Journal of Laser Applications

RESEARCH ARTICLE | AUGUST 26 2024

Analyzing multispectral emission and synchrotron data to evaluate the quality of laser welds on copper

Special Collection: Proceedings of the International Congress of Applications of Lasers & Electro-Optics (ICALEO 2024)

Jan Brüggenjürgen ; Christoph Spurk ; Marc Hummel ; Christoph Franz; Andrè Häusler ; Alexander Olowinsky ; Felix Beckmann ; Julian Moosmann

Check for updates

J. Laser Appl. 36, 032032 (2024)
https://doi.org/10.2351/7.0001600

View Online
Export Citation

Articles You May Be Interested In

Development workflow based on *in situ* synchrotron investigations to optimize laser processing of copper pins

J. Laser Appl. (September 2024)

Analysis of laser beam welding with superimposed 445 and 1070 nm wavelength lasers on copper by *in situ* synchrotron diagnostics

J. Laser Appl. (September 2024)

Application of a physics-informed hybrid model with additional output constraints for the prediction of the threshold of deep-penetration laser welding

J. Laser Appl. (February 2025)

12 February 2025 11:49:34

ALIA THE LASER INSTITUTE Journal of Laser Applications
Learn More
RAPID TIME TO ACCEPTANCE
COMMUNITY DRIVEN
EXPANSIVE COVERAGE
PRESTIGIOUS EDITORIAL BOARD
EXTENSIVE MARKETING

ALIA THE LASER INSTITUTE

Journal of Laser Applications ARTICLE pubs.aip.org/lia/jla

Analyzing multispectral emission and synchrotron data to evaluate the quality of laser welds on copper

Cite as: J. Laser Appl. **36**, 032032 (2024); doi: 10.2351/7.0001600
Submitted: 1 July 2024 · Accepted: 7 August 2024 ·
Published Online: 26 August 2024

View Online Export Citation CrossMark

Jan Brüggenjürgen,[1,2,a)] Christoph Spurk,[2] Marc Hummel,[2] Christoph Franz,[3] Andrè Häusler,[1] Alexander Olowinsky,[1] Felix Beckmann,[4] and Julian Moosmann[4]

AFFILIATIONS

[1]Fraunhofer Institute for Laser Technology ILT, Steinbachstraße 15, Aachen 52074, Germany
[2]Chair for Laser Technology LLT, RWTH Aachen University, Steinbachstraße 15, Aachen 52074, Germany
[3]4D Photonics GmbH, Burgwedeler Str. 27a, Isernhagen 30916, Germany
[4]Institute of Materials Physics, Helmholtz-Zentrum Hereon, Max-Planck-Str.1, Geesthacht 21502, Germany

Note: Paper published as part of the special topic on Proceedings of the International Congress of Applications of Lasers & Electro-Optics 2024.

[a)]**Author to whom correspondence should be addressed; electronic mail:** jan.brueggenjuergen@ilt.fraunhofer.de

12 February 2025 11:49:34

ABSTRACT

The validation of laser welding of metallic materials is challenging due to its highly dynamic processes and limited accessibility to the weld. The measurement of process emissions and the processing laser beam are one way to record highly dynamic process phenomena. However, these recordings always take place via the surface of the weld. Phenomena on the inside are only implicitly recognizable in the data and require further processing. To increase the validity of the diagnostic process, the multispectral emission data are synchronized with synchrotron data consisting of *in situ* high-speed images based on phase contrast videography. The welding process is transilluminated by synchrotron radiation and recorded during execution, providing clear contrasts between solid, liquid, and gaseous material phases. Thus, dynamics of the vapor capillary and the formation of defects such as pores can be recorded with high spatial and temporal resolution of $<5\,\mu$m and >5 kHz. In this paper, laser welding of copper Cu-ETP and CuSn6 is investigated at the Deutsches Elektronen-Synchrotron (DESY). The synchronization is achieved by leveraging a three-stage deep learning approach. A preprocessing Mask-R-CNN, dimensionality reduction PCA/Autoencoders, and a final LSTM/Transformer stage provide end-to-end defect detection capabilities. Integrated gradients allow for the extraction of correlations between defects and emission data. The novel approach of correlating image and sensor data increases the informative value of the sensor data. It aims to characterize welds based on the sensor data not only according to IO/NIO but also to provide a quantitative description of the defects in the weld.

Key words: deep Learning, multi-spectral, emission, sensor, laser, welding, copper, quality control, synchrotron

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). https://doi.org/10.2351/7.0001600

I. INTRODUCTION

Sensor data play a pivotal role in numerous manufacturing domains by enabling comprehensive monitoring, analysis, and optimization of complex systems and processes. One such domain is laser-based metal processing, where precise control of parameters such as laser power, beam position, and welding speed is essential to achieve high-quality parts. These integrated data provide valuable insights into the dynamics and interactions occurring during welding, facilitating the identification of optimal process settings and potential anomalies. Once fine-tuned, the process can be reproduced without high-cost *in situ* supervision, which, in turn, can drive down cost and complexity further.

Careful consideration must be given to strike a balance between the number of sensors, their cost, and the overall effectiveness in

© Author(s) 2024

detecting anomalies. Factors such as the complexity of the process, desired speeds compared to amounts of data, the level of accuracy required, and the financial constraints of the plant or facility play pivotal roles in determining the optimal sensor selection strategy.

While a blanket approach into fitting all process-relevant equipment with sensors is viable, a once-performed calibration with high-cost data can lead to comparable results at a reduced sensor complexity during runtime.

II. STATE OF THE ART

The idea of correlating defect formation and emission data in laser welding is an active field of research. Determining which process emissions contribute the highest information gain is of particular interest to reduce setup cost and subsequent computational resources.

