

# Deep Learning Algorithms for Process Monitoring in Plastics Absorber-Free Laser Transmission Welding

Nam-Phong Nguyen<sup>\*1</sup>, Chaitra Nayak<sup>1</sup>, Maximiliani Brosda<sup>1</sup>, Alexanderi Olowinsky<sup>1</sup>

One benefit of laser transmission welding is the non-contact, very accurate energy deposition. This makes it possible to fabricate seams with intricate and restricted geometries. Joints must be reproducibly made in the lowest dimensions while retaining good quality, especially in the case of high-tech devices. Here, pyrometric sensors are used to implement a process control. However, the signal is only indirectly detected, and it relies on a number of factors like the material's qualities, the area of the heat-affected zone, and its thermal characteristics. Furthermore, since the pyrometer only outputs a spatially integrated signal, neither the contour of the welding seam nor the shape of the melt pool can be determined. This study uses Deep Learning techniques to analyze the laser transmission welding process. Image frames that depict the region where the laser beam interacts with the substance are recorded. Semantic segmentation will then be used to automatically process the picture. This makes it possible to estimate common locations like seam geometries or weld pools. With intersection over union values  $> 0.92$ , the findings demonstrate a good agreement between the prediction and the actual data. The laser power is then predicted using the geometric data that was retrieved. For laser powers 15 W, excellent prediction outcomes are obtained in this case.

**Keywords:** laser transmission welding, process monitoring, deep learning, semantic segmentation, plastics

## 1. Introduction

Plastics cannot only be manufactured easily but can also possess competitive properties compared to glass and silicon [1]. The production of high quality plastic components can be expensive and they often consist of multiple parts which need to be joined together. Therefore, there is a need for a joining process where the plastic components can be welded in a reliable way. One such promising technology is laser transmission welding (LTW) [2]. The temporal, spatial and contactless energy deposition leads to sealing with low thermal load on the neighbouring regions [1]. Additionally, laser technology enables a highly automatic operation and online process monitoring techniques can be easily integrated. Recent developments enable a LTW process of transparent plastics where no absorber is needed. Here, beam sources are used which emit radiation in the natural absorption range of polymers ( $\lambda = 1.6 - 2.1 \mu\text{m}$ ). This

method is called absorber-free LTW and it is used for medical or biotechnology applications [3]. The motivation for process monitoring is to ensure the consistent quality of the produced goods by identifying or avoiding defects on the welding seam. Another goal is to reduce the manual control by an operator which also ensures reduced costs. At present, pyrometric sensors are used to measure the material temperature and to detect the heat radiation emitted during the welding process [3,4]. The signals generated from the process depend on factors such as the material properties, heat affected zone, beam parameters etc. time-resolved observation of the welding process [5,6]. AI is a branch of computer science that is progressing rapidly with an intent to make a machine think in an intelligent way similar to the thinking process of human beings in order to

<sup>1</sup>Fraunhofer Institute for Laser Technology ILT, Steinbachstraße 15, 52074 Aachen, Germany

<sup>2</sup>Visual Information Analysis Group, TU Kaiserslautern, 67653 Kaiserslautern, Germany

\*Corresponding author's e-mail: phong.nguyen@ilt.fraunhofer.de

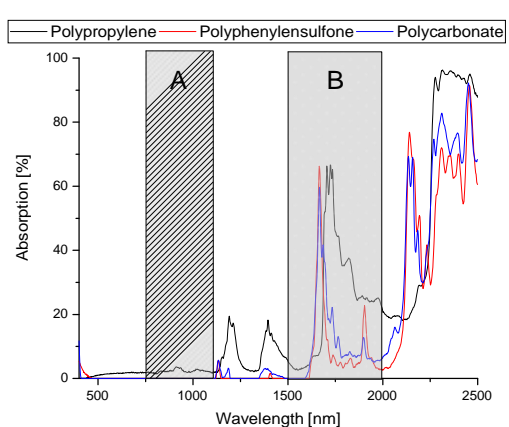
perform different tasks depending on the actual data. The demand for AI in applications such as natural language processing, vision systems, expert systems, speech recognition, robotics, sensor-based applications, handwriting recognition etc. has been dominant [7]. Similarly, the research with regard to AI in laser technology is promising which aims at improving the process productivity. The presented work is partly based on previous works of Knaak et al. for laser welding of metals [5]. The recorded image frames are automatically analysed using semantic segmentation which is a pixel-wise classification method. Each pixel is analysed and assigned to a pre-defined class such as weld seam, bubbles or burnings. This is carried out by training deep neural networks (NN). With this method, it is possible to visualise the geometrical features.

