



Real-time monitoring and quality assurance for laser-based directed energy deposition: integrating co-axial imaging and self-supervised deep learning framework

Vigneashwara Pandiyan¹ · Di Cui¹ · Roland Axel Richter¹ · Annapaola Parrilli² · Marc Leparoux¹

Received: 5 August 2023 / Accepted: 10 November 2023
© The Author(s) 2023

Abstract

Artificial Intelligence (AI) has emerged as a promising solution for real-time monitoring of the quality of additively manufactured (AM) metallic parts. This study focuses on the Laser-based Directed Energy Deposition (L-DED) process and utilizes embedded vision systems to capture critical melt pool characteristics for continuous monitoring. Two self-learning frameworks based on Convolutional Neural Networks and Transformer architecture are applied to process zone images from different DED process regimes, enabling in-situ monitoring without ground truth information. The evaluation is based on a dataset of process zone images obtained during the deposition of titanium powder (Cp-Ti, grade 1), forming a cube geometry using four laser regimes. By training and evaluating the Deep Learning (DL) algorithms using a co-axially mounted Charged Couple Device (CCD) camera within the process zone, the down-sampled representations of process zone images are effectively used with conventional classifiers for L-DED process monitoring. The high classification accuracies achieved validate the feasibility and efficacy of self-learning strategies in real-time quality assessment of AM. This study highlights the potential of AI-based monitoring systems and self-learning algorithms in quantifying the quality of AM metallic parts during fabrication. The integration of embedded vision systems and self-learning algorithms presents a novel contribution, particularly in the context of the L-DED process. The findings open avenues for further research and development in AM process monitoring, emphasizing the importance of self-supervised in situ monitoring techniques in ensuring part quality during fabrication.

Keywords Directed energy deposition · Process monitoring · Self-supervised learning · Melt pool · Convolutional Neural Network

Introduction

Directed Energy Deposition (DED), similar to other Additive Manufacturing (AM) methods, entails the continuous melting and deposition of material onto a designated surface, following a Computer-Aided Design (CAD) model, where it solidifies and fuses to create a dense part (Mazumder,

2015). DED is a highly versatile AM technique, enabling precise material deposition and finding extensive applications across diverse industries. The three primary categories of DED applications encompass fabricating near-net-shape parts, enhancing features, and performing repairs (Dutta et al., 2011). The advantages of DED over conventional manufacturing techniques are noteworthy, as it allows for the production of near-net-shape components, thereby reducing material wastage. This method also facilitates the fabrication of new moderately bulk parts with minimal or no tooling requirements and location-dependent properties. Furthermore, it proves instrumental in augmenting existing components by adding material to enhance their performance. The repair and refurbishment of worn-out or damaged areas in components are efficiently achieved through DED,

✉ Vigneashwara Pandiyan
vigneashwara.solairajapandiyan@empa.ch

¹ Laboratory for Advanced Materials Processing (LAMP), Swiss Federal Laboratories for Materials Science and Technology (Empa), 3602 Thun, Switzerland

² Center for X-Ray Analytics, Swiss Federal Laboratories for Materials Science and Technology (EMPA), 8600 Dübendorf, Switzerland

where additional material is built up. Moreover, the capability to combine dissimilar metals and create functionally graded structures with varying material properties further underscores DED's versatility. A key strength of DED lies in its ability to process new materials and multi-materials with exceptional efficiency (Hauser et al., 2020). Industries such as aerospace, automotive, tooling, and medical have widely adopted DED for diverse applications, including rapid prototyping, tool and die repair, coating, surface modification, and even the production of large-scale components. Notably, DED's adaptability extends to a wide array of materials, including metals, alloys, and composites.

In DED, a metal powder or wire is introduced at the focal point of a thermal energy source, where it undergoes melting and is fused onto the substrate or previously deposited layers. Various thermal sources can be employed in DED processes, with the selection based on factors like the material being processed, the desired application, and the specific DED technology utilized. The primary thermal sources used in DED are lasers and plasma or electric arcs. Laser-based DED (L-DED) is more prevalent, offering precise control and exceptional accuracy, making it suitable for a diverse range of materials and applications. The intense heat from the laser beam rapidly melts the material, facilitating solid bonding with the substrate or the preceding layer. In the L-DED process, using powder as the raw material, the powder is conveyed toward the process zone and irradiated by the laser source either co-axially or non-coaxially, promoting melting and deposition. Typically, the part to be fabricated remains stationary while a robotic arm moves to deposit material. Alternatively, a movable platform with stationary arms can be used to reverse this arrangement. The motion control system effectively guides the energy source and material deposition, enabling the fabrication of intricate geometries. The process is commonly conducted within a controlled inert chamber with low oxygen levels. Shielding the part with protective gas is also practised to prevent contamination during printing. Laser-based DED offers numerous advantages, including high deposition rates, precise control, material adaptability, minimal waste, in situ metallurgical bonding, repair capabilities, and design flexibility. It finds extensive applications in aerospace, automotive, tooling, and repair industries, where it is employed for manufacturing complex components, repairing damaged parts, and enhancing material properties through localized deposition.

The primary process parameters crucial for ensuring the quality of L-DED parts are laser power, travel speed, powder mass flow rate, powder focus, and laser beam focus (Liu et al., 2021a, 2021b; Song et al., 2012a, 2012b). Proper calibration of these parameters is essential to achieve the desired bead geometry and layer height for subsequent layer buildup. Deviations in these parameters can affect subsequent layers,

leading to changes in printing distance, subsequently inducing printing in a defocus zone and causing cumulative shape deviation. Uniform track morphology is vital for consistent material properties and minimal defects, but achieving it is challenging due to the intricate interplay of multiple process parameters. The thermal activity during layer stacking also influences part geometry, microstructure, and mechanical properties. Deviations from the suitable process space can result in various anomalies like pores, fractures, cavities, oxidation, delamination effects, geometric aberrations, and distortion (Kim et al., 2017; Liu et al., 2021a, 2021b). Some parameters may need real-time adjustments during the process. Hence, monitoring and control strategies are explored, including monitoring powder flow, thermal activity, and melt pool characteristics. Precise alignment of powder flow and laser beam focus with the printing surface is necessary to optimize process efficiency, minimize porosity, increase density, achieve better mechanical properties, and reduce distortion and thermal stresses. The right balance between laser power and feed speed is crucial to attaining a homogenous temperature distribution. Deviations during the process can reduce powder catchment, delivered laser power, melt pool depth, and area, resulting in minimized layer height and increased defects. Consequently, this directly affects the final material's mechanical properties and 3D geometry. To ensure high-quality products, precise control and monitoring of process parameters are of utmost importance.

The deposition occurs by means of the melt pool, which is co-related to all those fundamental parameters governing the L-DED process, so it is reasonable to expect that its state represents a significant portion of the quality of the built part. Co-axial and off-axial process zone sensing are two approaches used in DED to capture emissions from the process zone, i.e. the melt pool, as encapsulated through a didactic illustration as shown in Fig. 1. Analyzing the emissions from the melt pool gives the current state of the DED process. Co-axial process monitoring involves placing the sensing and monitoring equipment on the same axis as the energy beam used for material deposition. Co-axial process monitoring involves integrating sensors and cameras along with the energy beam to monitor the process in real time, enabling immediate feedback, defect detection, and control adjustments. On the other hand, off-axial process monitoring places sensors at a distance from the deposition zone to capture specific process characteristics without interfering with the energy source. These sensors can be positioned to capture specific aspects of the process, such as the plume of molten material or the heat-affected zone. They can also monitor the temperature distribution on the workpiece or track the growth of the deposited material. Off-axial monitoring allows for collecting focused data on specific process characteristics, such as temperature gradients or solidification rates,

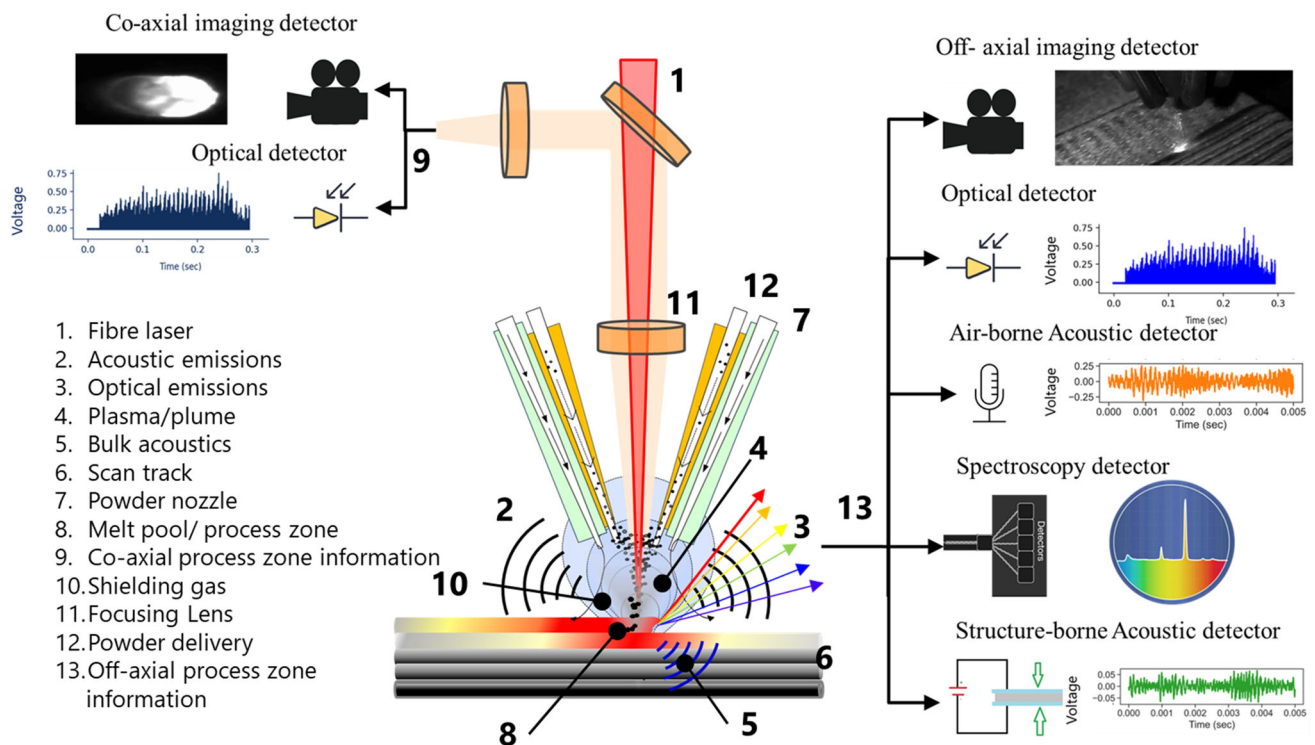


Fig. 1 Co-axial and off-axial process zone sensing are two approaches used in DED (Color figure online)

which can be vital for understanding the material's properties and optimizing the process. Both approaches are essential in ensuring DED's quality, accuracy, and optimization, providing valuable data for understanding material properties and achieving high-quality additively manufactured parts. It's important to note that co-axial and off-axial sensing approaches can complement each other. The advantage of co-axial sensing is that it offers a top view of the melt pool or the process zone independent of the feed direction.

Utilizing non-intrusive sensors in the process zone and coupling them with advanced algorithms enables the collection and decoding of crucial data, leading to improved decision-making and enhanced process efficiency. Moreover, these non-intrusive sensors do not alter the process's environment, allowing for undisturbed observations. One key advantage of such systems is their continuous monitoring capability, which surpasses intermittent machine diagnostics and post-mortem inspection techniques. This real-time monitoring allows immediate intervention if poor part quality is detected, preventing subsequent anomalies and offering better control over the manufacturing process (Stutzman et al., 2018; Xia et al., 2020). Ultimately, these advancements contribute to greater precision, reliability, and quality in DED (Bi et al., 2007), elevating its value as a manufacturing technology for various industries. Table 1 displays a list of non-intrusive sensors documented in the literature that can

record different aspects of the process zone from laser-based DED and differentiate between steady state and abnormality.

Due to the complicated physics involved in laser-based processes, most events in the process zone are transient. In order to ensure that online diagnostics is a prudent alternative for quantification, the multifaceted data from sensors that are capable of capturing such events are to be interpreted seamlessly, and decisions have to be made simultaneously with minimal human intervention (Pandiyan, Masinelli, et al., 2022). The paradigm of Machine Learning (ML) and soft computing techniques provides such a platform since it can recognize nonlinear patterns in the data, learn from them, and ultimately make decisions. Previous research into monitoring the DED process by integrating sensing methods and ML has demonstrated the benefits of doing so.