Gao *et al.* have shown in Ref. 1 that the recorded visible spectrum of electromagnetic emissions provides a high information landscape for defect detection and process stability estimation on stainless steels. Vakili-Farahani *et al.* performed wavelet analysis[2] on optical recordings spanning a spectral space from 450 to 1700 nm. They also identified changes in the visual spectrum to be a viable indicator for keyhole and plume geometry with the possibility for blowout. In their work with Ti6Al4V alloy, they showed that 250–1700 and 450–850 nm contain mostly similar information, giving insights into keyhole formation and collapse during the welding process. Eriksson *et al.* identified the self-illumination rather than reflective characteristics of the material plume.[3] They postulated its self-emitting radiation to be in the infrared range and highly correlated with the visible spectrum even when the welding process deviates from stable conditions with Pearson correlation coefficients above 0.7.

Given the physical constraints of laser welding being an ongoing process, the focus is on time series-centric evaluation models. Previous attempts by researchers such as Shevchik *et al.*[4] showed that laser back reflection signals and acoustic emission monitoring during laser welding yield valuable process information suitable for defect prediction. For surface-level defects such as blowout, accuracies of up to 95% are reported, while subsurface categories such as pore formation are detected with 73% accuracy. Time-dependent recordings were sliced using a sliding window schema, converted into a spectrogram, fed through a CNN architecture, and supervised training was used on predefined defect classes. Heilmeier and his team[5] used a pool of statistical metrics to evaluate recordings of combined spectral and acoustic emissions from the welding of battery cells. By provoking defect formation and comparing recordings against reference processes, they identified statistical measures for the detection and evaluation of weld seam defects.

III. OBJECTIVES

Current research has demonstrated that numerous defects in laser beam welding processes can be detected through process emissions. However, determining which emissions are relevant requires expert process knowledge.[2,3] This research paper proposes a novel approach to leverage the cost-effectiveness of simple sensors while performing calibration with high-cost *in situ* data. Through a three-stage pipeline, emission data are deduplicated, correlated with defects, and finally ranked according to importance.

IV. MATERIALS AND METHODS

For this study, spectral emissions and high-speed *in situ* imaging during laser beam welding of CuSn6 and Cu-ETP samples between 2 and 3 mm with a variety of infrared (IR) laser sources (1030 and 1070 nm) were used (see Table I). This includes a total of 195 experiments with a combined database row count of 394 057 samples. For this work, the whole set of existing historical experimental data was (re-)evaluated.

A. Synchrotron investigations at DESY

As part of Laser-Meets-Synchrotron,[6] regular research campaigns are scheduled at the Deutsches Elektronen-Synchrotron DESY. High-energy x-ray radiation from the particle accelerator is used to gain insights into metal probes during welding operations. After the x rays pass through the material, they are converted into visible light by a scintillator. This image is recorded by a high-speed camera iSpeed727 on the opposite side. Meanwhile, the multispectral sensor 4D.TWO from 4D Photonics records a broad set of emission data from an axis angle pointing at the welding area. It offers cumulative spectral band counts at 5 kHz from 353 to 1850 nm with 32 individual bands (16 around the VIS and 16 IR) spaced out at regular intervals to focus on different band groups interesting for the welding industry.

As a practical implementation, the material is moved, with a steady advance speed, by a linear axis through the cross-point of a laser and synchrotron beam [see Fig. 1]. The resulting images are flat field corrected and contrast enhanced. Annotating these high-speed images is a labor-intensive process. An internal Mask-RCNN network with a custom backbone was used to generate per image statistics of keyhole geometry and material defects [see Fig. 2]. Recorded data were normalized based on process parameters as much as possible. An assortment of unbiased preselections was

12 February 2025 11:49:34

TABLE I. Technical specifications of utilized lasers.

Parameter	Coherent ARM-FL	IPG YLR-1500	IPG YLR-3000	IPG YLR-4000	IPG 1000-WC	IPG 6000-WC	TRUMPF TD6001	TRUMPF TD8001
Wavelength λ (nm)	1030	1070	1070	1070	1070	1070	1030	1030
Power P (W)	600–800	500–750	360–1000	3000	360–800	3000–5800	350–4000	1250–7700
Operating mode (–)	cw	cw	Cw	cw	cw	Cw	cw	cw

© Author(s) 2024

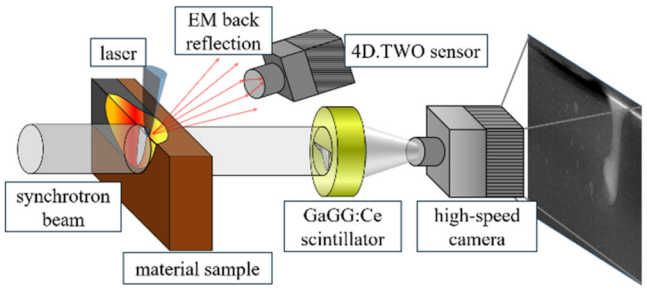

FIG. 1. Abstract data acquisition setup at DESY (Ref. 7).

used to discard experiments from our database. These included but were not limited to the following: experiment length to be sufficiently long, the detected keyhole center to be stable, no trend in laser power being apparent, and the material moving and not static.

IV. RESULTS AND DISCUSSION

A. Defect label construction

A key trait of supervised machine learning is the provision of ground truth data to compute gradient losses from. As the two main categories, a regression or a classification approach can be used. For this work- and domain-specific context, classification was deemed to be the more appropriate path. To facilitate quantitative defect descriptions, geometric per image statistics were combined to devise seven meta-defect columns as multihot vectors.