In the next step, the geometrical features which are generated from semantic segmentation are used to predict the laser power. Here, the number of pixels of each geometrical feature is passed into a neural network regressor which is a difference compared to semantic segmentation where the image data is processed.

## 2. Fundamentals

### 2.1 Absorber-free laser transmission welding

In absorber-free laser transmission welding (LTW), beam sources are used which emit radiation in the intrinsic absorption bands of polymers ( $\lambda = 1.6 - 2.1 \mu\text{m}$ ). Because of that, plastics can be welded without using any absorbing additives (see Fig. 1). A typical beam source is a thulium fibre laser with an emission wavelength of  $\lambda = 1.94 \mu\text{m}$ . Since transparent plastics can be welded with this method, absorber free LTW is especially suited for the encapsulation of multilayer polymeric films for sensitive electronic devices such as organic photovoltaics or organic light emitting diodes. Other promising application areas are life science applications such as sealing of microfluidic devices [2,4,8].

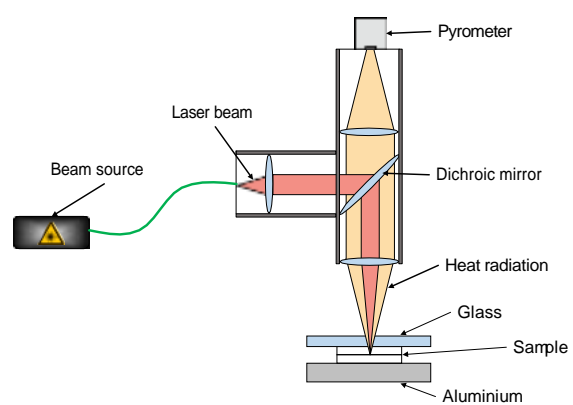


**Fig. 1** Absorption curves in the wavelength area of classic (A) and absorber-free (B) LTW. In (B), no absorbing additives are needed since most polymers show a natural absorption behaviour in this range.

Damages inside the produced goods have to be identified during the production process. In LTW, typical defects are thermal damages which are caused due to air gaps or contaminations on the sample. Furthermore, local heat accumulation which appears during the acceleration and de-

celeration of the laser beam movement can lead to thermal damages. Seam damages can lead to an early failure of the component. Especially in areas such as medical applications, a premature failure of the component can have dramatic consequences for the patients. Hence, there is a high need for the development of a process control in absorber-free LTW [3].

Typically, pyrometric sensors are integrated coaxially inside the beam path to monitor the welding process. The emitted heat during the welding process is then measured by the sensor and converted into an electric signal [3,9]. A typical setup is shown in the following figure.



**Fig. 2** The coaxial integration of a pyrometric sensor in LTW allows the observation of the welding process based on the emitted heat radiation.

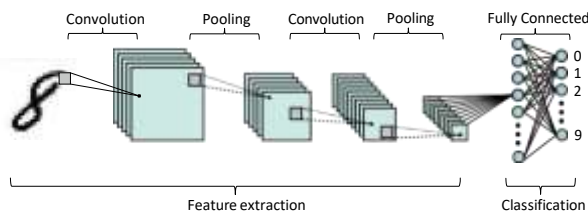
Thermal damages inside the plastic can be detected via a signal peak. Furthermore, a closed-loop control of the laser power can be easily established with the help of the pyrometric signal. However, due to the overlap between the emission wavelength of the beam sources ( $\lambda = 1.6 - 2.1 \mu\text{m}$ ) and the measurement range of the pyrometer ( $\lambda = 1.5 - 2.1 \mu\text{m}$ ), both systems cannot be operated simultaneously as the pyrometer would only measure the laser emission. Sensors with a measuring range at higher wavelengths can only measure the temperature on the surface and not in the interface as polymers have low transmission properties in this range. Mamuschkin et al. showed that it is necessary to operate the beam source in a pulsed mode whilst the thermal radiation is measured between the laser pulses. The disadvantage is that an ideal pulse parameter has to be identified. The longer the pulse duration, the shorter the measurement duration becomes. If the pulse duration is too short, a disruption of the welding seam can occur [3].

Another aspect is the spectrally and spatially integrated measurement method of the pyrometer. Thermal damages can be detected as a signal peak but it is not clear what kind of thermal damage (e.g. bubble formation, burnings etc.) is existing as well as the reason for its occurrence (e.g. contamination, air gaps etc.). Additionally, the detection of the radiation becomes more difficult with smaller laser spot sizes. Especially for the encapsulation of microfluidic devices, the spot diameter can reach values between 20 – 50  $\mu\text{m}$ . Furthermore, the emitted heat radiation is being attenuated through optical components such as lenses or mirrors.