Zhang et al. (2020) estimated melt pool temperature using extreme gradient boosting (XGBoost) and long short-term memory (LSTM) models with process parameters as input. Sampson et al. (2020) suggested an image-based edge identification technique for determining the shape of the melt pool. Khanzadeh et al., (2018a, 2018b) demonstrated the monitoring of process irregularities using principal component analysis (PCA) based on melt pool thermal imagery. Thermal images have also been employed to locate porosity during the DED process as input using Self-Organizing Maps (Khanzadeh et al., 2019). Convolutional Neural Networks (CNNs)

Table 1 Non-intrusive sensors used for monitoring different characteristics in all the variants of laser-based DED systems

Sensor	Remarks on process zone information	Reference
Charged Couple Device (CCD) and Complementary Metal–Oxide Semiconductor (CMOS) based imaging	• Geometric characteristics of the deposit	Iravani-Tabrizipour and Toyserkani, (2007), Meriaudeau and Truchetet (1996), Ribeiro et al. (2023), and Toyserkani and Khajepour (2006)
	• Powder deposition parameters/powder clogging	Doubenskaia et al. (2004), Lee et al. (2022), and Lei et al. (2012)
	• Process zone dynamics	Abe et al. (2013), Gharbi et al. (2013), Haley et al. (2018), Nassar et al. (2015), and Vandone et al. (2019)
	• Melt pool characteristics	Pandiyani et al., (2022a, 2022b, 2022c), Sampson et al. (2020), Vykhatar and Richter (2021), Zhang et al. (2021), and Zijue et al. (2019)
	• Process monitoring	Li et al. (2020) and Zhang et al. (2019)
Photodiodes	• Effect of build geometry and powder morphology	Bi et al. (2006b)
	• Powder flow rate	Ding et al. (2016), and Hu and Kovacevic (2003)
Pyrometers	• Infrared thermography-based melt pool characterization	Esfahani et al. (2022), Khanzadeh et al., (2018a, 2018b), and Smurov et al. (2012)
	• Porosity prediction	Khanzadeh et al., (2018a, 2018b), Khanzadeh et al., (2018a, 2018b), and Tian et al. (2020)
	• Melt pool surface temperature	Bi et al. (2006a), Doubenskaia et al. (2004), Hua et al. (2008), and Song et al., (2012a, 2012b)
Acoustic Emission (AE)	• Process stability	Bond et al. (2019) and Hauser et al. (2022)
	• Porosities and cracks	Chen et al. (2023) and Gaja and Liou (2018)
	• Powder deposition parameters	Koester et al. (2018) and Whiting et al. (2018)
Spectroscopy	• Melt pool composition monitoring	Bartkowiak (2010), Lednev et al. (2018), Schmidt et al. (2021), Song et al. (2016), Song and Mazumder (2011), Song, Wang, et al. (2012), and Wasmer et al. (2023)
	• Process quality	Ren et al. (2022) and Stutzman et al. (2018)
Hyperspectral camera	• Melt pool temperature	Devesse et al. (2016), Ertveldt et al. (2020), and Lison et al. (2019)

have been trained on raw medium wavelength Infrared co-axial images to correlate process dynamics and predict quality (Gonzalez-Val et al., 2020). Bayesian approaches and least squares regression have been used with photodiode signatures to detect and localise defects efficiently (Felix et al., 2022). Heterogeneous sensing techniques comprising an optical camera, spectrometer, and Supervised Support Vector Machine (SVM) have been demonstrated to detect lack-of-fusion defects (Montazeri et al., 2019). Monitoring strategies using CNN-based contrastive learners and generative models using process zone images have also been reported by the authors of this work (Pandiyan et al., 2022a, 2022b, 2022c; Pandiyan et al., 2022a, 2022b, 2022c). Wang et al. (2008) used AE and advanced pattern recognition to detect fractures. In contrast, Gaja and Liou (2018) used AE to assess the overall deposition quality. Using the Mel-Frequency Cepstral Coefficients (MFCCs) of the acoustic emission signals generated during the laser-material interaction in DED, defects such as cracks and keyholes could be predicted by training a CNN (Chen et al., 2023). Ribeiro et al. (2023) proposed a hybrid approach combining CNN and Support Vector Regression (SVR) utilizing melt pool images from an NIR camera to detect unexpected changes in the Z-axis and optimize the layer height in the build process. Besides monitoring, sensor information has been effectively used with soft computing techniques to control the process states. High speed dual color pyrometer sensor was used to collect melt pool information, which was sent into a closed-loop controller, which changed process inputs such as laser intensity to preserve part dimensions (Dutta et al., 2011). Using three high-speed charged couple device cameras in a triangulation arrangement and a dual color pyrometer, a hybrid controller was created to provide steady layer development by preventing overbuilding and underbuilding (Song et al., 2012a, 2012b).

Most approaches are trained using supervised and semi-supervised methods in the literature addressing monitoring strategies for DED processes based on process zone imaging (Zhang et al., 2022). However, the work discussed here by the authors represents a pioneering effort in developing a framework for monitoring part quality in terms of build density using a CNN trained in a self-supervised manner. This idea was inspired by the Bootstrap Your Own Latent (BYOL) approach (Grill et al., 2020). The motivation behind this work arises from the challenges encountered in labelling the dataset, mainly when dealing with process dynamics that are not discrete and involve semantic understanding. The authors recognized the limitations of traditional supervised approaches, which rely on explicit labels provided by human experts. In the context of DED processes, the complexities involved in defining discrete labels for quality assessment make the task challenging. To overcome these challenges, the authors adopted a self-supervised learning approach based on

the BYOL method. Self-supervised learning not only simplifies the training process but also reduces the time and cost associated with algorithm setup while enabling the model to learn meaningful representations from the data, leading to improved quality assessment and monitoring capabilities in the DED process. Once the model is trained using self-supervised learning, the learned representations can be transferred to downstream tasks. These representations serve as a powerful feature extractor that captures relevant information from the input data, enabling improved performance on various tasks. This opens up opportunities for discovering new insights and enhancing the monitoring and assessment of DED processes, such as evaluating build density in this case.

The scientific contribution presented in this paper is structured into six main sections. "Introduction" section serves as an introduction, providing a concise overview of the DED process and outlining the specific problem statement addressed in the study. "Theoretical basis: bootstrap your own latent (BYOL)" section offers a theoretical overview of the self-learning technique employed, which is based on BYOL. This section aims to provide readers with a brief understanding of the underlying principles and concepts involved in the self-learning approach used in the study. The experimental setup, data collection pipelines and the proposed methodology of self-supervised learning are introduced in "Experimental setup and methodology" section. "L-DED monitoring" and "Self-supervised learning" sections presents and discusses the results obtained from machine-vision-based build quality monitoring using the self-learning CNN and Transformer architectures, comparing them to similar architectures trained using supervised learning. These section analyses the performance and effectiveness of the self-learning approach in the context of build quality monitoring. Finally, "Conclusion" section concludes the paper by summarizing the main findings and contributions of the study. It also suggests potential avenues for future research and development in this field, highlighting areas that could be explored further based on current research outcomes.

Theoretical basis: bootstrap your own latent (BYOL)

The training of ML algorithms is usually supervised. Given a dataset consisting of an input and corresponding label, under a supervised paradigm, a typical classification algorithm tries to discover the best function that maps the input data to the correct labels. On the contrary, self-supervised learning does not classify the data to its labels. Instead, it learns about functions that map input data to themselves (Hendrycks et al., 2019; Liu et al., 2021a, 2021b). Self-supervised learning

helps reduce the amount of labelling required. Additionally, a model self-supervisedly trained on unlabeled data can be refined on a smaller sample of annotated data. (Jaiswal et al., 2020; Pandiyan et al., 2023).

BYOL is a state-of-the-art self-supervised method proposed by DeepMind and Imperial College researchers that can learn appropriate image representations for many downstream tasks at once and does not require labelled negatives like most contrastive learning methods (Grill et al., 2020). The BYOL framework consists of two neural networks, online and target, that interact and learn from each other iteratively through their bootstrap representations, as shown in Fig. 2. Both networks share the architectures but not the weights. The online network is defined by a set of weights θ and comprises three components: Encoder (f_θ), projector (g_θ), and predictor (q_θ). The architecture of the target network is the same as the online network but with a different set of weights ξ , which are initialized randomly. The online network has an extra Multi-layer Perceptron (MLP) layer, making the two networks asymmetric. During training, an online network is trained from one augmented view of an image to predict a target network representation of the same image under another augmented view. The standard augmentation applied on the actual images is a random crop, jitter, rotation, translation, and others. The objective of the training was to minimize the distance between the embeddings computed from the online and target network, as shown in Eq. (1).

$$\text{Loss function} \rightarrow \mathcal{L}_{\theta, \xi} \triangleq \|\overline{q_\theta}(z_\theta) - \overline{z}_\xi\|^2. \quad (1)$$

The update of target network parameter ξ is given by

$$\xi = \tau \xi + (1 - \tau)\theta, \text{ where } \tau \text{ is a decay parameter.}$$

The weights are updated only for the online network through gradient descent, and weights for the target network are updated only through an exponential moving average based on the weights from the online network as shown in Eq. (1).

BYOL's popularity stems primarily from its ability to learn representations for various downstream visual computing tasks such as object recognition (Afouras et al., 2022; Mitash et al., 2017) and semantic segmentation (Tung et al., 2017) or any other task-specific network (Zhou et al., 2020), as it gets a better result than training these networks from scratch. As far as this work is concerned, based on the shape of the process zone and the four quality categories that were gathered, BYOL was employed in this work to develop appropriate representations that could be used for in situ process monitoring.

Experimental setup and methodology

DED experimental setup

The experiments described in this study utilized a 5-axis industrial DED laser system manufactured by BeAM (Mobile 1.0, BeAM, France). The DED system was equipped with a co-axial process zone imaging setup, as depicted in Fig. 3. For the fabrication process, four cubes were built on a 4 mm thick baseplate made of 99.6% pure Ti6Al4V titanium alloy sourced from Zapp AG, Germany. The titanium powder used in the process was Cp-Ti grade 1 from AP&C (Advanced Powders & Coatings, Inc., Canada), consisting of particles ranging in size from 45 to 106 μm . The elemental composition of the powder is given in Table 2.

The fabrication process employed a serpentine scan pattern, and the DED process was carried out in an argon environment with a controlled oxygen concentration of 5 ppm. The argon gas was supplied by Carba gas, with a purity of 99.6%. The laser used was a ytterbium-doped fiber laser manufactured by IPG Photonics, USA, operating at a wavelength of 1068 nm. The laser deposition head was capable of focusing the laser beam to a nominal $1/e^2$ diameter of 800 μm . The beam profile followed a Gaussian distribution with a Rayleigh range of 18 mm. The deposition of layers was achieved by vertically moving the deposition head along the Z-axis, while movement in the X–Y plane enabled the formation of the desired 2D geometry. More comprehensive information regarding the elemental composition of the powder and detailed specifications of the experimental setup can be found in previously published works (Pandiyan et al., 2022a, 2022b, 2022c).

In terms of part density, four different build qualities were achieved by keeping the scan speed constant at 2000 mm/min and varying the laser power from 175 to 475 W, as depicted in Fig. 4. In other words, four distinct deposited energies induced the build quality. The offline computed tomography (CT) analysis confirmed all four quality grades (*P1–P4*). An EasyTom XL Ultra 230-160 micro/nano CT scanner (RX Solutions, Chavanod, France) was used to perform CT scan observations on the cube cross-sections. The samples were scanned with an 8 μm voxel size. Images were captured at a frame rate of 8 frames/s with an average of ten frames per projection. Images were reconstructed using filtered back projection based on the FDK algorithm for 3D cone-beam tomography. Figure 4a shows a 3D CT scan of the *P1* state, confirming the low density due to insufficient laser power, which results in incomplete powder adhesion, i.e., Lack of Fusion (LoF) porosity. However, the density increases with increasing laser power. Although the densities of regions *P3* and *P4* are high due to high heat storage, an undesirable microstructure evolution was observed, as reported in previous studies (Pandiyan et al., 2022a, 2022b, 2022c; Pandiyan

Fig. 2 An illustrative representation of the workflow for a self-supervised classification pipeline (Color figure online)

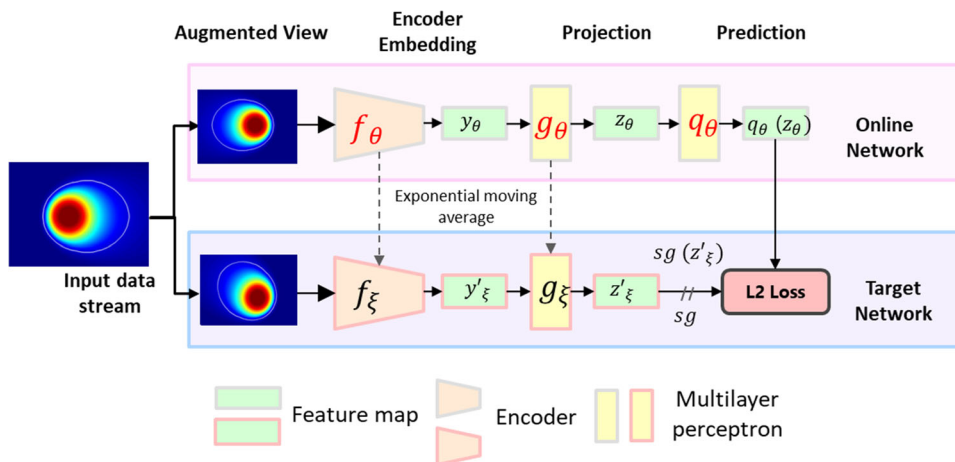


Fig. 3 Visual representation of the experimental setup for DED process incorporating co-axial imaging capability (Color figure online)

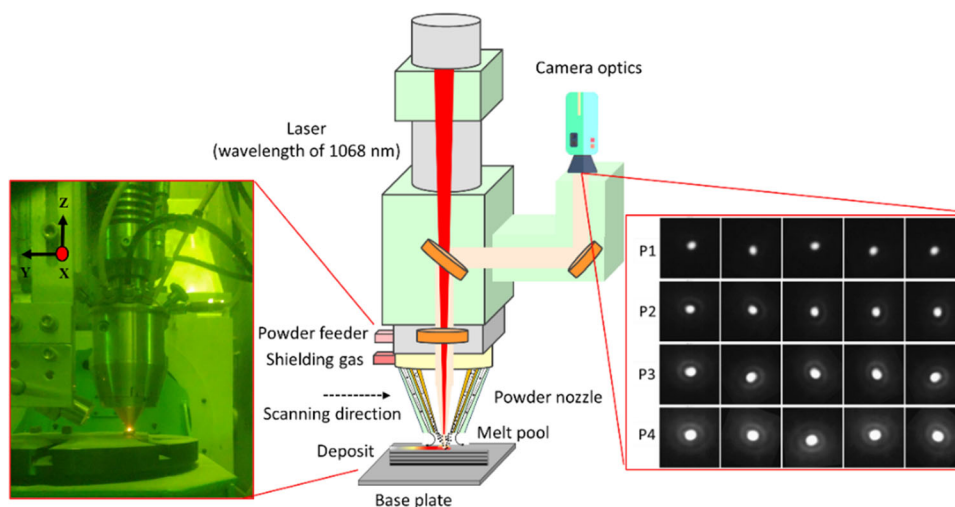


Table 2 Minimum and maximum percentage of elements comprising Ti grade 1, ASTM B348 powder

Material	Carbon C	Oxygen O ₂	Nitrogen N ₂	Hydrogen H ₂	Iron Fe	Other each, max	Other total, max	Titanium Ti
Composition (wt%)	0.02	0.08	0.02	0.002	0.03	< 0.1	< 0.4	Balance

et al., 2022a, 2022b, 2022c). Of the four parameters considered in this work, *P2* corresponds to favourable build conditions. The selection of ‘*P2*’ is justified by its alignment with a conduction regime that prevents the formation of defects. The *P2* condition results in a uniform primary α -microstructure, which is crucial for attaining the desired quality. In contrast, *P3* and *P4*, while having higher density, tend to exhibit the generation of primary α -acicular and secondary α -microstructures (Pandiyani et al., 2022a, 2022b, 2022c; Pandiyani et al., 2022a, 2022b, 2022c). These microstructures are considered undesirable as they indicate unstable melt pools and non-uniform material characteristics within the process zone.