In laser-based metal welding, keyhole integrity and pore formation are the key factors in obtaining high product yields. While certainly interesting, the exact percentage of pore volume inside the product is not needed to deem a weld successful or a failure. A classification task of monotone increasing porosity brackets of different sizes better reflects the industry needs. After inspection of the available dataset, the first four categories define the porosity of 0%–1%, 1%–5%, 5%–15%, and 15%–100% per 2D area, respectively, where the first bracket is a proxy for near no pores given a margin of detection error. The remaining steps were chosen arbitrarily and can be adjusted at will by abiding to individual quality constraints or ISO norms.

Additionally, three keyhole-related defects were introduced: bulging, piercing, and splashing of the keyhole. Together, they form a multiclass binary-encoded label vector. While the porosity quantiles are exclusionary, the keyhole labels can appear simultaneously.

The porosity of the experiment was calculated by determining the percentage of visible pore area relative to the area of the melt [see Fig. 3]. The underlying concept involves recognizing that while the pore area is absolute, the reference area should not encompass the entire image size but rather only the region where a pore can manifest physically.

This defining rectangle was established with its upper boundary aligned with the surface of the probes, as pores can exclusively manifest within the workpiece. The right side of the rectangle is constrained by the averaged x-coordinates of the keyhole centers, considering the keyhole as the sole source of pores. The lower side is delimited by the upper five percent quantile of keyhole depths for the same rationale. Conversely, the left side is determined by the lower five percent quantile of pore center x-coordinates. This design choice is rooted in the Gaussian x-ray beam distribution, where the brilliance of the beam peaks in the center of the image and gradually diminishes toward the edges, leading to a reduction in the signal-to-noise ratio. During normal operation, detected pores may consequently blend with the background noise before reaching the image border.

As evidenced in prior research,[4] a clear relation exists between the instability of a keyhole and the formation of pores, particularly when the keyhole enters a bulging phase.

During such bulging episodes, the keyhole expands its volume into the horizontal plane of the workpiece, only to subsequently collapse, trapping vapors within and resulting in the creation of new pores [see Fig. 4]. While it is not a certainty that every bulging keyhole will lead to pore formation, such occurrences are decidedly undesirable in standard operating conditions and are classified as defects if the lower portion of the keyhole contributes to 70% or more of the overall volume.

The second keyhole irregularity centers on the rapid increase in keyhole depth, serving as an indicator of inhomogeneous weld

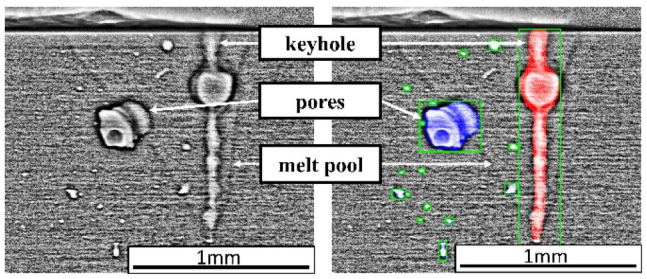

FIG. 2. Example feature extraction from a preprocessed DESY x-ray image (left). Pore detections (blue) and the central capillary (red) were visualized from pixel segmentation data (right).

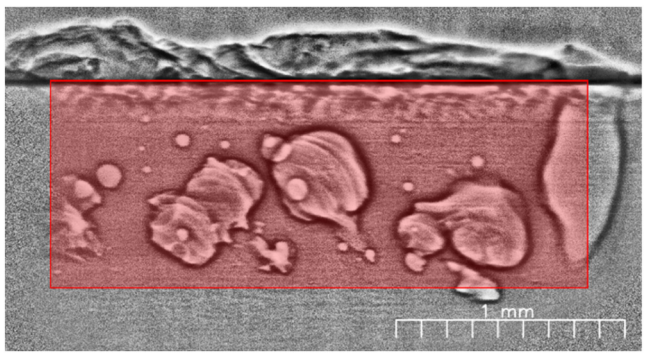

FIG. 3. Visualization of melt area considered for porosity calculations.

12 February 2025 11:49:34

© Author(s) 2024

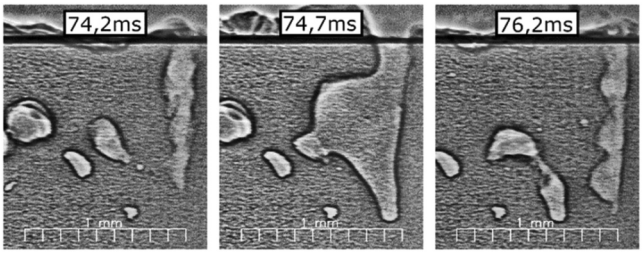

FIG. 4. Bulging weld at 18 kHz. Cropped, flat-field corrected, and contrast enhanced.

depth. During the piercing phase, the depth fluctuates by at least 30% per time step, signaling a significant deviation in the consistency of the weld line as mirrored by the unpredictable movements of the surrounding weld pool [see Fig. 5].

Lastly, a splashing keyhole is characterized by a swift contraction or expansion covering at least 30% of the surface length. This phenomenon serves as an indicator of uneven surface penetration and is often accompanied by a material blowout, as higher layers grapple with the full force of the laser energy, prompting rapid melting and partial vaporization (see Fig. 6).

B. Pipeline stage 1: Dimensionality reduction

Preliminary statistical analysis of the dataset displayed a high Pearson correlation coefficient, which can be indicative of input data being repetitive in their informational content. To yield an optimal usage of training time in terms of performance, pruning or combining correlated features is advised. Two dimensionality reduction techniques were investigated, namely principal component analysis (PCA)[8] and autoencoders (AE).[9]

Plotting every correlation coefficient permutation in order can give an assessment into what reduction can be expected and what method is advisable to use. It is important to note that this measure of correlation can only show linear correlations due to the inherent nature of the used covariance itself. Furthermore, it is limited in scope to the underlying process used, including the laser wavelengths and selected copper metal.

