Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 124 (2024) 793–796

www.elsevier.com/locate/procedia

13th CIRP Conference on Photonic Technologies [LANE 2024], 15-19 September 2024, Fürth, Germany

In-situ monitoring and online prediction of keyhole depth in laser welding by coaxial imaging

Henrique H. L. Núñez[a], Li-Wei Hsu[a], Kandice S. B. Ribeiro[a,*], Antti Salminen[a], Wallace M. Bessa[a]

[a]*Department of Mechanical and Materials Engineering, University of Turku, 20520 Turku, Finland*

* Corresponding author. Tel.: +358-0503-081-750. *E-mail address:* ksbari@utu.fi

Abstract

A comprehensive understanding of welding penetration and the role of process parameters is crucial for ensuring high-quality joints in laser welding. In-situ process monitoring can aid in detection of defects, reducing material usage and time-consuming inspection operations. In this study, we propose a novel approach for online prediction of keyhole depth in laser welding operations. Using in-process images captured with a coaxial camera and active illumination, our software employs pre-trained CNNs from the EfficientNet and DenseNet families to extract features. These features serve as input for data-efficient regression models, trained to predict the keyhole depth. The results have shown that both methods yield percentage errors of approximately 3%. Ultimately, this methodology facilitates real-time analysis of welding operations.

Keywords: laser welding; meltpool monitoring; computer vision; optical coherence tomography.

1. Introduction

Laser beam welding (LBW) is one of the joining processes that uses a concentrated heat source such as laser beam to melt materials and join structures as the melt pool solidifies. Its high energy density and operational versatility makes it a process that has been adopted increasingly in the manufacture of large metallic structures and shipbuilding [1].

Typically, when mounted in a robotic arm, it is one of the processes with extensive research in the field of digital manufacturing, as it allows for computer parametrized control. Following the current trend of digital manufacturing to assure a good quality, one way is to reduce the intervention of humans in the process and perform parameter value optimization and monitoring. In process monitoring, several sensors have been studied in the literature, with special attention to cameras [2]. A multi sensor monitoring approach using camera, Optical Coherence Tomography (OCT) and optical microphone data has been proposed [3].

Recent advances in Artificial Intelligence (AI), in particular the ever-growing Deep Learning (DL) models, both in terms of application and model performance, combined with a concurrent increase of computing availability and hardware acceleration, makes AI highly attractive for the field [4].

Coaxial process images have been used in the monitoring of other laser-based processes such as Directed Energy Deposition [5], where images from the melt pool were used to predict the focal distance between the machine head and the manufactured part. Another application was conducted with the same setup to estimate the final clad width [6].

Monitoring the keyhole behavior during the welding process is of great importance given the complex thermal interaction that takes place during the process [7]. The lack of penetration in the laser welding process can cause severe structural defects, compromising the whole structural properties and therefore generating more costs due to the time of rework and additional material spent in the manufacturing [8].

In this study, we propose a camera-based monitoring system that predicts the keyhole depth during LBW, composed by a

10.1016/j.procir.2024.08.227

feature extraction step followed by a supervised learning regression model. This work brings a new perspective towards the estimation of keyhole depth during laser welding processes, through the monitoring of melt pool images. Once validated, this method can simplify in-situ monitoring and provide a more economical solution. In the industry, the combination of visual inspection with estimated keyhole depth can make the monitoring process more robust and efficient.

Nomenclature

CNN	Convolutional Neural Network
OCT	Optical Coherence Tomography
KHD	Keyhole Depth
LBW	Laser Beam Welding

2. Methodology

Supervised machine learning algorithms require a label attributed to each example in the input data, so that the parameter optimization can happen. In this work, the input of the system were images from the melt pool during laser welding, while the labels are attributed by a coupled OCT system. An outline of the software methodology is presented in Figure 1.

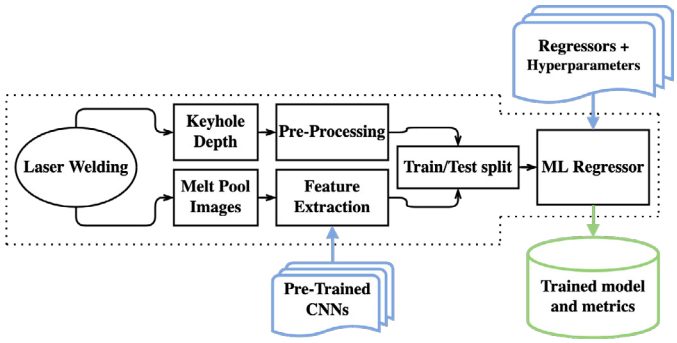

Fig. 1. Diagram of the data flow for model evaluation.