Dataset

The experimental setup utilized a co-axial color Charged Couple Device (CCD) camera from Watec, Japan, integrated into the laser deposition head. This camera operates at a frame rate of 30 frames per second and captures the morphology of the process area. The captured images consist of three RGB channels with a 640×480 pixels resolution. To enable the camera to capture the radiation from the process zone, a beam splitter is installed on Precitec’s laser applicator head. An optical notch filter within the 650–675 nm range also blocks the laser wavelengths. The captured images are extracted from a continuous video stream using Image J software

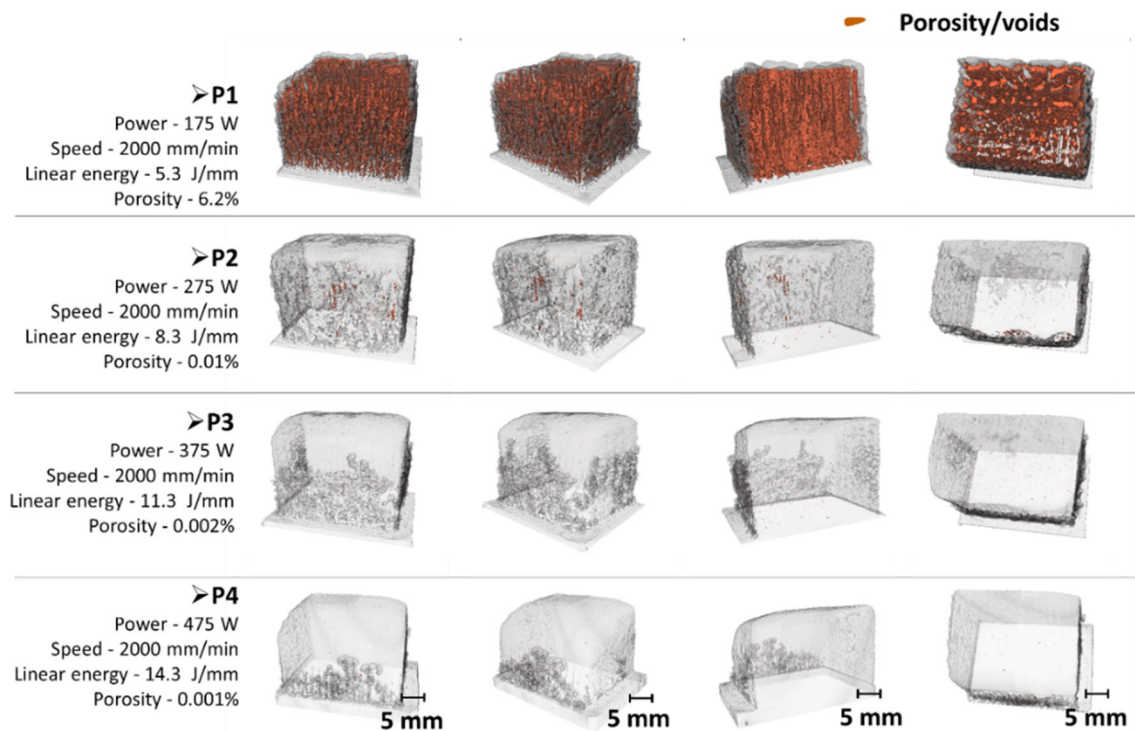


Fig. 4 Quantitative analysis of porosity in cubes fabricated using four different process regimes through Computed Tomography (CT) scans (Color figure online)

for offline processing, as mentioned in reference (Schneider et al., 2012). Figure 3 illustrates the variations in the melt pool and process zone morphology for each quality grade under the four different conditions. The dataset used in the study comprises a total of 48,995 process zone images representing the four quality grades. The dataset is randomly divided into three training, validation, and testing subsets. The first subset contains 60% of the training images, the second subset (10%) is used for validation, and the remaining images are reserved for testing the self-supervised algorithm's performance. To facilitate training and evaluation, the raw images are rescaled to a resolution of 480×320 pixels while maintaining the three-channel input. This rescaling operation aims to reduce computational resources while preserving classification accuracy. It should be noted that further reduction in image resolution may adversely affect the classification results.

Methodology

This work's suggested self-supervised learning methodology is schematically depicted in Fig. 5 for learning the lower dimensional representations from process zone morphological images. In order to extract appropriate representations unsupervisedly, we used a self-supervised learning method based on Jean-Bastien et al. (Grill et al., 2020), who first discussed BYOL. We expanded on this first effort to extract

appropriate representations to classify process zone morphology images that correlate with the quality of the part built. The advantage of using learned representations from self-supervised learning is that they capture rich and meaningful information from the data. This can improve downstream task performance, even when labelled data is limited or unavailable. By leveraging the learned representations, models can generalize better, extract relevant features, and make more accurate predictions.

Training a CNN using the self-supervised method typically involves three steps. The first step in training a CNN using the self-supervised method is data preparation. In this phase, the process zone images from different categories are gathered and combined to form the dataset. It is important to note that no ground truth labels are utilized at this stage. The process zone images are typically obtained from the DED process and represent different aspects or variations in the manufacturing process. Both online and target networks get input from this unlabeled dataset. Random transformations and scaling are applied before inputting the raw process zone image into the networks, as shown in Fig. 6. These augmentations provide an augmented view of the raw process zone image, which helps enhance the model's ability to learn invariant and discriminative features. These augmented images capture different perspectives and variations of the original image, which aids in training a more robust and generalized model.

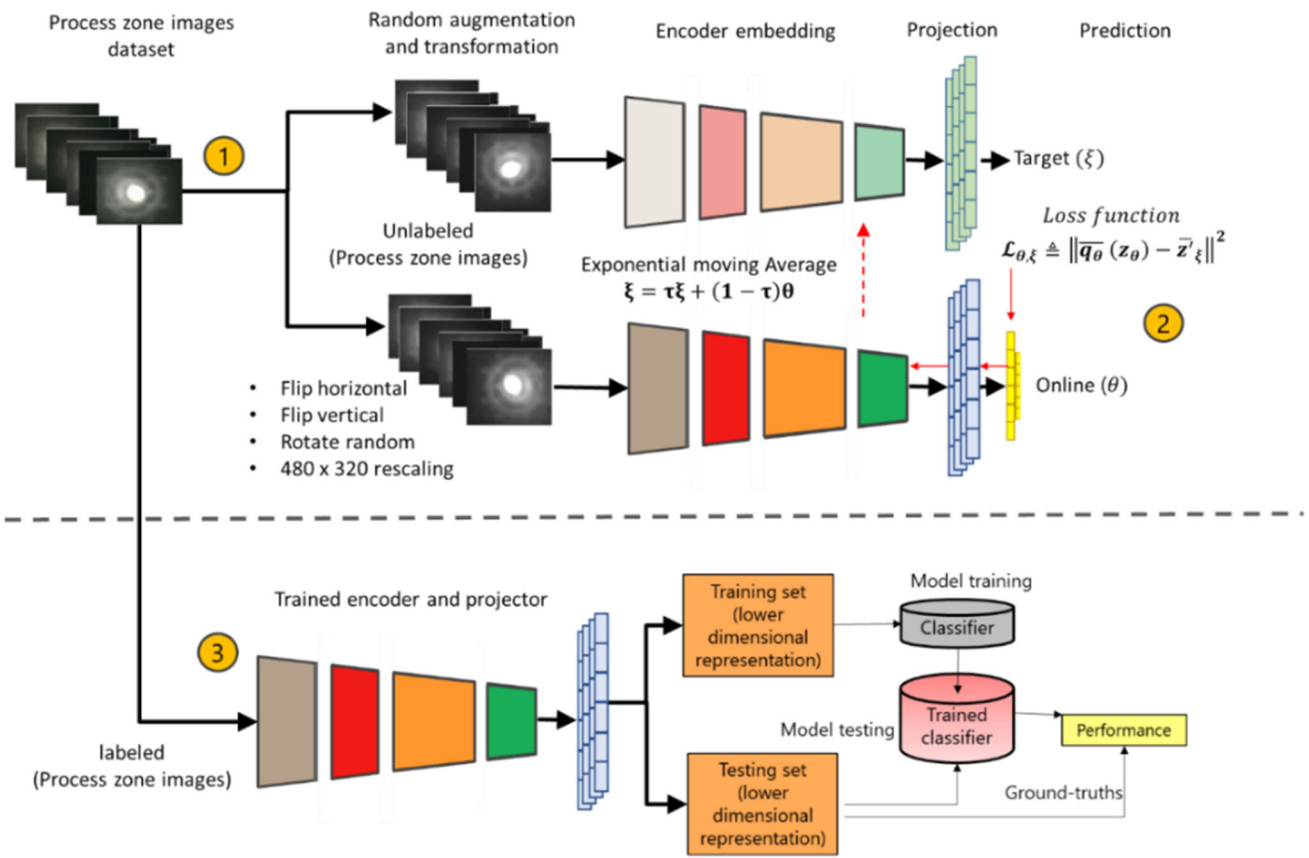


Fig. 5 Illustrative framework depicting the proposed methodology of self-supervised learning for process zone images (Color figure online)

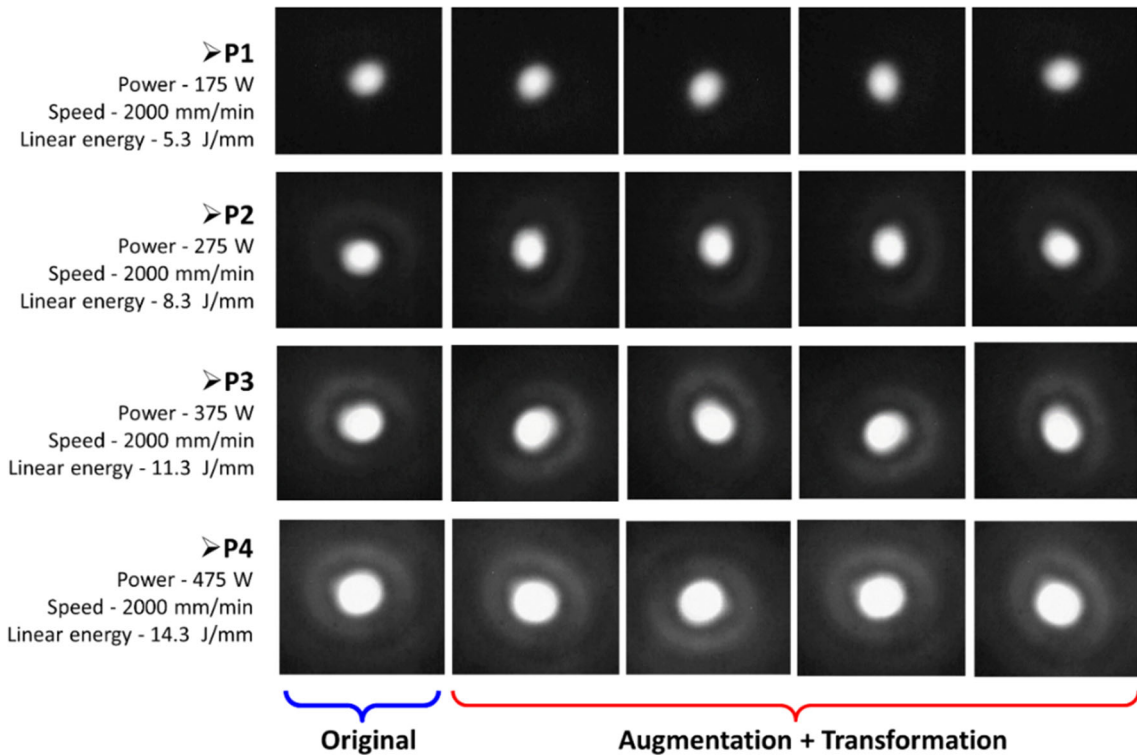
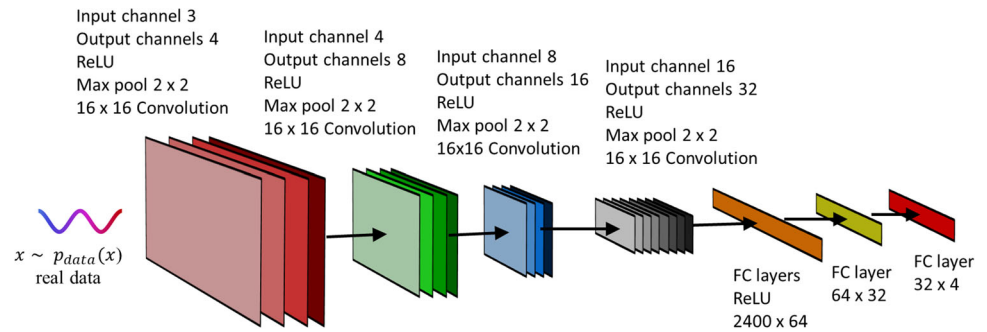


Fig. 6 Application of data augmentation techniques (flipping and rotation) on process zone images of four build conditions

Fig. 7 Architectural overview of the CNN employed for supervised classification (Color figure online)



The second step of the methodology involves training the online and target networks to extract lower-dimensional representations from the augmented process zone images. This is achieved by minimizing the distance between the representations, which is calculated using the Mean Squared Error (MSE) loss function. The loss is backpropagated only in the online network. As the online network outperforms the target network, the weights from the online network are progressively transferred to the target network using an exponential moving average. Throughout the training process, it is ensured that the magnitude of the training loss decreases, indicating the convergence of the network learning. This signifies that the network is effectively capturing the distinct embedding among the four categories without prior knowledge of their labels. In the third and final step, the network's performance trained using the self-supervised technique is assessed on a labelled dataset. For training, the lower-level feature map computed from each image and their corresponding ground truths is fed into simple classifiers such as Random Forest (RF), SVM, or K-Nearest Neighbors (k-NN). Once these classifiers are trained, the models are utilized to classify the feature maps acquired for samples in the labelled

dataset that were not part of the training process. The utilization of the lower-level feature map in classification tasks results in higher accuracy due to the network's ability to learn the distinct embedding among the four categories without relying on explicit labels. Furthermore, beyond classification, the lower-level feature map can be employed for other downstream tasks, such as segmentation and process control, which are crucial for industries relying on these processes.

L-DED monitoring

Supervised learning: baseline: CNN architecture

To establish a baseline for self-supervised learning on a Convolutional architecture, a CNN network, referred to as *CNN-1*, was trained in a supervised manner to classify the four quality categories using the dataset described in previous sections. The *CNN-1* architecture was implemented using the PyTorch package provided by Meta (USA) (Paszke et al., 2019). The architecture of *CNN-1* comprised four convolutional layers and three fully connected layers, as illustrated in Fig. 7. This architectural design was selected after an extensive search process.