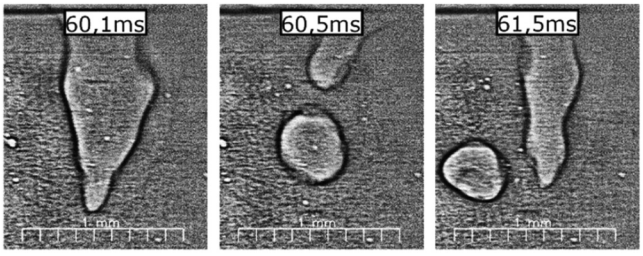

FIG. 5. Piercing weld at 18 kHz. Cropped, flat-field corrected, and contrast enhanced.

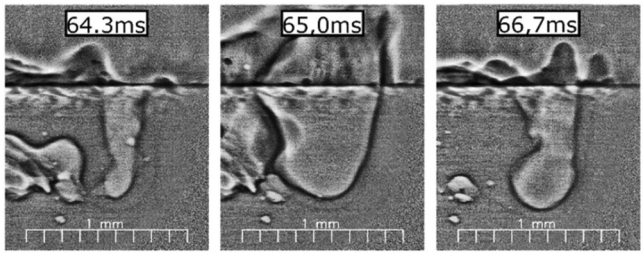

FIG. 6. Splashing weld at 18 kHz. Cropped, flat-field corrected, and contrast enhanced.

Inspecting the pairwise coefficients (see Fig. 7) displays the expected self-correlations of 1 across the diagonal. Additionally, there are two distinct patches in the lower left and upper right. The former part shows a high correlation between the ultraviolet (UV) (below 380 nm), visible (380–750 nm), and near-infrared (NIR) (755–1000 nm) with slight reaches into the short-wave infrared (SWIR) (1000–2500 nm) wavelength range. At the border to the second patch, minor steps following the transition can be observed. In accordance with previous research,[2,3] high linear correlation between the features exists, which can be exploited to perform dimensionality reduction techniques target at linear-dependent data.

Given this preliminary analysis, a PCA performed with a component count of $n = 2$ yields a 95% reproduction of data variance at a mean squared error of $\mathrm{MSE}_{\mathrm{PCA}(n=2)} \approx 2.32 \times 10^{-3}$. Increasing the component count to n = 4 raises the captured data variance to 99.42% and $\mathrm{MSE}_{\mathrm{PCA(n\,=\,4)}} \approx 3.62 \times 10^{-4}$, as more intricate spectral band correlations (as seen in Fig. 8) get

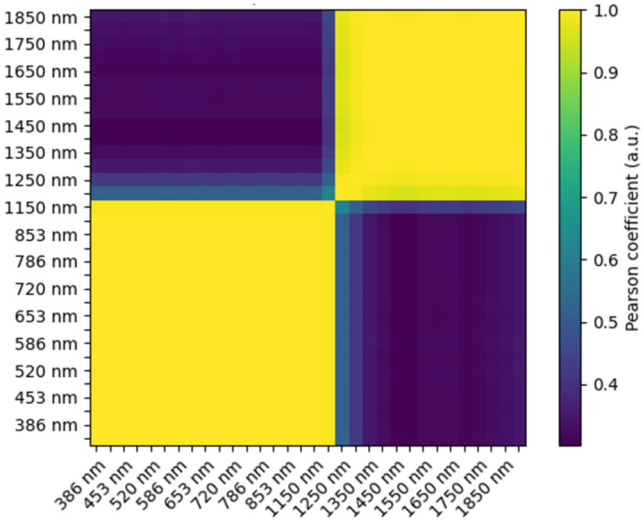

FIG. 7. Pearson correlation coefficient pairwise across all wavelength samples available in the dataset.

12 February 2025 11:49:34

© Author(s) 2024

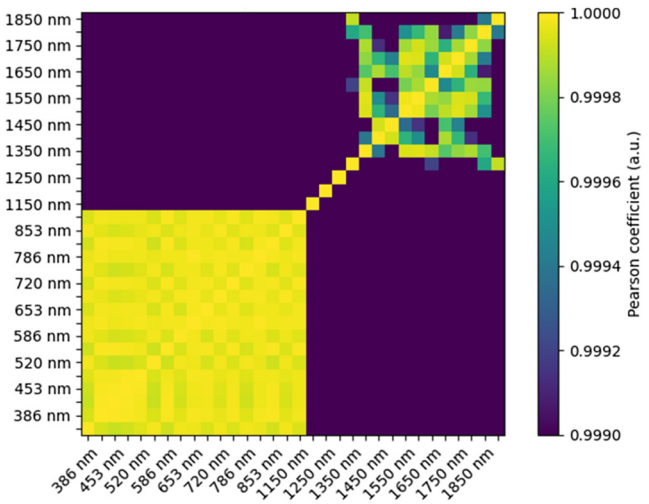

FIG. 8. Pearson correlation coefficient pairwise across all wavelength samples available in the dataset clamped between 0.999 and 1.

approximated. Here, special caution must be emphasized, as Fig. 7 can give the impression of only needing two components.

Comparably, an autoencoder outperforms the PCA at a latent space size of $n \geq 2$ with an $MSE_{AE(n=2)} \approx 1.86 \times 10^{-3}$ and $MSE_{AE(n=4)} \approx 1.34 \times 10^{-4}$. As an autoencoder is capable of learning nonlinear functional mappings with just one hidden layer,[10] this performance increase can be attributed to smaller nonlinear correlations in between spectral bands.

C. Pipeline stage 2: Defect prediction

To perform fault detection, state-of-the-art deep learning models, especially LSTM networks, and Transformers are implemented. Both are well-suited for capturing temporal and long-range dependencies in sequential data, making them applicable for analyzing fault patterns.