Starting with the laser welding, the keyhole depth and melt pool images were acquired and synchronized. Each image received a weld coordinate based on marks in the test body, frame index and acquisition rate. The collected OCT data was then interpolated with respect to the weld coordinate, using linear interpolation of the closest points. Then, feature extraction is performed using a pre-trained CNN, generating a feature vector per input image. These sets of pairs of feature vector and KHD were then divided into "Train" and "Test" subsets of data in a 3-fold cross validation, with each fold containing 20% of the dataset as test values. Training and evaluating the model are the next step, which results in a trained model and associated performance metrics. The described data processing is fixed, surrounded in the diagram by a dashed line. The remaining parameters were explored using a grid-search approach. This model evaluation pipeline enables the analysis of different combinations of feature extractors and regressors, and by analyzing the search space of "pre-trained CNNs" and "Regressors and Hyperparameters", it is possible to determine the combination with highest performance.

The experiments were conducted in a robot arm laser system equipped with an IPG YLS-10000S multimode fiber laser with a maximum laser power of 10kW. In terms of optics, the fiber core diameter was 300µm, collimation lens 150mm and focusing lens focal length 348mm. The welding speed was of 50 mm/s, and the test bodies were 5mm thick ASTM A36 plates. The 10 sets of 360 mm welds were performed with different laser powers, as stated in Table 1. The data acquisition was composed by two main sensors, an OCT IPG LDD-700 Inline Weld Monitor device, and one welding camera from Cavitar, model C400, both mounted coaxially with the welding laser beam. A preprocessing step was performed to select the data points corresponding only to the instants in which the welding process was active when the laser was operating.

Table 1. Sets of parameters for the experiments.

Weld	Laser Power (kW)	Weld (cont.)	Laser Power (kW)
1	4.6	6	5.1
2	4.7	7	5.2
3	4.8	8	5.3
4	4.9	9	5.4
5	5	10	5.5

Regarding the feature extraction CNN models, two different models were used: "DenseNet169" and "EfficientNetB6". The specific implementation was chosen based on the reported scores achieved by those networks on the ImageNet classification benchmark. Although the aforementioned strategies have inner peculiarities, from a top-level perspective both are composed by a sequence of blocks that perform image operations such as convolutions and pooling, handing the result to a final dense layer that performs image classification.

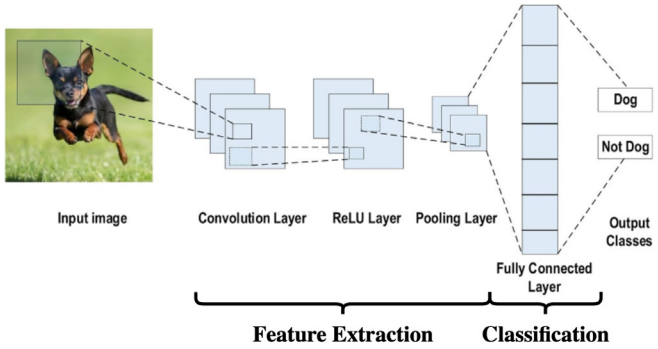

Fig. 2. Emphasis on the "Feature Extraction" and "Classification" sub-components of an abstract CNN architecture. Adapted from [4].

The transformation of a general image classification CNN into a generic feature extractor is performed by getting an intermediate result computed by the network, before being passed to the densely connected classification layer. A visual example of the different steps is shown in Figure 2.

For the present work, the pre-trained weights were acquired from training against the "ImageNet" dataset, which is an object classification dataset that provides images of 1000 different classes. This large set of features is then used by regression models that require less training data, as has been

demonstrated in other studies with different application domains [5].

The selection of regression models used in this study was based on low training data requirements and satisfactory performance. Beyond the algorithm itself, each regression model is tunable by a set of hyperparameters that can be optimized to improve model generalization and overall performance. In total, five regression algorithms were explored: Linear Regression (LR), with default parameters, as a baseline model, Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), Elastic Net Regressor (ENR) and Support Vector Regressor (SVR). The implementation was from the scikit-learn library, version 1.4.1.

The GBR and RFR are models composed of a specified number of trees that optimize for different sets of input variables. The searched hyperparameters for were:

- Number of estimators: 100, 250, 350.
- Maximum estimator depth: 5, 10 and default strategy.
- Learning rate (specific to GBR): 0.1 and 0.01.
- Minimum samples to split (specific to RFR): 2, 5 and 10.

In SVR, the used kernel was Radial Basis Function, and the regularization constant values were 100, 1000, 1500, 3000, 3500. As for the ENR, the "alpha" parameter possibilities were 1.0, 0.1, 0.01, 0.001, and "L1 ratio" were 0.1, 0.5, 0.9.