The *CNN-1* architecture consisted of four 2D convolutional layers with a kernel size of 16. The number of kernels gradually increased from four in the first layer to 32 in the fourth layer. Max pooling with a kernel size of two was applied to downsample the feature map after each convolutional layer. For classification, three fully connected layers were utilized. The Rectified Linear Unit (ReLU) activation function was employed for all layers of *CNN-1*. The dataset was divided probabilistically into training (60%), validation (10%), and testing (30%) sets. The input to the first *CNN* layer was a tensor of size $N \times 320 \times 480 \times 3$, where N represented the batch size (set to 256). This tensor corresponded to the rescaled process zone image after applying augmentation techniques such as flipping and rotation. The output of *CNN-1* was a feature map with dimensions $N \times 1 \times 4$, which was compared with the ground truth labels using the cross-entropy loss. The model was trained for 100

Table 3 Training parameters of the CNN-1 model trained supervisedly with cross-entropy loss

Training parameters	<i>CNN-1</i>
Analysis type	Supervised classification
Solver used	'AdamW'
Learning rate used	0.01
Training epochs	100
Batch size (N)	256
Dropout	5%
Loss function	Cross entropy
Shuffle	Every-epoch
Labelled dataset	60% Training, 30% testing, 10% Validation
Trainable weights	0.60 Million

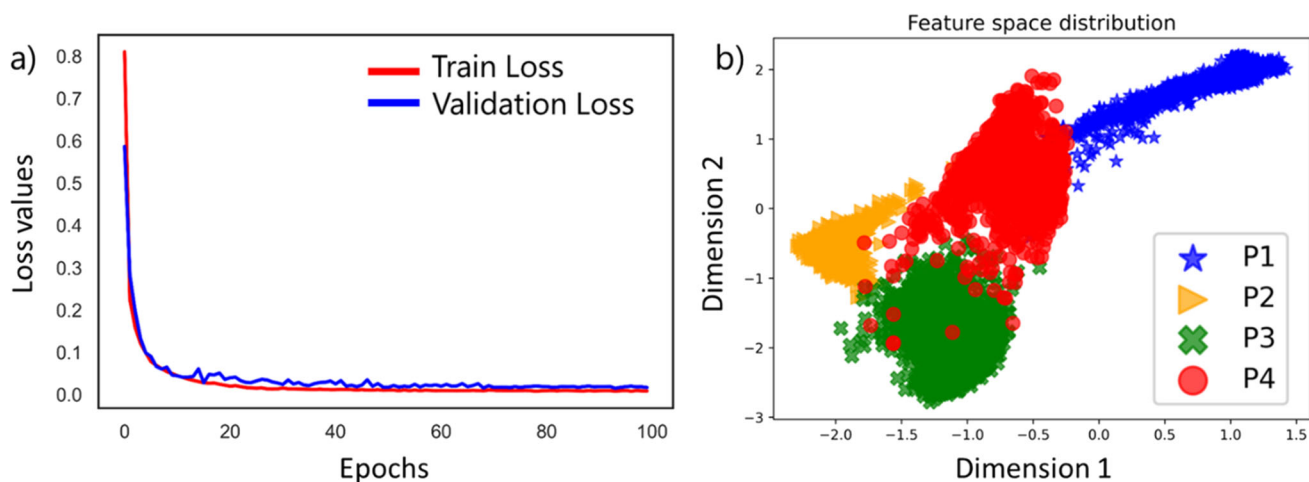


Fig. 8 Inference plots after model training **a** Learning curves illustrating model training progress: evaluating performance and convergence

of the *CNN-I* model and **b** Visualization of lower-dimensional representations of process zone images for four build grades: leveraging the encoding power of the *CNN-I* model (Color figure online)

epochs, with each epoch taking approximately six hours. The learning rate was adjusted every 20 epochs, starting at 0.01 and multiplied by a gamma value of 0.5. To prevent overfitting, batch normalization was applied across the layers of the *CNN* model. The datasets were randomized between epochs, and the training optimizer used was AdamW. A 5% dropout was applied during training between each cycle. The *CNN-I* model was trained on a Graphical Processing Unit (GPU), specifically the Nvidia RTX Titan and the training parameters are summarized in Table 3. The total number of tuned parameters in *CNN-I* was 0.6 million.

Figure 8a depicts the visualization of the *CNN-I* model’s training loss values over 100 training epochs. The *CNN-I* model has figured out how to decode the distributions encoded in images, as seen by the progression of the loss values. Additionally, the validation loss resembles the training loss, demonstrating that the model appropriately generalizes to new data. Visualizing the lower-dimensional representation, or the feature maps of the final fully connected layer, is another technique to verify the effectiveness of the trained *CNN* models. Figure 8b displays a 3D visualization of the weights from three nodes in the last fully connected layer

of the *CNN-I* model that was trained using the validation dataset. This diagram shows distinct clusters corresponding to the four various build grades. The clustered space on the validation dataset images demonstrates that the models have learned and separated themselves from other distributions by learning the distributions corresponding to the same type.

Table 4 displays the classification results as a confusion matrix. The classification accuracy is calculated by dividing the total number of tests by the true positives. On the test set, *CNN-I*’s average classification accuracy was 99.5%. According to the classification accuracy results in Table 4, *CNN-I* has learnt the differences and similarities between the process zone morphologies that correspond to the four build grades.

Supervised learning: baseline: transformer architecture

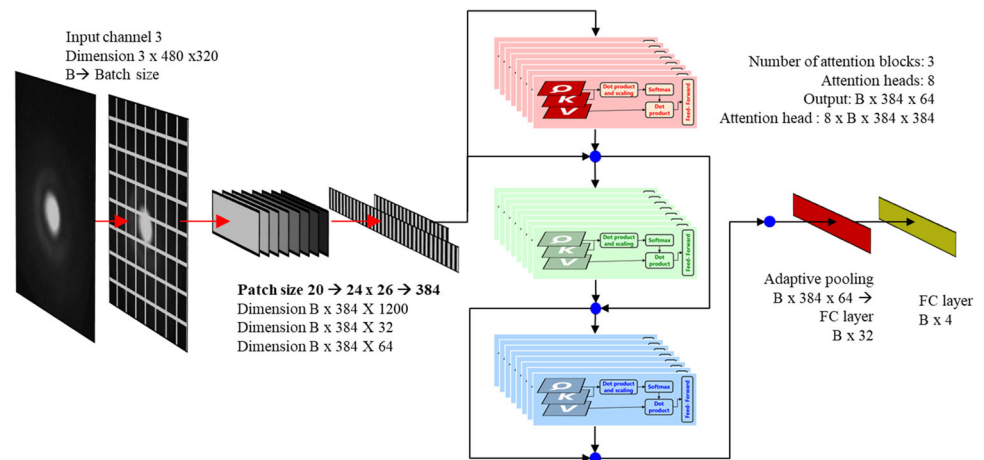
To establish a baseline for self-supervised learning also on a Transformer architecture, the Transformer network, denoted as *TF-I*, was trained in a supervised manner to classify the

Table 4 Confusion matrix on prediction outcomes of trained *CNN-I* model

Classification [Accuracy (%)]	Ground truth			
	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>
<i>P1</i>	100	0	0	0
<i>P2</i>	0	100	0	0
<i>P3</i>	0	0	99.5	0.5
<i>P4</i>	0	0	0.5	99.5

The bold, Normal and italics are used to separate three methodologies as is consistent throughout the document

Fig. 9 Architectural overview of the *TF-1* employed for supervised classification (Color figure online)



four quality categories using the dataset described in previous sections. The *TF-1* architecture is based on the vision transformer (ViT) (Dosovitskiy et al., 2020) and takes input images with dimensions $3 \times 480 \times 320$, where each image has 3 color channels, a width of 480 pixels and a length of 320 pixels. The *TF-1* architecture comprises three attention blocks, each equipped with eight attention heads, as illustrated in Fig. 9.

The *TF-1* network adopts a patch-based approach with self-attention mechanisms to process the input image efficiently. The image is divided into patches of size 20×20 pixels, resulting in 24 patches width-wise (480 pixels/20 pixels) and 16 patches length-wise (320 pixels/20 pixels). These patches are then linearly transformed and reshaped into a tensor of dimensions Batch size (B) \times 384 \times 64, where 384 is calculated by multiplying the number of width-wise patches and length-wise patches. This step allows the network to extract meaningful features from each patch effectively. The transformed patches undergo three attention blocks, each containing eight attention heads. These attention blocks enable the model to focus on relevant parts of the image and effectively capture local and global dependencies. The output of each attention block is a tensor of dimensions $B \times 384 \times 64$. After the attention blocks, adaptive pooling is applied to reduce the spatial dimensions of the tensor, resulting in a tensor of dimensions $B \times 1 \times 64$. This pooling operation aggregates information from different patches, creating a more concise representation of the input. Following the pooling step, the tensor is processed by two linear layers, resulting in an output tensor of dimensions $B \times 1 \times 4$. This tensor represents the extracted features from the input image after being transformed by the Transformer architecture. The cross-entropy loss function compares the output tensor with the ground truth labels. This loss function measures the dissimilarity between the predicted features and the actual labels, facilitating the model's learning and improvement during training. The *TF-1* model is trained for

Table 5 Training parameters of the *TF-1* model trained supervisedly with cross-entropy loss

Training parameters	<i>TF-1</i>
Analysis type	Supervised classification
Solver used	'AdamW'
Learning rate used	0.01
Training epochs	100
Batch size (B)	256
Dropout	5%
Loss function	Cross entropy
Shuffle	Every-epoch
Labelled dataset	60% Training, 30% testing, 10% Validation
Trainable weights	0.4 million

100 epochs, with each epoch taking approximately ten hours. The learning rate is adjusted every 20 epochs, starting at 0.01 and multiplied by a gamma of 0.5 value. To prevent overfitting, batch normalization is applied across the layers of the *TF-1* model, and a 5% dropout is used during training between each cycle. The *TF-1* model is trained on an Nvidia RTX Titan GPU, and the training parameters are summarized in Table 5. The total number of tuned parameters in *TF-1* amounts to 0.4 million. Overall, the *TF-1* network architecture with self-attention and attention blocks is a compelling feature extractor, allowing the model to capture relationships between different parts of the image and leading to accurate predictions for the image classification task.

Figure 10a illustrates the training curve of the *TF-1* PyTorch model, displaying its training loss values across 100 epochs. The *TF-1* model exhibits a progressive reduction in loss, indicating its ability to decode the underlying distributions present in the images effectively. Moreover, the similarity between the training loss and the validation loss

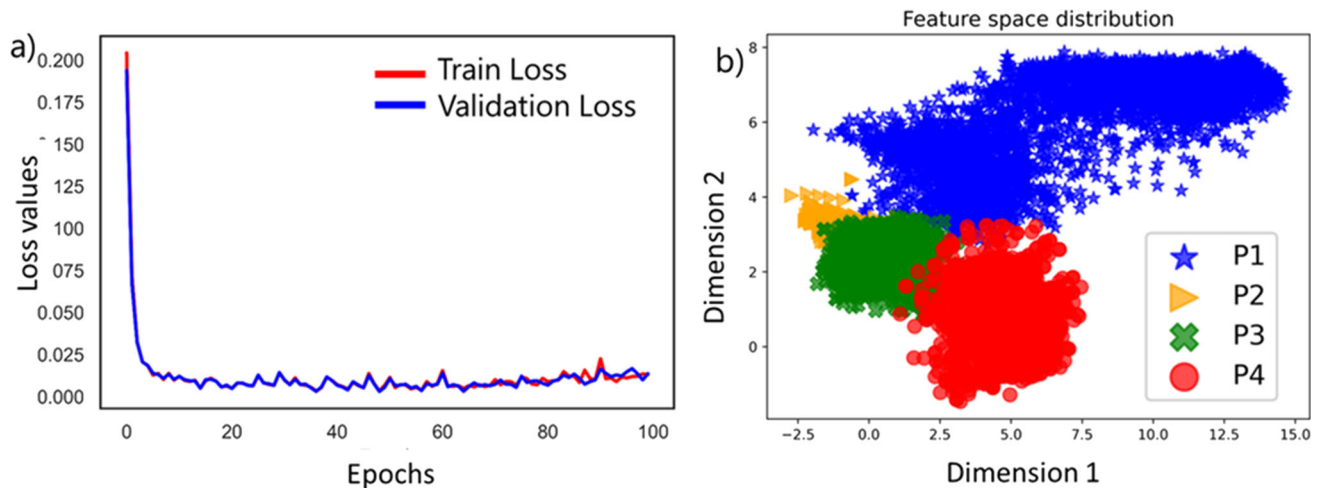


Fig. 10 Inference plots after model training **a** Learning curves illustrating model training progress: evaluating performance and convergence

of the *TF-1* model and **b** Visualization of lower-dimensional representations of process zone images for four build grades: leveraging the encoding power of the *TF-1* model (Color figure online)

signifies that the model generalizes well to new data. To further validate the efficacy of the trained *TF* model, lower-dimensional representations, or feature maps, from the final fully connected layer can be visualized. Figure 10b presents a 3D visualization of the weights obtained from three nodes in the last fully connected layer of the *TF-1* model, trained using the validation dataset. The distinct clusters observed in this visualization correspond to the four different build grades. The clear separation of data points in the validation dataset demonstrates that the model has learned to distinguish and categorize distributions accurately, aligning with their respective types.

Table 6 presents the confusion matrix depicting the classification results. The classification accuracy is determined by dividing the total number of correct predictions (true positives) by the total number of tests. In the case of *TF-1*, the average classification accuracy on the test set was found to be 99.7%, as depicted in Table 6. This high accuracy demonstrates that *TF-1* has successfully learned to distinguish and recognize the distinct process zone morphologies corresponding to the four different build grades.

Self-supervised learning

Self-supervised learning: CNN architecture

The architecture of the self-supervised network (*CNN-2*) used in this research work is shown in Fig. 11. The design of the online and target networks was inspired by the *CNN-1* architecture used for supervised classification in previous sections above. The idea of using a similar *CNN* architecture

between supervised and self-supervised learning (*CNN-2*) was to allow a fair comparison of the two techniques. The online network generates a representation of $N \times 32$ (N is the batch size) as images are fed sequentially and get transformed as they pass through the encoder ($f_{C\theta}$), projector ($g_{C\theta}$), and predictor ($q_{C\theta}$). For the case of the target network, the size of the output representation is similar to the online network. However, it is to be noted that images are fed sequentially only through an encoder ($f_{C\xi}$) and projector ($g_{C\xi}$) in the target network. Though the dimensions of the encoder (five 2D convolutional layers) and projector (fully collected layers) are similar across the online and target network, their weight ($\theta \neq \xi$) are dissimilar.