Annotated data are batched inside the dataset classes after sorting by experiment time to produce time series training data on the fly. Following inheritance-based design, trained PCA- and AE-based dimensionality reductions are integrated on top as separate datasets. For training on temporal data, an amount of lag is defined to batch samples together, forming a time series sample. While some deep learning architectures can operate on varying input length, such as the Transformer, a fixed time step was chosen for comparability between all presented models and seen as an inherent result from usually fixed sample rates of sensors. It is important to prevent cross-experiment batching to allow for temporal coherence, and this will render all trailing samples of any experiment not dividable by the selected time step unusable.

Through the previous synchronization step, the data were resampled to around 5000 Hz or 0.2 ms. As mentioned by Shevchik,[4] relevant events regarding laser-based metal welding occur in the perceptible time frame between 0.2 and 10 ms. This limits the interesting time lag, regarding defect formation, from 1 up to 50 samples. Due to the collection of different advance speeds of work materials during recording, as well as different gain settings selected for the spectral sensor, no further normalization or regularization was performed to mitigate absolute value drift. Instead, a grid search for all lag values through the range of interest was performed, and the highest accuracy was observed with a lag of 33 samples or 6.6 ms. The expressive power, therefore, solely relies on the relative signal changes rather than absolute indications.

To establish a baseline for defect label predictability and demonstrate the effectiveness of simpler architectures, an LSTM will act as the starting point for the investigation into different deep learning models at this pipeline stage. Due to its age of inception in 1997,[11] they are a staple in sequence learning methods and included in virtually all modern machine learning libraries as function blocks. They should always be considered when evaluating deep learning applicability, and speed of development is paramount. In this work, the default pytorch LSTM implementation was used, followed by a linear layer to condense the output down to seven labels. Through hyperparameter grid search, a stacked configuration of 3 LSTMs with an inner layer dropout of 0.5 was determined, where each block contains a hidden layer size of 128 neurons. In addition, BiLSTM[12] comprising two perpendicular-oriented LSTM networks were tested to compare against and evaluate the influence of vanishing gradients on feature attribution analysis.

To assess the viability of variable input length models, the Transformer[13] was implemented as well. A typical Transformer acts as a sequence-to-sequence model, transforming one sequence of features into another, by encoding an embedded representation, applying multiple attention steps, and decoding the results again. For a classification task, the last stage is not needed but rather the attention-based outputs are used as inputs for a classifier stage directly. Therefore, the Transformer in this work consists of three distinct stages. A preprocessing, an attention, and a classification stage.

D. Reweighting training data

Defect introduction in machining is usually undesirable, actively avoided, and meticulously looked out for, which challenges the usage of normal operation data as training data. It can lead to drastic imbalances between positive (1) and negative (0) classifications. In Fig. 9, such an imbalance in the utilized dataset class distribution can be observed. Training will become ineffective as models can simply drop their weights to zero and collapse to a constant function while still retaining high accuracy.

Ideally, additional experiments would be performed to intentionally create samples for the missing class labels. If this option is unavailable, as in this case, due to plant unavailability, cost, or other reasoning, alternative solutions exist to balance the dataset. Data augmentation aims to artificially inflate the dataset prior to training based on injecting noise, manipulating statistical properties, or utilizing generative models to generate new and plausible time series data. Wen *et al.*[14] provide a comprehensive overview of different time series augmentation methods to explore.

Due to real data availability in future DESY campaigns, an alternative route of in-training adjustment of class label weights was implemented. Pytorch offers the ability to introduce positive weights during loss computations with *torch.nn.BCEWithLogitsLoss*.

12 February 2025 11:49:34

© Author(s) 2024

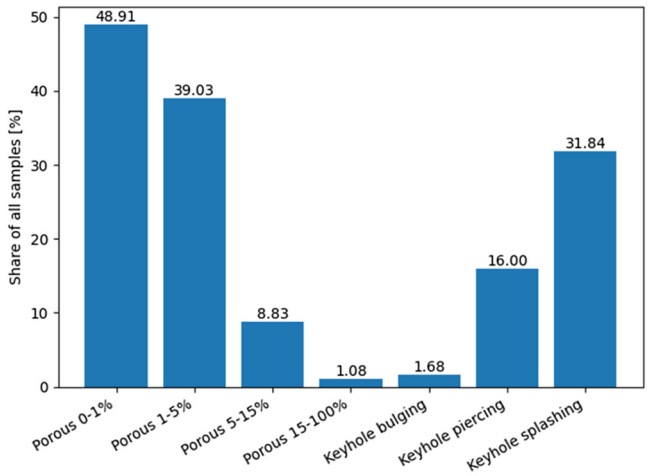

FIG. 9. Distribution of labels across the dataset.

Conceptually, sparse label classes contribute to higher losses than abundant ones. Weights for the positive class probabilities are provided as the ratio of negative to positive class imbalance.

While this reweighting will solve the dataset skew regarding the per label imbalance, a multiclass multilabel dataset introduces an interlabel imbalance where labels correlate together, which can be undesirable. The most reliable solution would be to split the task into multiple classifiers to work on each label in a collective and rejoin their results later. Alternatively, if one model is desired, the loss at every training step can be weighted by the label class contributions to reduce the overall imbalance. Weights are adjusted by the label classes' contribution to the overall positive classifications and renormalized to a weight tensor of one. By disabling the default reduction technique at loss computation, multiplying the weight tensor with the loss matrix, and reducing the loss, the per class weighting is achieved. Both reweighting techniques presented were implemented into this work's training code.

