow level-wise (horizontally), leading to faster convergence. The model is built by iteratively choosing the leaf with max delta loss to grow. It attempts to minimize the following objective:

$$min \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{m} \lambda \left|\left|\omega_j\right|\right|^2 \qquad (3)$$

where l is a differentiable convex loss function measuring the difference between ground truth and predicted values, and λ is the regularization parameter. $\left|\left|\omega_j\right|\right|^2$ represents L_2 regularization term for weights of the leaves.

These algorithms were selected for their effectiveness in handling complex, non-linear relationships in data, making them suitable for the complex task of MPW prediction.

III. Results and Discussions

A. Regression model performance evaluation

The evaluation of regression model performance is critical in proving the effectiveness of models in MPW prediction. Fig. 6 illustrates this by plotting predicted values against ground truth data.

The RF using a combination of acoustic signals and robot TCP motion data (Fig. 6(b)) emerged as the superior model. It achieved the highest $\boldsymbol{R^2}$ score of 0.73, indicating a strong positive correlation between the predicted and actual values. This model also yielded the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE), suggesting it has the best predictive accuracy and is most consistent with the ground truth.

A comparative analysis across all models reveals a clear trend: models trained on combined acoustic and robot TCP motion data outperform those relying solely on acoustic signals. For instance, when augmented with robot motion data, the XGBoost Regressor (XGBRegressor) and LightGBM Regressor (LGBMRegressor) exhibit a significant boost in performance (about 20% and 15% in $\boldsymbol{R^2}$ score, respectively) compared to their acoustic-only counterparts (Figs. 6(c) and 6(e)).

The evaluation confirms that integrating spatiotemporal robot TCP motion data with acoustic features improves model predictions for MPW. The next sub-section will visualize the ground truth MPW data with the predicted values for different printing scenarios.

B. Inference of melt pool visual characteristics

Fig. 7 presents a detailed view of the RF models' performance in capturing the dynamics of melt pool visual characteristics (i.e., MPW) over time for different LDED test samples. Each graph depicts the ground truth data against the predicted MPW values, demonstrating the models' performance

with and without additional robot TCP motion information across varying scenarios.

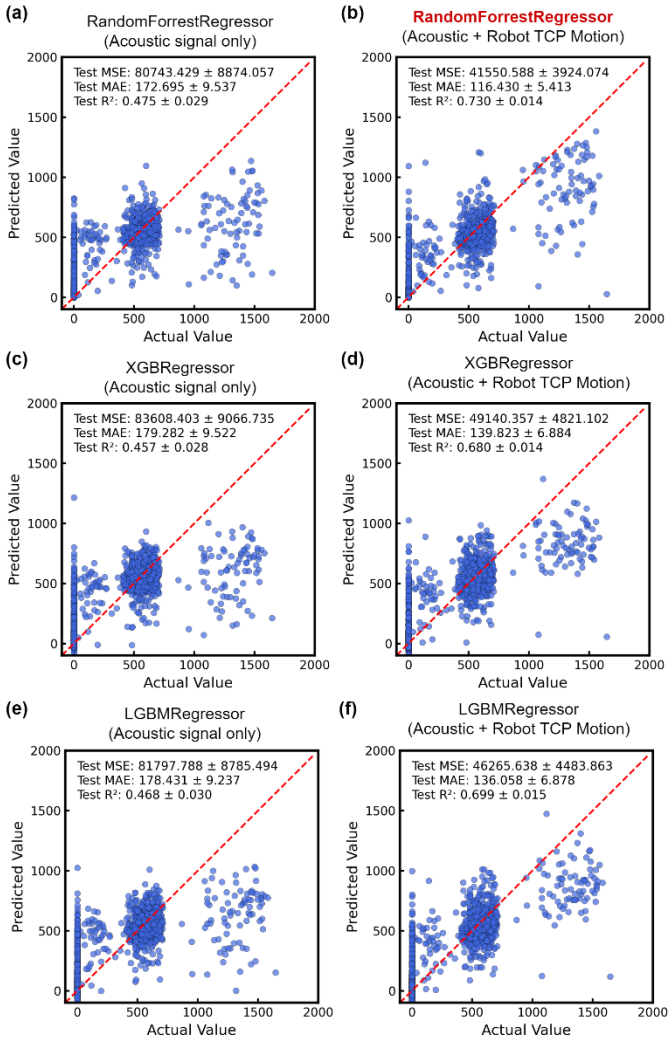

Fig. 6. Ground truth data plotted against predicted data points for melt pool width prediction. (Note: the higher R^2 score indicates better performance, while a lower MAE and MSE score indicates better performance.).