As stated in previous sections, the dataset was randomly divided into 60% for training, 10% for validation, and 30% for testing. Although the dataset's ground truth was known, it was purposely left out of the training and validation datasets to enable self-supervised learning. However, to check the predictability of the trained model, the ground truths were retained for the test dataset. During training across each batch, the encoder and decoder were supplied individually with two sets of random transformations computed from the same batch in the training set. Three common transformations were simultaneously applied to the images: horizontal flipping, vertical flipping and random rotation between (0° and 180°). A probability score of 0.5 was also set to avoid biasing with the transformations. The objective of the training was to decrease the distance between the representations computed between the encoder and decoder network on the same image subjected to two different transformations. The distance metric to be minimized was the MSE. The MSE loss used as a similarity metric was back-propagated to update the

Table 6 Confusion matrix on prediction outcomes of trained TF-1 model

Classification [Accuracy (%)]]	Ground truth			
	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>
<i>P1</i>	100	0	0	0
<i>P2</i>	0	100	0	0
<i>P3</i>	0	0	99.8	0.2
<i>P4</i>	0	0	0.3	99.7

The bold, Normal and italics are used to separate three methodologies as is consistent throughout the document

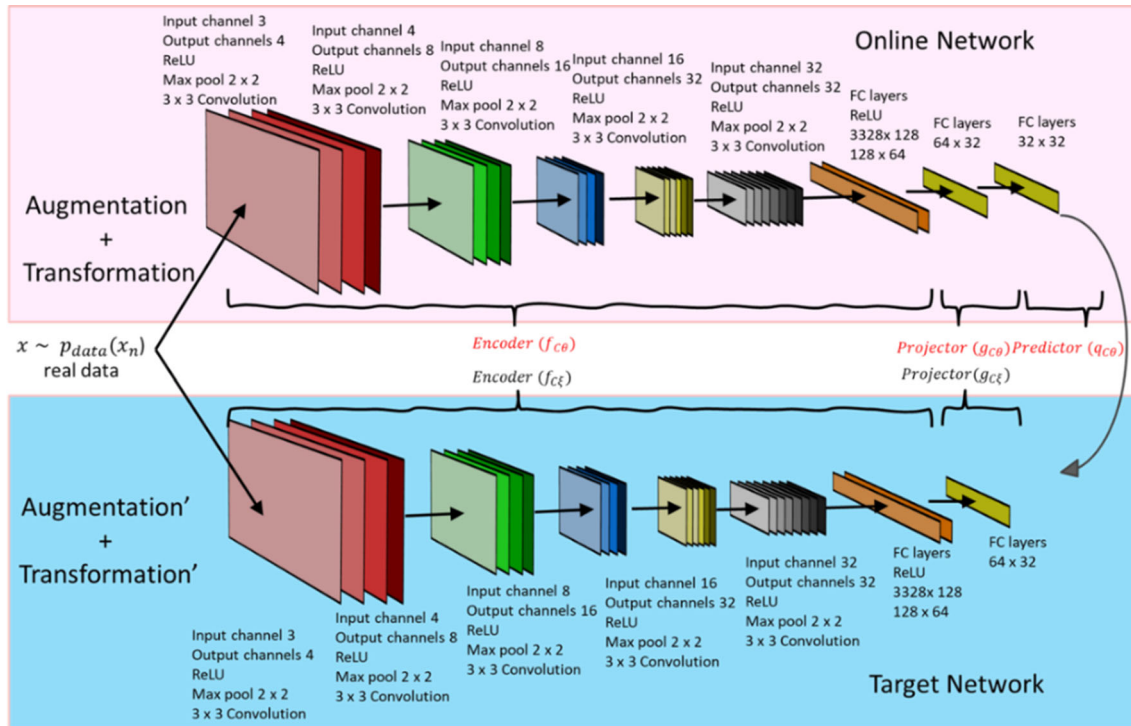


Fig. 11 Proposed CNN-2 architecture for self-supervised computation of lower-dimensional representation (Color figure online)

weights of the online network ($C\theta$), whereas the weights of the target network ($C\xi$) are only updated periodically using an exponential moving average based on the weights of the online network, as stated in Eq. (1) with a decay parameter ($\tau = 0.999$). To prevent overfitting, a 5% dropout was employed between each epoch for the two networks. The model was trained over 100 epochs using a GPU (NVIDIA RTX Titan, NVIDIA, USA). Table 7 shows the parameters utilized during training. On CNN-2, there were 0.95 million parameters tweaked (target and online network combined).

Figure 12a depicts the visualization of the CNN-2 model's loss values, basically the distance between the online network and the target network over 100 training epochs. As seen by the progression of the loss values and its saturation after 20 epochs, it is evident that the CNN-2 model has figured out how to decode the distributions encoded in transformed images. Additionally, the validation loss resembles

the training loss, demonstrating that the model appropriately generalizes to similar distributions. Another method for validating the effectiveness of trained CNN-2 models is to visualize the lower-dimensional representation from the projector ($g_{C\theta}$). The testing dataset with known ground truths was used in this study.

Figure 12b displays a 3D visualization of the lower-dimensional representations on the output from the projector ($g_{C\theta}$) of the CNN-2 model computed using t -distributed stochastic neighbour embedding (t -SNE) with a perplexity value of 20. This diagram shows distinct clusters corresponding to the four various build grades. The clustered space on the test dataset images demonstrates that the models have learned and separated themselves from other distributions by learning the distributions corresponding to the same type. Additionally, distributions produced by individual nodes of the projector ($g_{C\theta}$) was also computed on the test dataset

Table 7 Training parameters for self-supervised training of *CNN-2* model

Training parameters	<i>CNN-2</i>
Type of analysis	Lower-dimensional representation
Solver used	'AdamW'
Learning rate	0.001
Epochs	100
Batch size	100
Dropout	5%
Similarity metric	Mean Squared Error (MSE)
Shuffle	Every-epoch
Training/Validation set	60%/10% (Unlabelled)
Testing set	30% (Labelled)
Trainable weights	0.95 Million (Online and target network combined)

for visualization, as shown in Fig. 13. The distribution of representations by each node of the projector ($g_{C\theta}$) on the four categories had distinct overlaps between *P1* and *P2* from others suggesting self-supervised learning without labels is feasible on the process zone image for process monitoring, which was the primary objective of this work. It is also to be noted that there were overlaps between *P3* and *P4* among them.

In this work, supervised classification among the categories was accomplished from various down-streaming tasks that could be performed on the lower dimensional representation. The trained online network (*CNN-2*) was truncated to the projector ($g_{C\theta}$) and was used as a feature extractor (*CNN-3*). The whole dataset was passed to the network, and the weights, i.e. a feature map of $N \times 32$ dimensions

corresponding to each image, were stored along with the ground truth. This new down-sampled dataset consisting of feature maps was split into the training and testing set at a ratio of 70:30. Supervised classification was performed using three traditional classifiers: *RF*, *SVM* and *k-NN*. The training parameters for the chosen classifiers were determined using empirical guidelines as outlined in Table 8 and their performance was assessed by considering prediction accuracy. It should be emphasized that subsequent research endeavours will prioritize the refinement of these training parameters, considering additional factors like training time and prediction accuracy. By conducting more comprehensive experiments, exploring diverse combinations of hyperparameters, and employing advanced optimization techniques, the goal is to enhance the classifiers' performance and generalization capabilities.

Table 9 displays the classification outcomes from the three classifiers as a confusion matrix. *CNN-3*'s average classification accuracy on the test set was 97% on all three classifiers. The observed misclassifications were more prevalent in scenarios denoted as *P3* and *P4*, which was also evident from the plot depicted in Figs. 12b and 13. These scenarios involve higher energy density and are positioned in close proximity to each other in terms of process dynamics, as evidenced by our tomography data and microstructure analysis. Of the three classifiers trained, *RF* showed higher predictability than the other two.

Self-supervised learning: transformer architecture

The architecture of the self-supervised Transformer network (*TF-2*) proposed in this research work is shown in Fig. 14. The design of the online and target networks was inspired by the *TF-1* architecture used for supervised classification.

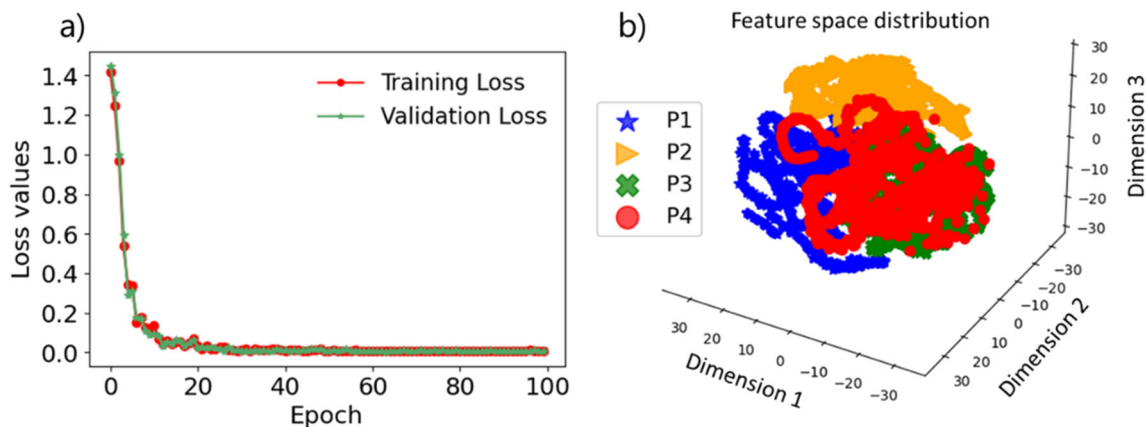


Fig. 12 Inference plots after model training **a** Learning curves illustrating model training progress: evaluating performance and convergence

of the *CNN-2* model and **b** Visualization of lower-dimensional representations of process zone images for four build grades, computed using t-SNE on the encodings of the *CNN-2* model (Color figure online)

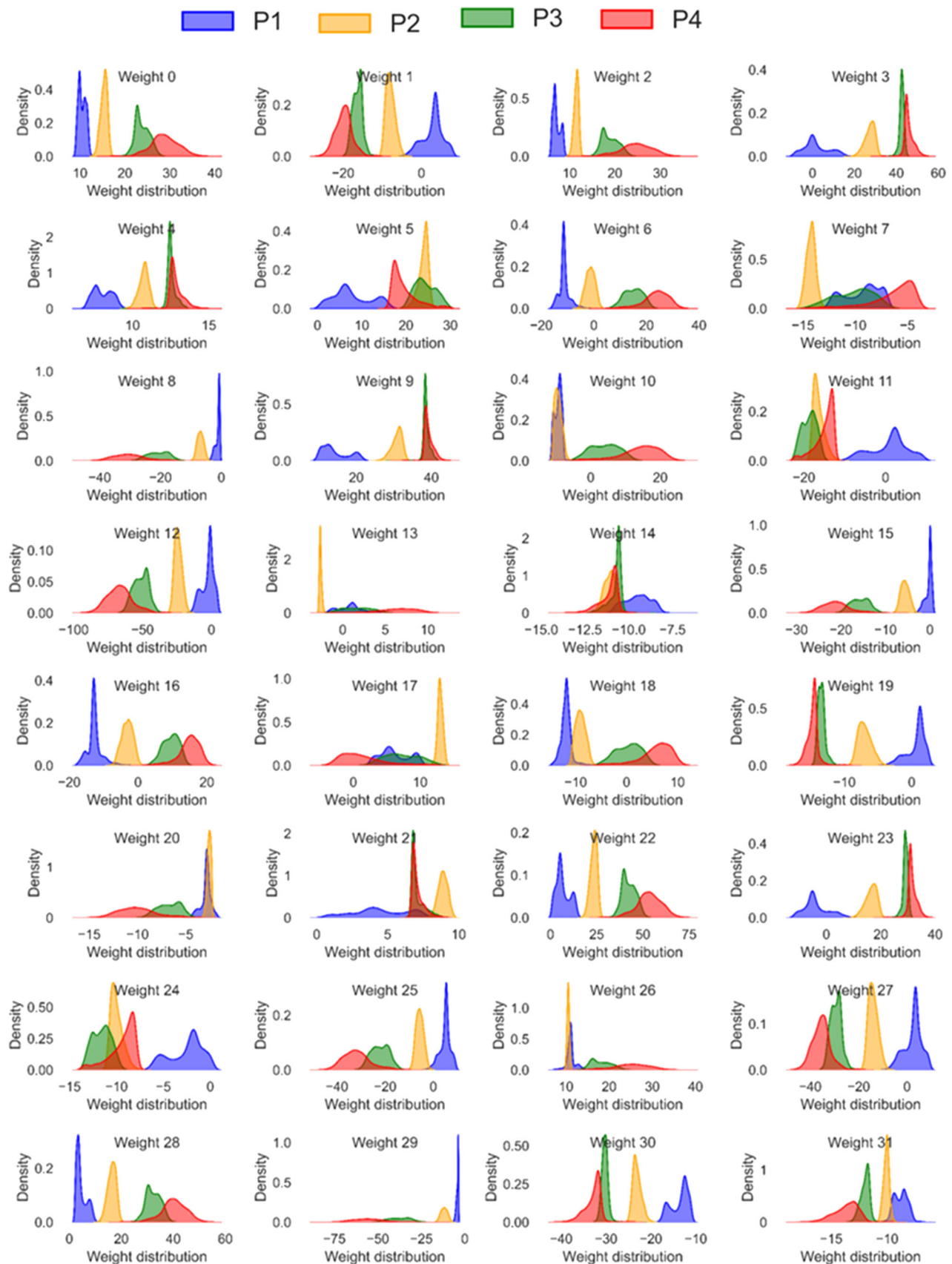


Fig. 13 Distribution analysis of lower-dimensional representations computed using CNN-2 on four grades of process zone images from the projector ($gC0$) (Color figure online)

Table 8 Training parameters of classifiers for predicting build conditions (*RF*, *SVM* and *k-NN*)

Classifier type	Parameters	Values
Random Forest (RF)	n_estimators	100
	Split criterion	Gini's diversity
	Maximum-features	Sqrt
Support Vector Machine (SVM)	SVM kernel function	Radial basis function
	Gamma scale	Automatic
	Multiclass method	One-vs-rest
	Standardize data	True
k-Nearest Neighbor (k-NN)	Classifier	Minkowski
	Number of neighbours	10
	Distance metric	Distance
	Distance weight	Uniform

Table 9 Confusion matrix depicting the comparative performance of the three traditional classifiers (*RF*, *SVM* and *k-NN*)

Classification [Accuracy (%)]	Ground truth			
	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>
<i>P1</i>	100	0	0	0
	100	0	0	0
	<i>100</i>	<i>0</i>	<i>0</i>	<i>0</i>
<i>P2</i>	0	100	0	0
	0	100	0	0
	<i>0</i>	<i>100</i>	<i>0</i>	<i>0</i>
<i>P3</i>	0	0	97	3
	0	0	98	2
	<i>0</i>	<i>0</i>	<i>98</i>	<i>2</i>
<i>P4</i>	0	0	6	94
	0	0	10	90
	<i>0</i>	<i>0</i>	<i>7</i>	<i>93</i>

The bold, Normal and italics are used to separate three methodologies as is consistent throughout the document

Table displays the classification results obtained with the three traditional classifiers (**RF**, *SVM* and *k-NN*). The classification results in each cell from top to bottom are for **RF** (**CNN-3**), *SVM* (*CNN-3*) and *k-NN* (*CNN-3*). All values are in %

The idea of using a similar *TF-1* architecture between supervised and self-supervised learning (*TF-2*) was to allow a fair comparison of the two techniques. The online network generates a representation of $B \times 32$ (B is the batch size) as images are fed sequentially and get transformed as they pass through the encoder based on the Transformer backbone ($f_{T\theta}$), projector ($g_{T\theta}$), and predictor ($q_{T\theta}$). For the case of the

target network, the size of the output representation is similar to the online network. However, it is to be noted that images are fed sequentially only through an encoder ($f_{T\xi}$) and projector ($g_{T\xi}$) in the target network. Though the dimensions of the encoder and projector are similar across the online and target network, their weight ($\theta \neq \xi$) are dissimilar. As mentioned in previous sections, the dataset was partitioned randomly into three subsets: 60% for training, 10% for validation, and 30% for testing. The ground truth information was intentionally withheld from the training and validation datasets to enable self-supervised learning. However, the ground truths were retained for the test dataset to evaluate the trained model's predictive performance. During the training process for *TF-2*, similar to *CNN-2* training, each batch involved two sets of random transformations computed from the same training set. These transformations included horizontal flipping, vertical flipping, and random rotation within the range of 0° to 180° . Each transformation was applied with a probability score of 0.5 to prevent any biasing. The main training objective was to minimize the MSE distance between the representations generated by the encoder and decoder networks for the same image subjected to the two different transformations. The MSE loss, serving as a similarity metric, was used for back-propagation to update the weights of the online network ($T\theta$). In contrast, the target network's ($T\xi$) weights were updated periodically using an exponential moving average based on the online network's weights, as denoted by Eq. (1) with a decay parameter ($\tau = 0.999$). A 5% dropout was applied between each epoch for both networks to avoid overfitting. The model was trained for 100 epochs on a GPU (NVIDIA RTX Titan, NVIDIA, USA). Table 10 presents the specific parameters utilized during the training process. *TF-2* involved 0.94 million adjusted parameters (combining the target and online network).