The collected performance measures (see Tables II and III) follow the assumption that more sophisticated architectures provide better accuracy at the cost of training time. A basic LSTM utilizes the lowest parameter count at 348k and subsequently the shortest training time of 245 s. While providing a modest baseline performance with a total accuracy of 69.78%, it could surpass other models (in the same features category) in bulging keyhole detection. This anomaly could indicate further problems regarding the sparse definition of positive samples in the label class. At a significant increase in model parameters of nearly 200%, the BiLSTM could improve predictive accuracy only marginally. By trading a third of the total parameter count of the BiLSTM, the transformer model could still acquire the overall performance crown at a total accuracy of 76.53%. As the most complex architecture, it requires additional training time. This ranking of model performance gets blurred when introducing labels from previous time steps as features. On average, all models gained a significant boost in performance of 15%, whereas the Transformer is still marginally ahead at 90.80% (see Table IV). Surprisingly, both LSTM closed the lead with a total accuracy of 90.26% and 90.49%, respectively. The introduction of the previous prediction could introduce an additional dimension of time progression to aid the LSTMs, reducing the advantage Transformers get from their self-attention mechanism.

All trained time series models can be used to perform live or quasilive annotation of spectral data obtained during welding operations. In Fig. 10, a weld with the Trumpf TruDisk 6001 at P = 2000 (W), V = 100 $\left(\frac{\text{mm}}{\text{s}}\right)$, and λ = 1030 (nm) on Cu-ETP was performed. By overlaying the predicted defect label annotations onto a surface image, a clear coincidence of real-world defect formation with class predictions can be observed. Nonstable welding processes create irregular weld seams as shown by Kang *et al.*[15]

For porosity, the four classes are combined into a color code of green, yellow, orange, and red depending on the severity of the porosity. For the keyhole classes, only green (negative) and red

TABLE II. Explored model architecture statistics with 32 features as inputs and 33 time steps.

Metric	LSTM	BiLSTM	Transformer
Parameter (#)	348k	957k	599k
Epochs (#)	55	59	61
Training (s)	245.54	274.03	312.99
Inference (ns)	2.72	4.80	11.92

TABLE III. Model accuracies in percent with 32 features as inputs and 33 time steps (best results bold).

Defect class	LSTM	BiLSTM	Transformer
0%–1%	65.08	65.63	**75.03**
1%–5%	62.19	62.08	**70.92**
5%–15%	75.51	72.77	**77.74**
15%–100%	68.89	73.45	**90.10**
Bulging	78.63	75.13	**69.57**
Piercing	70.05	72.84	**80.17**
Splashing	68.15	67.35	**72.17**
Total	69.78	69.89	**76.53**

TABLE IV. Model accuracies in percent with 39 features (32 spectral, 7 previous predictions) as inputs and 33 time steps (best results bold).

Defect class	LSTM	BiLSTM	Transformer
0%–1%	93.82	94.18	**94.34**
1%–5%	91.91	92.27	**92.66**
5%–15%	95.06	**95.48**	94.93
15%–100%	97.27	**97.41**	97.36
Bulging	84.77	**86.65**	86.30
Piercing	88.20	88.55	**89.42**
Splashing	**80.79**	78.89	80.62
Total	90.26	90.49	**90.80**

12 February 2025 11:49:34

© Author(s) 2024

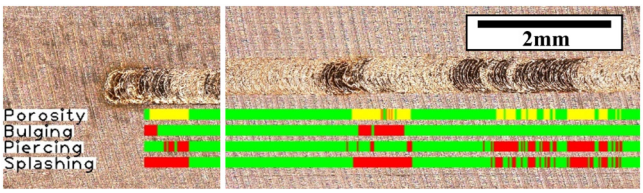

FIG. 10. Annotation of weld experiment Cu-ETP-ID3 superimposed on the weld-surface image (transformer model, wavelength data only, and truncated).

(positive) are used. The beginning of the weld cannot be annotated due to the imposed fixed input size constraint. In the middle, a combination of bulging and splashing with indications of medium levels of porosity are indicated following a blowout of material characterized by the irregular weld surface with some material missing. Inspecting the annotations on the right, a problem with stability becomes apparent. Piercing and bulging keyhole switch rapidly between positive and negative classifications, resulting in a barcode pattern. Improved model design or postprocessing in the form of a rolling average could be investigated further.

E. Pipeline stage 3: Attribution analysis

The final stage in the proposed pipeline utilizes feature attribution (also called salience) methods to infer relative importance scores for input features. By focusing on integrated gradients (IG),[16] a wide range of machine learning models can be used besides the one presented.

IG aims to provide a nuanced understanding of feature importance by considering the entire path of input changes from a baseline to the actual input. By performing a feature attribution on a per label basis and averaging the results over the whole dataset, emerging patterns specific to the labels' formation can be extracted. Attribution scores, especially normalized, are relative measures and should not be used to interpret features directly. Selecting an appropriate baseline vector to perform the path integration is paramount in deriving meaningful interpretations. Sturmfels *et al.* have shown that the selection of baseline vectors is not trivial, and no superior sampling method exists.[17] For this work, a Monte-Carlo-based resampling of the feature space combined with averaging the resulting vectors was used. The process is stopped when either the baseline vector does not change anymore up to a certain precision or the feature space is exhausted.

While our brute-force approach by extensive sampling of the feature space shows promising results, Hase *et al.* have shown that explainability functions such as IG suffer an error called out-of-distribution problem.[18] By removing input features to gauge their relative relevance to the output, model inputs often drop out of the distribution they were trained on, yielding subpar results and skewing attribution metrics in the process (called "socially misaligned" in the paper). Training models with attributions in mind can mitigate that problem by injecting those "counterfactual" inputs during training as part of the dataset. It is unclear whether this work is significantly impacted by this problem, which is still an active field of research. Considering that our results are numerically stable and physically sound, it is expected that the magnitude of this error is negligible. Introducing a vast number of features and not performing dimensionality reduction can also lead to attributions approaching or even surpassing random features selections, as shown by Hooker *et al.* from the Google Brain project in Ref. 19. In their performance test, they discovered that for the Resnet-50 architecture, 90% of input features had a near-zero impact on the overall decision making. This leads IG to perform poorly unless additional retraining for distribution alignment was performed in a technique they called ROAR. Like the results of Hase *et al.*, the introduction of "counterfactual" inputs during training is needed to align the models' input features distribution with the expected attribution distribution.