In the single-bead wall fabrication with 25 layers and without interlayer dwell (Fig. 7(a)), the actual MPW fluctuated significantly. The MPW became increasingly unstable over time due to heat accumulation and laser defocusing. The localized RF model (acoustic + robot TCP motion) prediction closely followed the actual MPW trend, indicating the model's responsiveness to changes in the process dynamics. The acoustic-only model, on the other hand, exhibited larger variations in predictions. However, both models tended to underestimate the MPW fluctuations in the higher layers (time > 40s). The initial five layers (0 – 20s) of the same single-bead wall fabrication (Fig. 7(b)) showed a more detailed view. The localized RF model predictions aligned better with the actual MPW values than the standard model trained only on acoustic feature.

When an interlayer dwell time of 5s was introduced (Fig. 7(c)), the actual MPW became zero during the laser-off period. Both acoustic-only model and localized model exhibited large variations in the MPW predictions. The poor performance in this test sample could be due to the difficulties of inference during the 'laser-off' condition, when the acoustic signal consists primarily of ambient sound. The ambient noise content was difficult to distinguish, causing the model to predict the melt pool state incorrectly. This indicates one of the areas for future improvement.

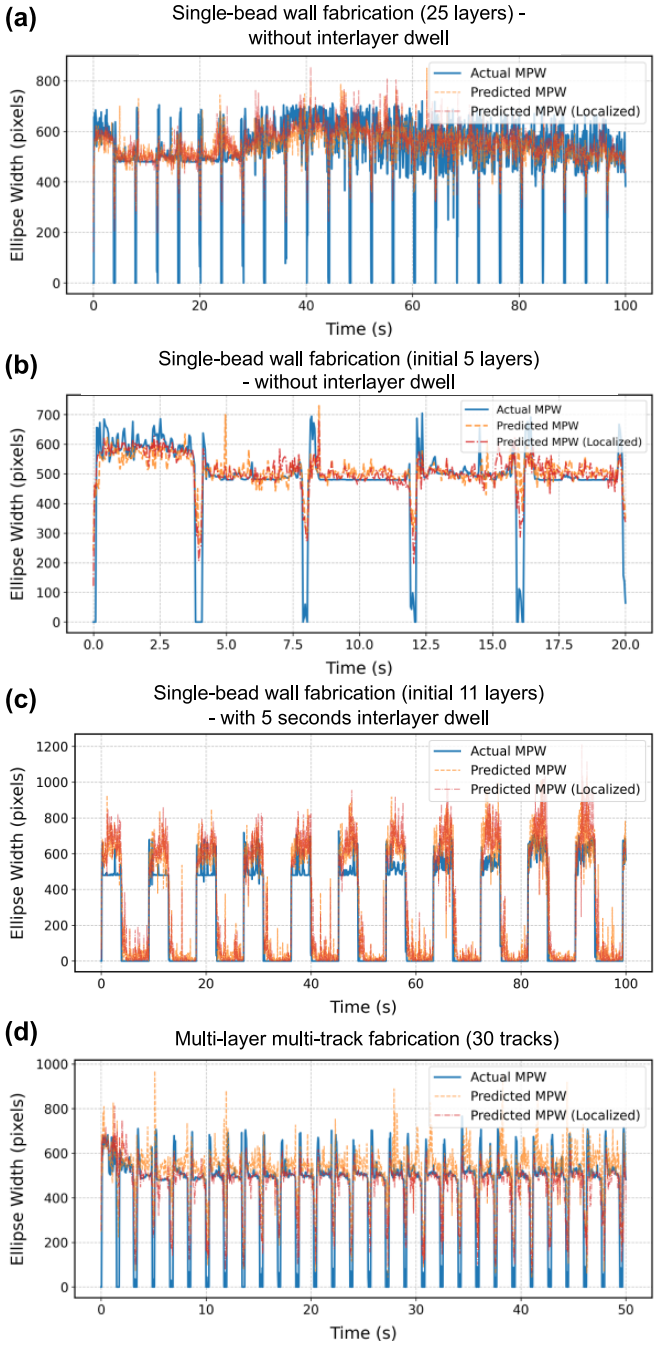

Fig. 7. Visualization of the melt pool width (ground truth vs. predicted) as a function of the time for different test samples. The tracks have been deposited with the parameters of 2.3 kW (powder), 25 mm/s (speed) and 12 g/min (powder feeding rate).

The multi-layer multi-track fabrication with 30 tracks (Fig. 7(d)) shows a complex deposition scenario in which the actual MPW exhibits a more complex pattern due to overlapping tracks. Both the standard and localized predictions have a strong correlation with the actual MPW, but the localized predictions clearly outperform the standard acoustic-only predictions.

IV. Conclusions

This paper presented a novel method that used acoustic signal features and robotic motion data to predict melt pool visual characteristics in LAM. The proposed method revealed that combining acoustic signals with spatiotemporal data from the robot's TCP movements considerably improves the prediction accuracy of melt pool width , which is one of the most commonly used features