Figure 15a illustrates the training curves of the *TF-2* model, representing the loss values, which essentially measure the distance between the online network and the target network over the course of 100 training epochs. The progressive decrease in loss values, followed by saturation after approximately 20 epochs, indicates that *TF-2* has effectively learned to decode the distributions encoded in transformed images. Moreover, the validation loss closely resembles the training loss, indicating that the model generalizes well to similar distributions. To further assess the effectiveness of the trained *TF-2* models, lower-dimensional representations obtained from the projector ($g_{T\theta}$) were visualized using t-SNE with a perplexity value of 20. Figure 15b displays a 3D visualization of these representations, showcasing distinct clusters corresponding to the four different build grades. The clustering on the test dataset demonstrates that the models have learned to effectively separate themselves from other distributions by learning the distributions associated with each specific type. Additionally, the distributions produced

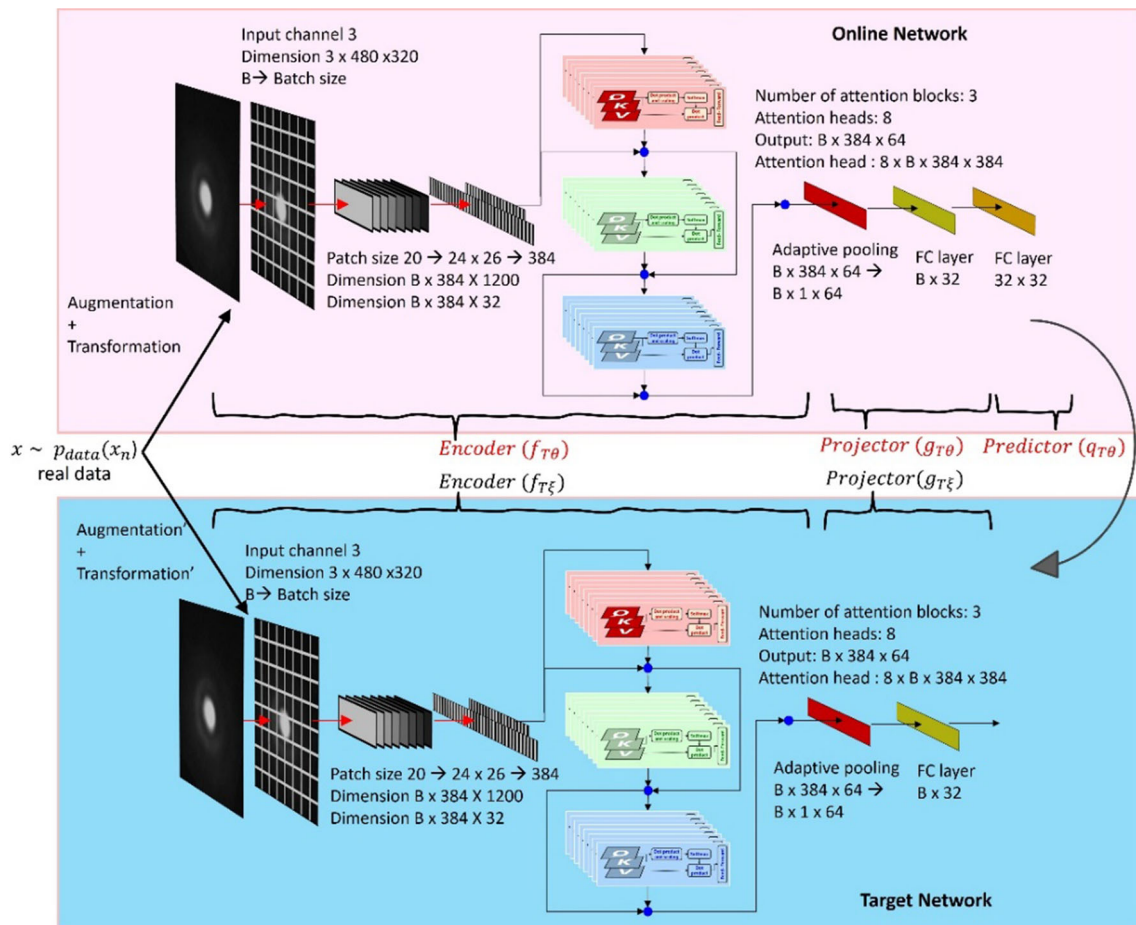


Fig. 14 Proposed TF-2 architecture for self-supervised computation of lower-dimensional representation (Color figure online)

by individual nodes of the projector ($g_{T\theta}$) were computed on the test dataset for visualization, as shown in Fig. 16. It was observed that the representations from each node of the projector ($g_{T\theta}$) for the four categories exhibited distinct overlaps between $P1$ and $P2$, indicating the feasibility of self-supervised learning without labels for process monitoring, which was the primary objective of this study. It should be noted that there were overlaps between $P3$ and $P4$ among them.

To carry out supervised classification among the categories, a down-sampled dataset consisting of feature maps was obtained by truncating the trained online network ($TF-3$) to the projector ($g_{T\theta}$). This down-sampled dataset was then split into training and testing sets at a ratio of 70:30. Three traditional classifiers, namely RF , SVM , and $k-NN$, were employed for supervised classification. The training parameters for these classifiers were determined using empirical guidelines outlined in Table 8, and their performance was evaluated based on prediction accuracy. It is important to note that future research endeavors will focus on refining these training parameters, considering additional factors such as training time and prediction accuracy. Table 11 presents

Table 10 Training parameters for self-supervised training of $TF-2$ model

Training parameters	$TF-2$
Type of analysis	Lower-dimensional representation
Solver used	'AdamW'
Learning rate	0.001
Epochs	100
Batch size	100
Dropout	5%
Similarity metric	Mean Squared Error (MSE)
Shuffle	Every-epoch
Training/Validation set	60%/10% (Unlabelled)
Testing set	30% (Labelled)
Trainable weights	0.94 Million (Online and target network combined)

the classification outcomes from the three classifiers as a confusion matrix. $TF-3$ achieved an average classification accuracy of 97.2% on the test set across all three classifiers.

The observed misclassifications were indeed more prevalent in scenarios denoted as *P3* and *P4*, as evidenced by the data presented in Figs. 15b and 16. These scenarios are positioned in close proximity to each other in terms of process dynamics, as further supported by our comprehensive tomography data and microstructure analysis. This proximity and similarity in process dynamics contributed to the challenges in accurately distinguishing these scenarios, which were the primary sources of the observed misclassifications. Among the three trained classifiers, *RF* and *k-NN* exhibited higher predictability than the others.

The high accuracy achieved by the self-supervised feature extractors of both the DL architectures provides strong encouragement for its utilization as a strategy for monitoring the L-DED process. Notably, no significant downside was observed when comparing the classification accuracy between supervised and self-supervised learning. While the self-supervised learning accuracy was slightly lower, the authors believe optimizing the transformation techniques during the training procedure could enhance the model's self-learning capability, particularly across overlapping classes. Furthermore, the self-supervised trained models hold the potential for refinement on a smaller sample of annotated data for other challenging downstream tasks. This implies that the model, which has learned meaningful representations from the self-supervised learning phase, can be further fine-tuned or adapted using a smaller set of labelled data specific to the target task. This approach leverages the benefits of both self-supervised learning, which requires only unlabeled data, and supervised learning, which benefits from labelled data. By refining the self-supervised learning process, such as by optimizing transformation techniques and exploring advanced training methodologies, it is expected that the model's performance across different classes and challenging scenarios can be further improved. This optimization can lead to enhanced self-learning capabilities and better generalization on overlapping classes, ultimately boosting the overall performance of the self-supervised monitoring strategy. The classifiers chosen for this study, namely *RF*, *SVM*, and *k-NN*, were selected to explore diverse classification paradigms. Random Forest exemplifies rule-based approaches, *SVM* represents kernel-based methods, and *k-NN* represents distance-based techniques. It's essential to emphasize that the objective was not to advocate for these particular algorithms solely but to evaluate how various techniques perform when applied to computed representations from *CNN-2* and *TF-2*. Ensemble techniques could also be used as an alternative for similar classification tasks (Bustillo et al., 2018). In the *CNN-2* model, the feature extraction process took approximately 6 ms, while for *TF-2*, it was around 9 ms due to its larger size. Subsequently, applying ML classifiers to the latent representation for decision-making took an additional 5 ms. It's important to note that the meltpool co-axial imaging system

operates at a frame rate of 30 frames/s, with each frame being acquired in roughly 33 ms. The decision-making pipeline is faster than the time required for each frame acquisition, making our system nearly real-time. However, it's essential to acknowledge that any change in the acquisition rate could introduce latency, potentially necessitating hardware improvements to handle higher acquisition rates.

While the primary focus in this work is on classification using traditional classifiers, it's worth noting that these acquired representations also have broader potential. They can be effectively applied to subsequent tasks, improving adaptability and performance in L-DED monitoring and control. Furthermore, there is room for future research, which could involve exploring transfer learning for more generalization. In this context, the *CNN-2* and *TF-2* models could serve as adaptable base models, supplemented with new fully connected layers to achieve specific monitoring objectives on newer data spaces. Nevertheless, it's important to clarify that this exceeds the current scope of our study. In L-DED, process control is indispensable for upholding ideal properties and specifications apart from process monitoring. This research is focused on continuous monitoring and sensor integration to detect process faults. Corrective actions can be initiated once deviations from the ideal process space are identified through ML models discussed in this work. These actions encompass adjusting parameters such as laser power, powder flow rate, and scan speed to guarantee accurate layer-by-layer material deposition, producing high-quality components while minimizing defects (González-Barrio et al., 2022; Pérez-Ruiz et al., 2023). Monitoring is the initial step, but following it with well-defined process control strategies that adjust process parameters when deviations arise is crucial.

Conclusion

This research developed a self-supervised deep learning framework to enable real-time monitoring of the L-DED process by imaging the emissions from the process zone. The experimental setup involved the utilization of a laser with a wavelength of 1068 nm and Cp-Ti grade 1 powder characterized by particle sizes ranging from 45 to 106 μm . These materials were employed to manufacture 15 mm \times 15 mm cubes representing four different build grades on a Ti6Al4V Grade 1 base plate. Among the four grades investigated in this work, one grade (*P1*) corresponds to *LoF pores*, a defective regime as it has high porosity. At the same time, the other three (*P2*, *P3*, and *P4*) are still in the conduction regime but have varying porosity levels, density and microstructure. The emissions from the process zone, corresponding to four different build grades, were co-axially captured using a CCD color camera in the visible spectral range to train the CNN and

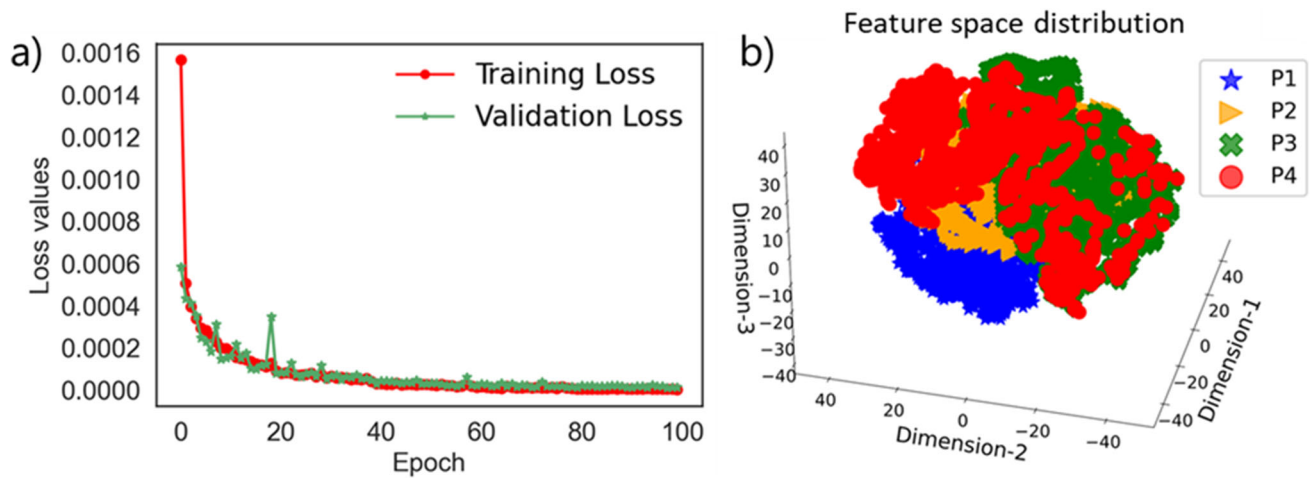


Fig. 15 Inference plots after model training **a** Learning curves illustrating model training progress: evaluating performance and convergence

of the *TF-2* model and **b** Visualization of lower-dimensional representations of process zone images for four build grades, computed using t-SNE on the encodings of the *TF-2* model (Color figure online)