In Fig. 11, the results of an IG attribution analysis for the previously trained 32 feature LSTM on the bulging defect class are shown. For the baseline vector, the average of all negative samples in the dataset regarding the class is picked. The resulting plot depicts the integrated path from a classification as nonbulging to a bulging keyhole. Positive values in attribution score indicate that a rise in a certain wavelength at a specific time step would increase the likelihood of switching from a negative (nonbulging) to a positive (bulging) class. Similarly, a negative attribution score indicates a decrease in certain spectral emissions at specific timeframes to be beneficial for a positive classification. For the presented example, an overall increase in spectral emissions over the whole spectrum is calculated with a peak at time step of −20 (−4 ms). This could indicate improper absorption through destabilizing of the keyhole area. It is followed by a decline in signal up until the time step of −10 (−2 ms). Here, the visible spectrum continues to decline while the UV stabilizes, and the SWIR is rising sharply. At the time step of 0 (0 ms), the bulging keyhole is fully developed and absorption is significantly enhanced via multiple reflections of the beam inside the keyhole and increased vertical expansion.

As visual inspection of a static 3D landscape can be difficult, projecting the results along either the time dimension (see Fig. 12)

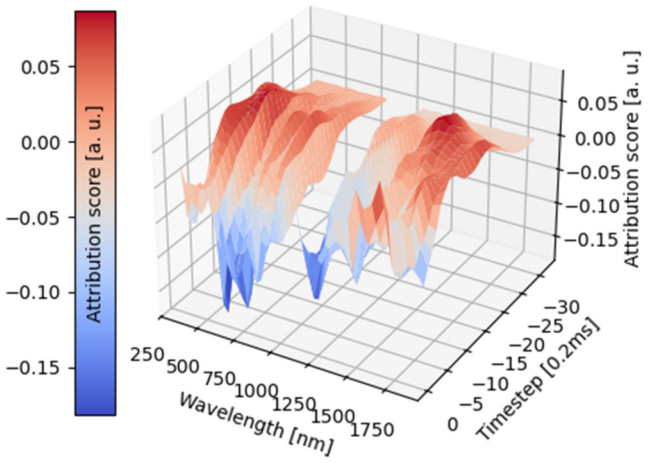

FIG. 11. Normalized LSTM IG attribution analysis for bulging keyhole defect over the whole dataset (negative attributions in blue, positive attributions in red).

12 February 2025 11:49:34

© Author(s) 2024

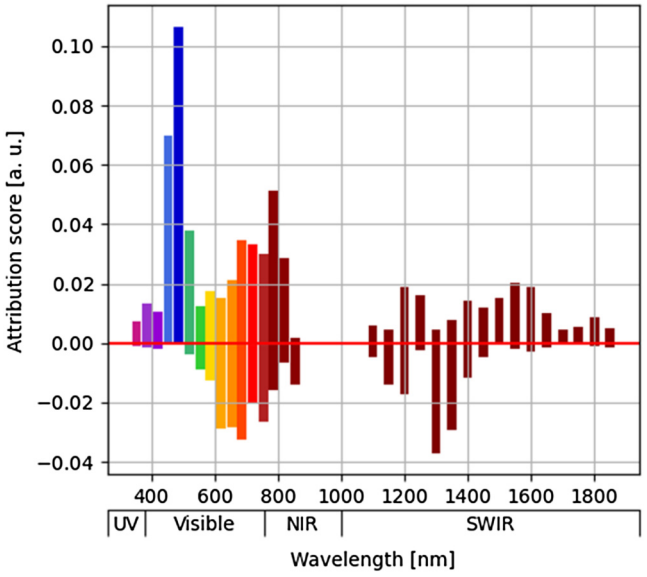

FIG. 12. Normalized LSTM IG attribution analysis for the bulging keyhole defect over the whole dataset, projected along the time axis and split into positive (above red 0 line) and negative (below red 0 line) contributions.

or along the feature dimension [see Fig. 13] can aid in further interpretation, which, in turn, also provides quantitative descriptions of which spectral range is most relevant for a class of defects.

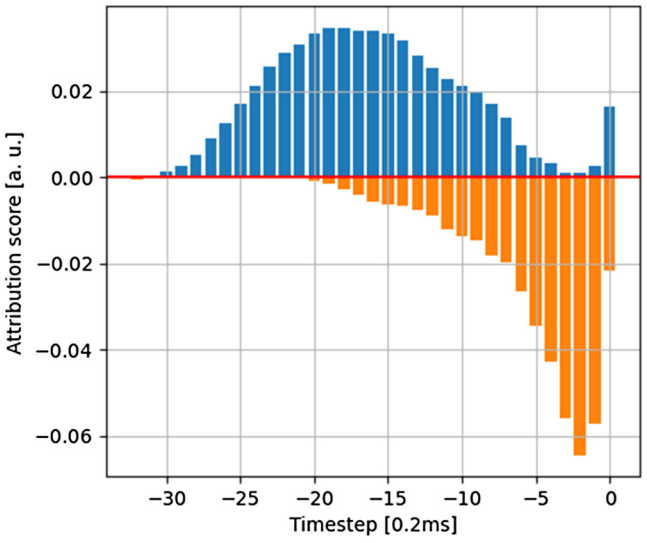

FIG. 13. Normalized LSTM IG attribution analysis for the bulging keyhole defect over the whole dataset, projected along the feature axis and split into positive (above red 0 line) and negative (below red 0 line) contributions.