for closed-loop control in LAM. The RF regressor achieved the best performance, with the highest $\boldsymbol{R^2}$ score (0.73) and the lowest MAE (116.4), indicating that the model is capable of monitoring complicated melt pool visual characteristics. The results are validated by a series of experiments across various LAM scenarios. The future research aims to incorporate this approach into an acoustic-based closed-loop control system, with the predicted visual feature serving as the feedback signal to adjust the laser power input. This technique has the potential to provide a versatile and cost-effective solution for robotic LDED quality monitoring and assurance without the need for time-consuming and computationally demanding visual feature extraction.

Acknowledgment

This research is funded by the Agency for Science, Technology and Research (A*STAR) of Singapore through RIE2025 MTC IAFPP grant (Grant No. M22K5a0045) and the Career Development Fund (Grant No. C210812030). It is also supported by Singapore Centre for 3D Printing (SC3DP), the National Research Foundation, Prime Minister's Office, Singapore under its Medium-Sized Centre funding scheme.

References

[1] J. Lee *et al.*, "Key Enabling Technologies for Smart Factory in Automotive Industry : Status and Applications," *International Journal of Precision Engineering and Manufacturing-Smart Technology*, vol. 1, no. 1, pp. 93–105, Jan. 2023.

[2] P. C. Priarone, A. R. Catalano, and L. Settineri, "Additive manufacturing for the automotive industry: on the life-cycle environmental implications of material substitution and lightweighting through re-design," *Prog Addit Manuf*, Jan. 2023, doi: 10.1007/s40964-023-00395-x.

[3] J. Su *et al.*, "Additive manufacturing of novel heterostructured martensite-austenite dual-phase steel through in-situ alloying," *Materials Today Communications*, vol. 33, p. 104724, Dec. 2022, doi: 10.1016/j.mtcomm.2022.104724.

[4] C. Tan, F. Weng, S. Sui, Y. Chew, and G. Bi, "Progress and perspectives in laser additive manufacturing of key aeroengine materials," *International Journal of Machine Tools and Manufacture*, vol. 170, p. 103804, Nov. 2021, doi: 10.1016/j.ijmachtools.2021.103804.

[5] L. Chen, X. Yao, P. Xu, S. K. Moon, and G. Bi, "Surface Monitoring for Additive Manufacturing with in-situ Point Cloud Processing," in *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*, Apr. 2020, pp. 196–201. doi: 10.1109/ICCAR49639.2020.9108092.