Table 11 Confusion matrix depicting the comparative performance of the three traditional classifiers (*RF*, *SVM* and *k-NN*)

Classification [Accuracy (%)]	Ground truth			
	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>
<i>P1</i>	100	0	0	0
	100	0	0	0
	<i>100</i>	<i>0</i>	<i>0</i>	<i>0</i>
<i>P2</i>	0	100	0	0
	0	100	0	0
	<i>0</i>	<i>100</i>	<i>0</i>	<i>0</i>
<i>P3</i>	0	0	97	3
	0	0	98	2
	<i>0</i>	<i>0</i>	<i>98</i>	<i>2</i>
<i>P4</i>	0	0	7	93
	0	0	9	91
	<i>0</i>	<i>0</i>	<i>7</i>	<i>93</i>

The bold, Normal and italics are used to separate three methodologies as is consistent throughout the document

Table displays the classification results obtained with the three traditional classifiers (**RF**, **SVM** and **k-NN**). The classification results in each cell from top to bottom are for **RF** (**TF-3**), **SVM** (**TF-3**) and **k-NN** (**TF-3**). All values are in %

Transformer networks. The objective of the self-supervised learning approach was to enable the CNN and Transformer model to extract a compact and condensed representation of the process zone images, effectively reducing their dimensionality without relying on explicit ground-truth knowledge. Training CNN and Transformer networks in a self-supervised manner was expected to extract meaningful features from the

process zone images, facilitating the differentiation between the different build grades. The following observations can be drawn from the training performance of the two models:

- The proposed self-supervised learning approach was evaluated for robustness with a *CNN-1* and *TF-1* that had been supervisedly trained and could categorize the four quality categories with good accuracy. The architectures of the CNNs and Transformers trained in a supervised and self-supervised manner were kept the same for a fair comparison.
- *CNN-2* and *TF-2*, trained in a self-supervised manner, efficiently clustered process zone images for the four quality grades based on lower-dimensional representations. This underscores their vital role in our research. Self-supervised learning streamlined training, reduced reliance on labelled data, and empowered the models to extract meaningful insights from the data, ultimately enhancing the L-DED process's quality assessment and monitoring capabilities.
- The performance of the two models trained self-supervisedly was evaluated by training traditional classifiers such as RF, SVM, and k-NN supervisedly on lower-dimensional representation computed on process zone images using *CNN-3* and *TF-3*. The prediction accuracy of all the classifiers was above $\approx 97\%$, indicating that the networks have learnt the representations corresponding to the quality grades without knowing the ground truth.
- The accuracy of the self-supervised frameworks opens up the opportunity of fine-tuning these models for other specific down-streaming tasks, such as segmentation and process control, apart from monitoring.

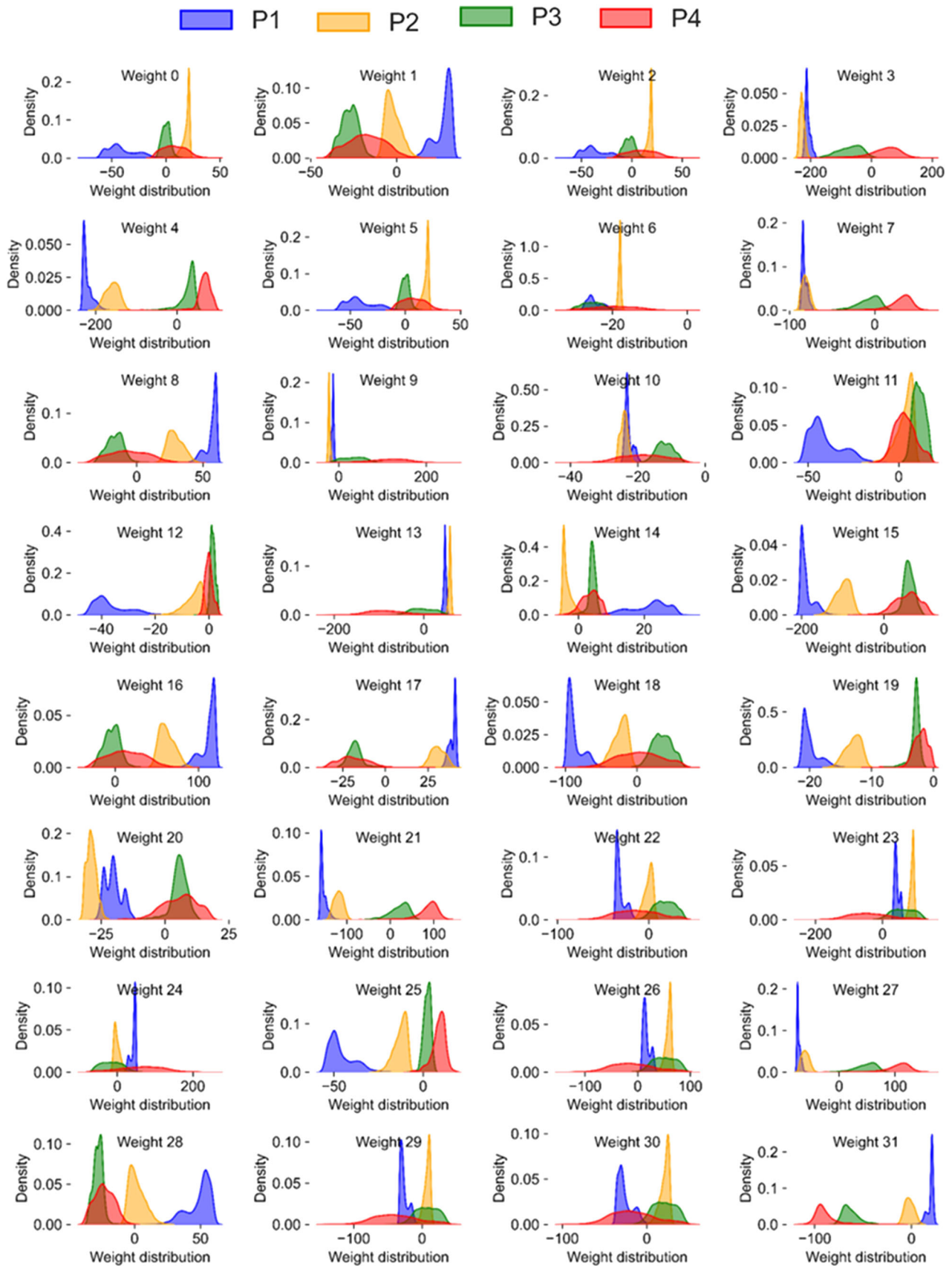


Fig. 16 Distribution analysis of lower-dimensional representations computed using *TF-3* on four grades of process zone images from the projector ($g_{T\theta}$) (Color figure online)

The primary goal of this study has been to explore self-supervised learning within the context of co-axial images, focusing on representation learning for classifying process zone images with different dynamics. However, further investigation to achieve a more profound understanding of the model and enhance its explainability is pending, and this will be a part of our future work. This contribution presents a monitoring approach based on 3-channel images (i.e., RGB) that captures morphological information at a lower frame rate of 30 frames/s. However, examining the composition of melt pools and plumes using multispectral imagery at a higher frame rate would aid in realizing a comprehensive monitoring system and is a continuation of this work. Furthermore, the L-DED technique has tremendous promise in printing functionally graded parts where different powder compositions are alloyed and modified in situ. Monitoring such multi-material AM printing is one of the work's following directions.

Acknowledgements The authors would like to thank Dr. Patrik Hoffmann for his input on the original version of the manuscript.

Funding Open Access funding provided by Lib4RI – Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL.

Data availability The datasets and source code utilized in this study are readily available and can be accessed through the repositories hosted at <https://c4science.ch/diffusion/12521/>.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abe, N., Tanigawa, D., Tsukamoto, M., Hayashi, Y., Yamazaki, H., Tatsumi, Y., & Yoneyama, M. (2013). Dynamic observation of formation process in laser cladding using high speed video camera. In *International congress on applications of lasers and electro-optics*, 2013. <https://doi.org/10.2351/1.5062915>
- Afouras, T., Asano, Y. M., Fagan, F., Vedaldi, A., & Metze, F. (2022). Self-supervised object detection from audio-visual correspondence. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022. <https://doi.org/10.48550/arXiv.2104.06401>
- Bartkowiak, K. (2010). Direct laser deposition process within spectrographic analysis in situ. *Physics Procedia*, 5, 623–629. <https://doi.org/10.1016/j.phpro.2010.08.090>
- Bi, G., Gasser, A., Wissenbach, K., Drenker, A., & Poprawe, R. (2006a). Identification and qualification of temperature signal for monitoring and control in laser cladding. *Optics and Lasers in Engineering*, 44(12), 1348–1359. <https://doi.org/10.1016/j.optlaseng.2006.01.009>
- Bi, G., Gasser, A., Wissenbach, K., Drenker, A., & Poprawe, R. (2006b). Investigation on the direct laser metallic powder deposition process via temperature measurement. *Applied Surface Science*, 253(3), 1411–1416. <https://doi.org/10.1016/j.apsusc.2006.02.025>
- Bi, G., Schürmann, B., Gasser, A., Wissenbach, K., & Poprawe, R. (2007). Development and qualification of a novel laser-cladding head with integrated sensors. *International Journal of Machine Tools and Manufacture*, 47(3–4), 555–561. <https://doi.org/10.1016/j.ijmachtools.2006.05.010>
- Bond, L. J., Koester, L. W., & Taheri, H. (2019). NDE in-process for metal parts fabricated using powder based additive manufacturing. In *Smart structures and NDE for energy systems and Industry 4.0*, 2019.
- Bustillo, A., Urbikain, G., Perez, J. M., Pereira, O. M., & de Lacalle, L. N. L. (2018). Smart optimization of a friction-drilling process based on boosting ensembles. *Journal of Manufacturing Systems*, 48, 108–121. <https://doi.org/10.1016/j.jmsy.2018.06.004>
- Chen, L., Yao, X., Tan, C., He, W., Su, J., Weng, F., . . . , Moon, S. K. (2023). In situ crack and keyhole pore detection in laser directed energy deposition through acoustic signal and deep learning. *Additive Manufacturing*, 69, 103547. <https://doi.org/10.1016/j.addma.2023.103547>
- Devesse, W., De Baere, D., Hinderdael, M., & Guillaume, P. (2016). Hardware-in-the-loop control of additive manufacturing processes using temperature feedback. *Journal of Laser Applications*, 28(2), 022302. <https://doi.org/10.2351/1.4943911>
- Ding, Y., Warton, J., & Kovacevic, R. (2016). Development of sensing and control system for robotized laser-based direct metal addition system. *Additive Manufacturing*, 10, 24–35. <https://doi.org/10.1016/j.addma.2016.01.002>
- Dosovitskiy, A., Beyler, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . , Gelly, S. (2020). An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929). <https://doi.org/10.48550/arXiv.2010.11929>
- Doubenskaia, M., Bertrand, P., & Smurov, I. (2004). Optical monitoring of Nd:YAG laser cladding. *Thin Solid Films*, 453, 477–485. <https://doi.org/10.1016/j.tsf.2003.11.184>
- Dutta, B., Palaniswamy, S., Choi, J., Song, L., & Mazumder, J. (2011). Direct metal deposition. *Advanced Materials and Processes*, 33. <https://api.semanticscholar.org/CorpusID:100065561>
- Ertveldt, J., Guillaume, P., & Helsen, J. (2020). MiCLAD as a platform for real-time monitoring and machine learning in laser metal deposition. *Procedia CIRP*, 94, 456–461. <https://doi.org/10.1016/j.procir.2020.09.164>
- Esfahani, M. N., Bappy, M. M., Bian, L., & Tian, W. (2022). In situ layer-wise certification for direct laser deposition processes based on thermal image series analysis. *Journal of Manufacturing Processes*, 75, 895–902. <https://doi.org/10.1016/j.jmapro.2021.12.041>
- Felix, S., Ray Majumder, S., Mathews, H. K., Lexa, M., Lipsa, G., Ping, X., . . . , Spears, T. (2022). In situ process quality monitoring and defect detection for direct metal laser melting. *Scientific Reports*, 12(1), 1–8. <https://doi.org/10.1038/s41598-022-12381-4>
- Gaja, H., & Liou, F. (2018). Defect classification of laser metal deposition using logistic regression and artificial neural networks for pattern recognition. *The International Journal of Advanced Manufacturing Technology*, 94(1), 315–326. <https://doi.org/10.1007/s00170-017-0878-9>