V. CONCLUSION

In this paper, a general framework for multispectral sensor calibration using sparsely available high-cost synchrotron data was presented. A three-step pipeline was presented consisting of data deduplication, defect prediction, and attribution analysis. Data deduplication was achieved through principal component analysis and autoencoders. Each provides benefits depending on the functional correlation type but yields compression ratios of more than 90% without a perceivable loss in defect predictability. By comparing LSTM, BiLSTM, and Transformer deep learning models, different architectures were explored including a way to mitigate severe dataset imbalance through the means of loss reweighting. Deploying a trained model in a real-world environment has shown promising results as automatically generated defect annotations coincide with optically observed surface deformations. Using integrated gradients spectral patterns indicates that defect formations were extracted from trained deep learning models. This provides a more robust understanding of feature importance and a semiautomated pathway for the selection of useful wavelength channels.

The multispectral sensor system 4D.TWO offers a huge variety of wavelength channels and is, therefore, a beneficial system for machine learning analysis. Depending on the material and process setup, the used channels can be individually chosen due to their contribution to the algorithm's accuracy. Further investigations can explore attribution scores more in-depth and create qualitative comparisons to other metals or for other laser wavelengths following this pipeline.

ACKNOWLEDGMENTS

The presented investigations were carried out within the cooperation "Laser Meets Synchrotron" (www.laser-meets-synchrotron.de). The experimental setup and its operation were funded by the Deutsche Forschungsgemeinschaft e.V. (DFG, German Research Foundation) within the framework of the Collaborative Research Centre SFB1120-236616214 "Bauteilpräzision durch Beherrschung von Schmelze und Erstarrung in Produktionsprozessen." The experiments were carried out in cooperation with Helmholtz-Zentrum Hereon in Hamburg at Beamline P07 of DESY PETRA III as part of proposal BAG-20211050, and we would like to thank F. Beckmann, J. Moosmann, and all people involved for their support. Thanks to 4D Photonics GmbH for providing the 4D.TWO sensor.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jan Brüggenjürgen: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Software (lead); Validation (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Christoph Spurk:** Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal). **Marc Hummel:** Conceptualization (equal); Funding acquisition (equal);

12 February 2025 11:49:34

© Author(s) 2024

Investigation (equal); Project administration (supporting); Supervision (equal). **Christoph Franz:** Conceptualization (supporting); Funding acquisition (equal); Resources (supporting). **Andrè Häusler:** Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal). **Alexander Olowinsky:** Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal). **Felix Beckmann:** Resources (equal); Supervision (supporting). **Julian Moosmann:** Resources (equal); Supervision (supporting).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Selected data are available after authorization in Coscine with the persistent identifier (PID), http://hdl.handle.net/21.11102/b37ef8ee-135f-4c07-a25b-b2931524c65a (accessed on July 27, 2024).

REFERENCES

[1]G. Xiangdong, Y. Deyong, and K. Seiji, "The high frequency characteristics of laser reflection and visible light during solid state disk laser welding," Laser Phys. Lett. **12**, 076003 (2015).
[2]F. Vakili-Farahani, J. Lungershausen, and K. Wasmer, "Wavelet analysis of light emission signals in laser beam welding," J. Laser Appl. **29**, 022424 (2017).
[3]I. Eriksson, J. Powell, and A. F. H. Kaplan, "Signal overlap in the monitoring of laser welding," Meas. Sci. Technol. **21**, 105705 (2010).
[4]S. Shevchik, "Supervised deep learning for real-time quality monitoring of laser welding with X-ray radiographic guidance," Sci. Rep. **10**, 3389 (2020).
[5]J. Heilmeier, M. K. Kick, S. Grabmann, T. Muschol, F. Schlicht, F. von Hundelshausen, H.-G. von Ribbeck, T. Weiss, and M. F. Zaeh, "Inline failure detection in laser beam welding of battery cells: Acoustic and spectral emission analysis for quality monitoring," J. Laser Appl. **36**, 022007 (2024).
[6]A. Olowinsky, A. Gillner, and C. Spurk, *Laser Meets Synchrotron* (Fraunhofer ILT, 2024), see https://laser-meets-synchrotron.de
[7]H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psychol. **24**, 417 (1933).
[8]G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science **313**, 504–507 (2006).
[9]K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Netw. **2**, 359–366 (1989).
[10]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**, 1735–1780 (1997).
[11]M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process. **45**, 2673–2681 (1997).
[12]A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv:1706.03762 [cs.CL] (2017).
[13]Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," arXiv:2002.12478 (2020).
[14]Y. Kang, Y. Zhao, Y. Li, J. Wang, and X. Zhan, "Simulation of the effect of keyhole instability on porosity during the deep penetration laser welding process," Metals **12**, 1200 (2022).
[15]M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," arXiv:1703.01365 [cs.LG] (2017).
[16]P. Sturmfels, S. Lundberg, and S.-I. Lee, "Visualizing the impact of feature attribution baselines," Distill **5**, e22 (2020).
[17]P. Hase, H. Xie, and M. Bansal, "The out-of-distribution problem in explainability and search methods for feature importance explanations," Adv. Neural Inf. Process. Syst. **34**, 3650–3666 (2021).
[18]S. Hooker, D. Erhan, P.-J. Kindermans, and B. Kim, "A benchmark for interpretability methods in deep neural networks," arXiv:1806.10758 [cs.LG] (2019).
[19]C. Meier, *Synchrotron-Strahlung zur Visualisierung und Analyse Wellenlängenabhängiger Prozesseigenschaften Beim Laserstrahlschweißen von Kupfer* (RWTH Aachen, Aachen, 2022).

12 February 2025 11:49:34

© Author(s) 2024