The metrics used to validate the performance of the model are MAE, MAPE and R^2. MAE and MAPE are metrics that work by subtracting the predicted value from the ground truth value, and the difference is that the former averages those differences, while the latter normalizes the differences by the ground truth data again. The formulations for those metrics are present in the Equations (1) and (2), where y and $\hat{y}$ represent, respectively, the ground truth and model predictions data vectors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i}\cdot 100\% \tag{2}$$

Another metric used to analyze regression performance is R^2, and its interpretation is the variance of the predicted variable that is captured by the regression model. A R^2 of 1 indicates the best possible fit of the model to the data, while a value of 0 means that the model does not explain the data, or that the input data itself is not enough to explain the target variable. The formulation for the R^2 metric is shown in Equation 3, where $\bar{y}$ is the average of the target values, y_i is the i-th label and $\hat{y}_i$ is the prediction for the i-th input.

$$R^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)}{\sum_i \left(y_i - \bar{y}\right)} \tag{3}$$

As an additional evaluation strategy, a visual analysis was conducted in order to verify generalization across the whole interval of labels. The test data was used to generate a scatter plot between the true and predicted values, where the best possible outcome is the $y = x$ line, also plotted, as well as the $\pm 5\%$ error lines for guidance.

Implementation-wise, TensorFlow/Keras was chosen as the DL framework to instance and run the pre-trained CNNs. All the steps carried out in the methodology were conducted using Python 3.11.

3. Results and discussion

All the sets of proposed experiments were collected, and the data was processed and analyzed according to the methodology. The set of experiments resulted in a dataset of approximately 430 images per weld, aggregating 5608 images of the keyhole. Two examples of coaxial images captured in the setup are depicted in Figure 3, with corresponding KHD of 3.92mm and 4.49mm. In the Figure, contrast and brightness correction were made to improve the visualization quality.

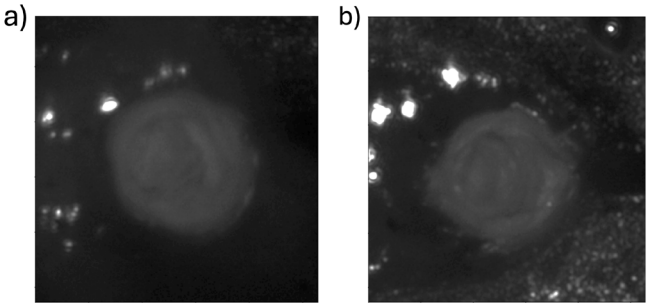

Fig. 3. Example of captured coaxial image for a KHD of a) 3.92mm and b) 4.49mm.

After exploring the different feature extraction and regression strategies, the best set of parameters for each regressor and feature set was organized in Table 2. The model code is formed by a prefix denoting the CNN model and a suffix denoting the regression model. The execution time per image was of approximately 30ms in a personal computer, translating into a processing capability of approximately 30fps.

Table 2. Best hyperparameters for each combination of models.

Model	Parameters
EN_GBR	Learning Rate: 0.1; Max. Depth: 5; N. Estimators: 350
EN_RFR	Min. Samples Split: 2; Max. Depth: auto; N. Estimators: 350
EN_ENR	Alpha: 0.1; L1 Ratio: 0.5
EN_SVR	C: 1000
DN_GBR	Learning Rate: 0.1; Max. Depth: 5; N. Estimators: 250
DN_RFR	Min. Samples Split: 5; Max. Depth: auto; N. Estimators: 350
DN_ENR	Alpha: 0.1; L1 Ratio: 0.1
DN_SVR	C: 1000

Alongside the best parameters for each model, the resulting metrics after performing the cross-validation step were organized in Table 3, with best and second-best model scores in bold.

As it is possible to analyze in Table 3, the best model was DenseNet169 combined with ENR, achieving a R^2 score of

0.59, a MAE of 130.78µm and a MAPE of 2.89%. Additionally, DenseNet169 combined with SVR has also demonstrated a comparable performance across the different assessed metrics, with 0.59, 131.40µm and 2.9% of R^2, MAE and MAPE respectively.

Table 3. Computed metrics for the combinations of models and feature extractors.

Model Code	MAE (µm)	MAPE (%)	R^2
EN_LR	135.21	3.00	0.56
EN_GBR	148.86	3.30	0.47
EN_RFR	162.60	3.61	0.38
EN_ENR	134.44	2.98	0.57
EN_SVR	138.77	3.08	0.54
DN_LR	131.93	2.92	0.58
DN_GBR	148.31	3.28	0.48
DN_RFR	163.82	3.63	0.37
DN_ENR	**130.78**	**2.89**	**0.59**
DN_SVR	**131.40**	**2.9**	**0.59**

An error of approximately 3% would be considered good for most applications, while a R^2 score in the achieved range is an indicator that the model either does not capture variance in the dataset, or that the data itself does not fully explain the label. As a general trend, GBR and RFR had a substantially poorer performance than SVR and ENR, regardless of the feature extraction methods.