- Gharbi, M., Peyre, P., Gorny, C., Carin, M., Morville, S., Le Masson, P., . . . , Fabbro, R. (2013). Influence of various process conditions on surface finishes induced by the direct metal deposition laser technique on a Ti–6Al–4V alloy. *Journal of Materials Processing Technology*, 213(5), 791–800. <https://doi.org/10.1016/j.jmatprotec.2012.11.015>
- González-Barrío, H., Calleja-Ochoa, A., De Lacalle, L. N. L., & Lamikiz, A. (2022). Hybrid manufacturing of complex components: Full methodology including laser metal deposition (LMD) module development, cladding geometry estimation and case study validation. *Mechanical Systems and Signal Processing*, 179, 109337. <https://doi.org/10.1016/j.ymssp.2022.109337>
- Gonzalez-Val, C., Pallas, A., Panadeiro, V., & Rodriguez, A. (2020). A convolutional approach to quality monitoring for laser manufacturing. *Journal of Intelligent Manufacturing*, 31(3), 789–795. <https://doi.org/10.1007/s10845-019-01495-8>
- Grill, J.-B., Strub, F., Althé, F., Tallec, C., Richemond, P., Buchatskaya, E., . . . , Gheshlaghi Azar, M. (2020). Bootstrap your own latent—A new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33, 21271–21284. <https://doi.org/10.48550/arXiv.2006.07733>
- Haley, J. C., Schoenung, J. M., & Lavernia, E. J. (2018). Observations of particle-melt pool impact events in directed energy deposition. *Additive Manufacturing*, 22, 368–374. <https://doi.org/10.1016/j.addma.2018.04.028>
- Hauser, T., Breese, P. P., Kamps, T., Heinze, C., Volpp, J., & Kaplan, A. F. (2020). Material transitions within multi-material laser deposited intermetallic iron aluminides. *Additive Manufacturing*, 34, 101242. <https://doi.org/10.1016/j.addma.2020.101242>
- Hauser, T., Reisch, R. T., Kamps, T., Kaplan, A. F., & Volpp, J. (2022). Acoustic emissions in directed energy deposition processes. *The International Journal of Advanced Manufacturing Technology*, 119(5), 3517–3532. <https://doi.org/10.1007/s00170-021-08598-8>
- Hendrycks, D., Mazeika, M., Kadavath, S., & Song, D. (2019). Using self-supervised learning can improve model robustness and uncertainty. In *Advances in Neural Information Processing Systems* (Vol. 32). <https://doi.org/10.5555/3454287.3455690>
- Hu, D., & Kovacevic, R. (2003). Sensing, modeling and control for laser-based additive manufacturing. *International Journal of Machine Tools and Manufacture*, 43(1), 51–60. [https://doi.org/10.1016/S0890-6955\(02\)00163-3](https://doi.org/10.1016/S0890-6955(02)00163-3)
- Hua, T., Jing, C., Xin, L., Fengying, Z., & Weidong, H. (2008). Research on molten pool temperature in the process of laser rapid forming. *Journal of Materials Processing Technology*, 198(1–3), 454–462. <https://doi.org/10.1016/j.jmatprotec.2007.06.090>
- Iravani-Tabrizipour, M., & Toyserkani, E. (2007). An image-based feature tracking algorithm for real-time measurement of clad height. *Machine Vision and Applications*, 18(6), 343–354. <https://doi.org/10.1007/s00138-006-0066-7>
- Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2020). A survey on contrastive self-supervised learning. *Technologies*, 9(1), 2. <https://doi.org/10.3390/technologies9010002>
- Khanzadeh, M., Chowdhury, S., Marufuzzaman, M., Tschopp, M. A., & Bian, L. (2018a). Porosity prediction: Supervised-learning of thermal history for direct laser deposition. *Journal of Manufacturing Systems*, 47, 69–82. <https://doi.org/10.1016/j.jmsy.2018.04.001>
- Khanzadeh, M., Chowdhury, S., Tschopp, M. A., Doude, H. R., Marufuzzaman, M., & Bian, L. (2019). In situ monitoring of melt pool images for porosity prediction in directed energy deposition processes. *IIEE Transactions*, 51(5), 437–455. <https://doi.org/10.1080/24725854.2017.1417656>
- Khanzadeh, M., Tian, W., Yadollahi, A., Doude, H. R., Tschopp, M. A., & Bian, L. (2018b). Dual process monitoring of metal-based additive manufacturing using tensor decomposition of thermal image streams. *Additive Manufacturing*, 23, 443–456. <https://doi.org/10.1016/j.addma.2018.08.014>
- Kim, H., Cong, W., Zhang, H.-C., & Liu, Z. (2017). Laser engineered net shaping of nickel-based superalloy Inconel 718 powders onto AISI 4140 alloy steel substrates: Interface bond and fracture failure mechanism. *Materials*, 10(4), 341. <https://doi.org/10.3390/ma10040341>
- Koester, L. W., Taheri, H., Bigelow, T. A., Bond, L. J., & Faierson, E. J. (2018). In situ acoustic signature monitoring in additive manufacturing processes. *AIP Conference Proceedings*. <https://doi.org/10.1063/1.5031503>
- Lednev, V., Tretyakov, R., Sdvizhenskii, P., Grishin, M. Y., Asyutin, R., & Pershin, S. (2018). Laser induced breakdown spectroscopy for in situ multielemental analysis during additive manufacturing process. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1109/1/012050>
- Lee, H., Heogh, W., Yang, J., Yoon, J., Park, J., Ji, S., & Lee, H. (2022). Deep learning for in situ powder stream fault detection in directed energy deposition process. *Journal of Manufacturing Systems*, 62, 575–587. <https://doi.org/10.1016/j.jmsy.2022.01.013>
- Lei, J. B., Wang, Z., & Wang, Y. S. (2012). Measurement on temperature distribution of metal powder stream in laser fabricating. *Applied Mechanics and Materials*. <https://doi.org/10.4028/www.scientific.net/AMM.101-102.994>
- Li, X., Siahpour, S., Lee, J., Wang, Y., & Shi, J. (2020). Deep learning-based intelligent process monitoring of directed energy deposition in additive manufacturing with thermal images. *Procedia Manufacturing*, 48, 643–649. <https://doi.org/10.1016/j.promfg.2020.05.093>
- Lison, M., Devesse, W., de Baere, D., Hinderdael, M., & Guillaume, P. (2019). Hyperspectral and thermal temperature estimation during laser cladding. *Journal of Laser Applications*, 31(2), 022313. <https://doi.org/10.2351/1.5096129>
- Liu, M., Kumar, A., Bukkapatnam, S., & Kuttolamadom, M. (2021a). A review of the anomalies in directed energy deposition (DED) Processes & Potential Solutions—Part quality & defects. *Procedia Manufacturing*, 53, 507–518. <https://doi.org/10.1016/j.promfg.2021.06.093>
- Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021b). Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.48550/arXiv.2006.08218>
- Mazumder, J. (2015). Design for metallic additive manufacturing machine with capability for “Certify as You Build.” *Procedia CIRP*, 36, 187–192. <https://doi.org/10.1016/j.procir.2015.01.009>
- Meriaudeau, F., & Truchetet, F. (1996). Control and optimization of the laser cladding process using matrix cameras and image processing. *Journal of Laser Applications*, 8(6), 317–324. <https://doi.org/10.2351/1.4745438>
- Mitash, C., Bekris, K. E., & Boularias, A. (2017). A self-supervised learning system for object detection using physics simulation and multi-view pose estimation. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2017. <https://doi.org/10.48550/arXiv.1703.03347>
- Montazeri, M., Nassar, A. R., Stutzman, C. B., & Rao, P. (2019). Heterogeneous sensor-based condition monitoring in directed energy deposition. *Additive Manufacturing*, 30, 100916. <https://doi.org/10.1016/j.addma.2019.100916>
- Nassar, A., Starr, B., & Reutzel, E. (2015). Process monitoring of directed-energy deposition of Inconel-718 via plume imaging. In *2014 International solid freeform fabrication symposium*, 2015. <https://pure.psu.edu/en/publications/process-monitoring-of-directed-energy-deposition-of-inconel-718-v>
- Pandiyani, V., Cui, D., Le-Quang, T., Deshpande, P., Wasmer, K., & Shevchik, S. (2022a). In situ quality monitoring in direct energy deposition process using co-axial process zone imaging and deep

- contrastive learning. *Journal of Manufacturing Processes*, 81, 1064–1075. <https://doi.org/10.1016/j.jmapro.2022.07.033>
- Pandiyan, V., Cui, D., Parrilli, A., Deshpande, P., Masinelli, G., Shevchik, S., & Wasmer, K. (2022b). Monitoring of direct energy deposition process using manifold learning and co-axial melt pool imaging. *Manufacturing Letters*, 33, 776–785. <https://doi.org/10.1016/j.mfglet.2022.07.096>
- Pandiyan, V., Masinelli, G., Claire, N., Le-Quang, T., Hamidi-Nasab, M., de Formanoir, C., . . . , Logé, R. (2022). Deep learning-based monitoring of laser powder bed fusion process on variable time-scales using heterogeneous sensing and operando X-ray radiography guidance. *Additive Manufacturing*, 58, 103007. <https://doi.org/10.1016/j.addma.2022.103007>
- Pandiyan, V., Wróbel, R., Axel Richter, R., Leparoux, M., Leinenbach, C., & Shevchik, S. (2023). Self-Supervised Bayesian representation learning of acoustic emissions from laser powder bed Fusion process for in situ monitoring. *Materials and Design*. <https://doi.org/10.1016/j.matdes.2023.112458>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . , Antiga, L. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*. <https://doi.org/10.48550/arXiv.1912.01703>
- Pérez-Ruiz, J., González-Barrio, H., Sanz-Calle, M., Gómez-Escudero, G., Munoa, J., & de Lacalle, L. L. (2023). Machining stability improvement in LPBF printed components through stiffening by crystallographic texture control. *CIRP Annals*. <https://doi.org/10.1016/j.cirp.2023.03.025>
- Ren, W., Wen, G., Zhang, Z., & Mazumder, J. (2022). Quality monitoring in additive manufacturing using emission spectroscopy and unsupervised deep learning. *Materials and Manufacturing Processes*, 37(11), 1339–1346. <https://doi.org/10.1080/10426914.2021.1906891>
- Ribeiro, K. S. B., Núñez, H. H. L., Venter, G. S., Doude, H. R., & Coelho, R. T. (2023). A hybrid machine learning model for in-process estimation of printing distance in laser Directed Energy Deposition. <https://doi.org/10.1007/s00170-023-11582-z>
- Sampson, R., Lancaster, R., Sutcliffe, M., Carswell, D., Hauser, C., & Barras, J. (2020). An improved methodology of melt pool monitoring of direct energy deposition processes. *Optics and Laser Technology*, 127, 106194. <https://doi.org/10.1016/j.optlas.2020.106194>
- Schmidt, M., Huke, P., Gerhard, C., & Partes, K. (2021). In-line observation of laser cladding processes via atomic emission spectroscopy. *Materials*, 14(16), 4401. <https://doi.org/10.3390/ma14164401>
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 Years of image analysis. *Nature Methods*, 9(7), 671–675. <https://doi.org/10.1038/nmeth.2089>
- Smurov, I., Doubenskaia, M., Grigoriev, S., & Nazarov, A. (2012). Optical monitoring in laser cladding of Ti6Al4V. *Journal of Thermal Spray Technology*, 21(6), 1357–1362. <https://doi.org/10.1007/s11666-012-9808-4>
- Song, L., Bagavath-Singh, V., Dutta, B., & Mazumder, J. (2012a). Control of melt pool temperature and deposition height during direct metal deposition process. *The International Journal of Advanced Manufacturing Technology*, 58(1), 247–256. <https://doi.org/10.1007/s00170-011-3395-2>
- Song, L., Huang, W., Han, X., & Mazumder, J. (2016). Real-time composition monitoring using support vector regression of laser-induced plasma for laser additive manufacturing. *IEEE Transactions on Industrial Electronics*, 64(1), 633–642. <https://doi.org/10.1109/TIE.2016.2608318>
- Song, L., & Mazumder, J. (2011). Real time Cr measurement using optical emission spectroscopy during direct metal deposition process. *IEEE Sensors Journal*, 12(5), 958–964. <https://doi.org/10.1109/JSEN.2011.2162316>
- Song, L., Wang, C., & Mazumder, J. (2012). Identification of phase transformation using optical emission spectroscopy for direct metal deposition process. In *High power laser materials processing: Lasers, beam delivery, diagnostics, and applications*. <https://doi.org/10.1117/12.908264>
- Stutzman, C. B., Nassar, A. R., & Reutzler, E. W. (2018). Multi-sensor investigations of optical emissions and their relations to directed energy deposition processes and quality. *Additive Manufacturing*, 21, 333–339. <https://doi.org/10.1016/j.addma.2018.03.017>
- Tian, Q., Guo, S., & Guo, Y. (2020). A physics-driven deep learning model for process-porosity causal relationship and porosity prediction with interpretability in laser metal deposition. *CIRP Annals*, 69(1), 205–208. <https://doi.org/10.1016/j.cirp.2020.04.049>
- Toyserkani, E., & Khajepour, A. (2006). A mechatronics approach to laser powder deposition process. *Mechatronics*, 16(10), 631–641. <https://doi.org/10.1016/j.mechatronics.2006.05.002>
- Tung, H.-Y., Tung, H.-W., Yumer, E., & Fragkiadaki, K. (2017). Self-supervised learning of motion capture. In *Advances in neural information processing systems* (Vol. 30). <https://doi.org/10.5555/3295222.3295276>
- Vandone, A., Baraldo, S., Valente, A., & Mazzucato, F. (2019). Vision-based melt pool monitoring system setup for additive manufacturing. *Procedia CIRP*, 81, 747–752. <https://doi.org/10.1016/j.procir.2019.03.188>
- Vykhtar, B., & Richter, A. M. (2021). Optical monitoring sensor system for laser-based directed energy deposition. In *Lasers in manufacturing conference*, 2021.
- Wang, F., Mao, H., Zhang, D., Zhao, X., & Shen, Y. (2008). Online study of cracks during laser cladding process based on acoustic emission technique and finite element analysis. *Applied Surface Science*, 255(5), 3267–3275. <https://doi.org/10.1016/j.apsusc.2008.09.039>
- Wasmer, K., Wüst, M., Cui, D., Masinelli, G., Pandiyan, V., & Shevchik, S. (2023). Monitoring of functionally graded material during laser directed energy deposition by acoustic emission and optical emission spectroscopy using artificial intelligence. *Virtual and Physical Prototyping*, 18(1), e2189599. <https://doi.org/10.1080/17452759.2023.2189599>
- Whiting, J., Springer, A., & Sciammarella, F. (2018). Real-time acoustic emission monitoring of powder mass flow rate for directed energy deposition. *Additive Manufacturing*, 23, 312–318. <https://doi.org/10.1016/j.addma.2018.08.015>
- Xia, C., Pan, Z., Polden, J., Li, H., Xu, Y., Chen, S., & Zhang, Y. (2020). A review on wire arc additive manufacturing: Monitoring, control and a framework of automated system. *Journal of Manufacturing Systems*, 57, 31–45. <https://doi.org/10.1016/j.jmsy.2020.08.008>
- Zhang, B., Liu, S., & Shin, Y. C. (2019). In-Process monitoring of porosity during laser additive manufacturing process. *Additive Manufacturing*, 28, 497–505. <https://doi.org/10.1016/j.addma.2019.05.030>
- Zhang, Y., Shen, S., Li, H., & Hu, Y. (2022). Review of in situ and real-time monitoring of metal additive manufacturing based on image processing. *The International Journal of Advanced Manufacturing Technology*, 123(1–2), 1–20. <https://doi.org/10.1007/s00170-022-10178-3>
- Zhang, Z., Liu, Z., & Wu, D. (2020). Prediction of melt pool temperature in directed energy deposition using machine learning. *Additive Manufacturing*. <https://doi.org/10.1016/j.addma.2020.101692>
- Zhang, Z., Liu, Z., & Wu, D. (2021). Prediction of melt pool temperature in directed energy deposition using machine learning. *Additive Manufacturing*, 37, 101692. <https://doi.org/10.1016/j.addma.2020.101692>
- Zhou, M., Bai, Y., Zhang, W., Zhao, T., & Mei, T. (2020). Look-into-object: Self-supervised structure modeling for object recognition. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020 (pp. 11774–11783). <https://doi.org/10.48550/arXiv.2003.14142>

Zijue, T., Weiwei, L., Zhaorui, Y., Hao, W., & Hongchao, Z. (2019). Study on evolution behavior of geometrical accuracy based on dynamic characteristics of molten pool in laser-based direct energy deposition. *Jixie Gongcheng Xuebao Chinese Journal of Mechanical Engineering*, 55, 39. <https://doi.org/10.3901/JME.2019.15.039>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.