[6] L. Chen, X. Yao, P. Xu, S. K. Moon, and G. Bi, "Rapid surface defect identification for additive manufacturing with in-situ point cloud processing and machine learning," *Virtual and Physical Prototyping*, vol. 16, no. 1, pp. 50–67, 2021, doi: 10.1080/17452759.2020.1832695.

[7] V. Azamfirei, F. Psarommatis, and Y. Lagrosen, "Application of automation for in-line quality inspection, a zero-defect manufacturing approach," *Journal of Manufacturing Systems*, vol. 67, pp. 1–22, Apr. 2023, doi: 10.1016/j.jmsy.2022.12.010.

[8] L. Chen, X. Yao, Y. Chew, F. Weng, S. K. Moon, and G. Bi, "Data-Driven Adaptive Control for Laser-Based Additive Manufacturing with Automatic Controller Tuning," *Applied Sciences*, vol. 10, no. 22, p. 7967, Nov. 2020, doi: 10.3390/app10227967.

[9] Z. Smoqi *et al.*, "Closed-loop control of meltpool temperature in directed energy deposition," *Materials & Design*, vol. 215, p. 110508, Mar. 2022, doi: 10.1016/j.matdes.2022.110508.

[10] Z. Ren *et al.*, "Machine learning–aided real-time detection of keyhole pore generation in laser powder bed fusion," *Science*, vol. 379, no. 6627, pp. 89–94, Jan. 2023, doi: 10.1126/science.add4667.

[11] L. Chen, X. Yao, and S. K. Moon, "In-situ acoustic monitoring of direct energy deposition process with deep learning-assisted signal denoising," *Materials Today: Proceedings*, Sep. 2022, doi: 10.1016/j.matpr.2022.09.008.

[12] L. Chen *et al.*, "In-situ crack and keyhole pore detection in laser directed energy deposition through acoustic signal and deep learning," *Additive Manufacturing*, vol. 69, p. 103547, May 2023, doi: 10.1016/j.addma.2023.103547.

[13] H. Liu *et al.*, "Inference of highly time-resolved melt pool visual characteristics and spatially-dependent lack-of-fusion defects in laser powder bed fusion using acoustic and thermal emission data." arXiv, Nov. 01, 2023. doi: 10.48550/arXiv.2310.05289.

[14] L. Chen, X. Yao, W. Feng, Y. Chew, and S. K. Moon, "Multimodal Sensor Fusion for Real-Time Location-Dependent Defect Detection in Laser-Directed Energy Deposition," presented at the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, Nov. 2023. doi: 10.1115/DETC2023-110284.

[15] L. Chen *et al.*, "Multisensor fusion-based digital twin for localized quality prediction in robotic laser-directed energy deposition," *Robotics and Computer-Integrated Manufacturing*, vol. 84, p. 102581, Dec. 2023, doi: 10.1016/j.rcim.2023.102581.

[16] L. Chen, X. Yao, W. Feng, Y. Chew, and S. K. Moon, "Multimodal sensor fusion for real-time location-dependent defect detection in laser-directed energy deposition." arXiv, May 22, 2023. doi: 10.48550/arXiv.2305.13596.

[17] L. Chen, X. Yao, K. Liu, C. Tan, and S. K. Moon, "MULTISENSOR FUSION-BASED DIGITAL TWIN IN ADDITIVE MANUFACTURING FOR IN-SITU QUALITY MONITORING AND DEFECT CORRECTION," *Proceedings of the Design Society*, vol. 3, pp. 2755–2764, Jul. 2023, doi: 10.1017/pds.2023.276.

[18] P. Xu, X. Yao, L. Chen, K. Liu, and G. Bi, "Heuristic Kinematics of a Redundant Robot-Positioner System for Additive Manufacturing," in *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*, Apr. 2020, pp. 119–123. doi: 10.1109/ICCAR49639.2020.9108047.

[19] D. Bogdanov *et al.*, "Essentia: an audio analysis library for music information retrieval," presented at the 14th Conference of the International Society for Music Information Retrieval (ISMIR'13), Curitiba, Brazil, 2013, pp. 493–498.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[22] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[23] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 3149–3157.