In addition to the regression metrics, Figure 4 illustrates a dispersion plot of the predictions compared to the ground truth data. The blue dashed line represents an ideal reference, while the red lines represent a 5% error. Visually, the majority of the points used in the test phase are withing the 5% deviation interval. The points that deviate from this interval do not exhibit a particular distribution across the labels of the dataset.

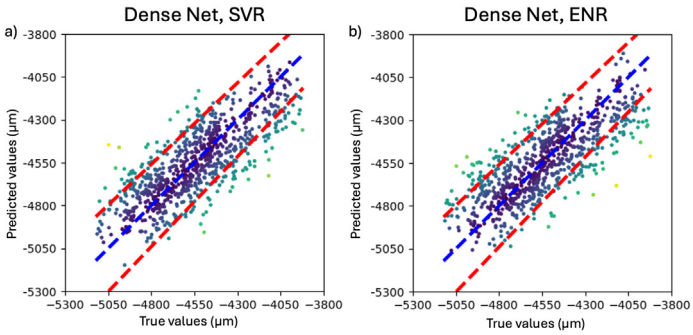

Fig. 4. Predicted and true labels using DenseNet with a) SVR and b) ENR.

Since this methodology is based on supervised learning, the performance of the model depends on the quality of the training data. Model robustness can be increased with the addition of outlier removal steps, for example.

4. Conclusions

This study demonstrated evidence for the success of laser welding monitoring by coaxial cameras and image processing to estimate the keyhole depth. In this regard, a supervised learning approach was proposed and employed in the study.

The data processing setup was composed of a pre-trained CNN feature extractor in tandem with a conventional Machine Learning regression model. The target variable was acquired by an OCT sensor attached to the robotic arm, coaxial with respect to the laser. Among the different approaches, the one yielding the best metrics used DenseNet169 for feature extraction with ENR as the regression model, achieving R^2, MAE and MAPE of 0.59, 130.78µm and 2.9%, demonstrating a satisfactory absolute error.

Additionally, the processing time implies that this method can be deployed without the need of special computers. Further optimization can be achieved by using a low-level implementation and specialized hardware.

The relevance of a camera-based monitoring system confirms the possibility to use a sensor commonly found in laser welding setups that shares the same optics as the laser. Additionally, the abundance of image-based monitoring data can provide a robust foundation for conducting various types of simultaneous operational analyses, enabling the use of the same sensor for a different set of measurements. As a follow up of this work, we suggest the extension of this methodology for a different set of welds with different materials and geometries in order to validate the generalization of this approach. Another continuation would be to conduct a large number of experiments enabling the development of a representative and curated dataset, allowing larger and more sophisticated DL models.

Acknowledgements

This work is part of the CaNeLis (Carbon-neutral lightweight ship structures using advanced design, production, and life-cycle services) project co-funded by Business Finland (3360/31/2022) and NEcOLEAP. The authors would also like to thank our industrial partners Meyer Turku Oy, Cavitar Oy and SSAB Oyj.

References

[1] Katayama S, Handbook of Laser Welding Technologies, 2013, pp. 3-16.
[2] Stavridis J, Papacharalampopoulos A, Stavropoulos P. Quality assessment in laser welding: a critical review. Int J Adv Manuf Technol 94: 1825–1847; 2018.
[3] Hsu LW, Salminen A. Laser welding monitoring with multisensory data fusion: A brief review. IOP Conf. Ser.: Mater. Sci. Eng., vol. 1296, no. 1, p. 012014; 2023.
[4] Alzubaidi L et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, vol. 8, no. 1, p. 53; 2021.
[5] Ribeiro KSB, Núñez HHL, Venter GS et al. A hybrid machine learning model for in-process estimation of printing distance in laser Directed Energy Deposition. Int J Adv Manuf Technol 127, 3183–3194; 2023.
[6] Mochi, V.H., Núñez, H.H.L., Ribeiro, K.S.B. et al. Real-time prediction of deposited bead width in L-DED using semi-supervised transfer learning. Int J Adv Manuf Technol 129: 5643–5654; 2023.
[7] Kawahito Y, Wang H. In-situ observation of gap filling in laser butt welding, Scripta Materialia, Volume 154: 73-77; 2018.
[8] Roland F, Manzon L, Kujala P, Brede M, et al. Advanced Joining Techniques in European Shipbuilding. J Ship Prod 20: 200–210; 2004.