## 2.2 Convolutional neural networks and semantic segmentation

Deep Learning is a sub-field of machine learning containing algorithms based on layers to create an artificial network which can learn on its own and predict the outcome. Deep complex neural networks can be used to create a human-like AI as the networks help in analysing large datasets. In this work convolutional neural networks (CNN) are used which will be presented in the following [7].

The purpose of CNNs is to identify the patterns in a given dataset of images by extracting the features for making predictions [10]. The layers in the network capture the spatial information by learning through kernels or filters. The filters have a small receptive field and are convolved across the image. Apart from these convolutional layers, the input is further processed via dropout and pooling layers as well as different activation functions. As a result, CNNs have the capability of identifying spatial patterns such as edges, shapes, curves, textures, objects, shading changes etc. The low-level features (simple) are detected at an initial stage followed by medium-level features and finally the high-level features (complex) are detected [7]. The following figure shows a simple CNN which is used for detecting handwritten digits.



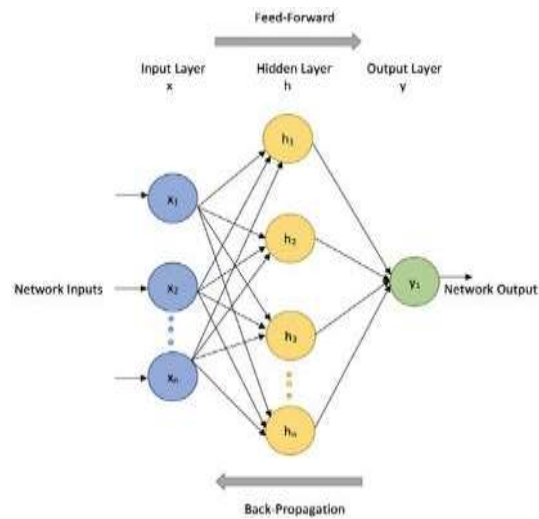
**Fig. 3** The architecture of a simple CNN for the classification of handwritten digits. The input picture is processed by different filters and functions. Modified according to [10].

The fully connected layers are responsible for the classification whilst the rest of the layers are responsible for the feature extraction. Each layer in a CNN is composed of weights and biases that sum up to the total number of parameters in the entire network [7]. In this work, the fully connected layers are replaced by convolutional layers which makes the network a fully convolutional network. Further information about this concept can be found in [11].

Applications of CNNs include image classification, object detection, object retrieval, segmentation, instance segmentation etc. In this work, semantic segmentation is used which is a pixel-wise classification. Here, certain areas in an image are segmented into different groups. Each group is marked with a particular colour so that the size and the location of each group can be estimated [7, 11].

## 2.3 Prediction of laser power using regression analysis

To predict the laser power, a regression analysis is used. Regression analysis is a statistical technique used to predict numerical values. They contain methods which help analyse the relationship between two or more continuous variables [12]. In this work, a NN is used to perform regression analysis. An example of such a NN is presented in the following figure. The model takes a vector as an input and outputs a scalar.



**Fig. 4** A neural network which can be used to perform regression analysis. The model takes a vector as an input and outputs a scalar.

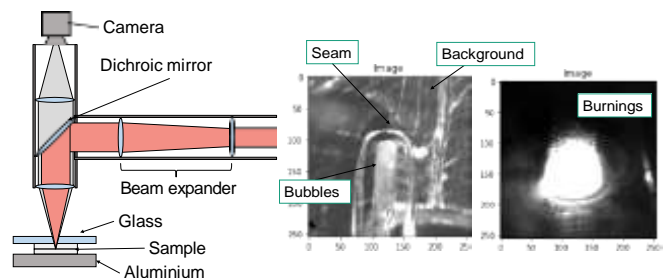
Backpropagation is used during the training process to adjust the weights with the aim to minimise the prediction error [7].

The output variable is the laser power whilst the input parameters are the number of pixels of each class which can be extracted based on the output of the semantic segmentation task. For instance, an increase of the laser power leads to a higher seam width. This on the other hand increases the number of pixel of the seam class which can be detected and analysed by the NN regressor. The prediction can be plotted against the real values to assess the performance of the network.

## 3. Implementation

### 3.1 Generation and preparation of the data set

Contour welding tests with a constant feed rate of  $v = 3.33 \text{ mm/s}$  are carried out to generate the data set. The laser power is varied between  $P = 6.6 - 47 \text{ W}$ . Apart from intact welding seams, thermal damages in the form of bubble formation or burnings are created in the polymeric sample to cover a wide range of welding scenarios. As testing material, polystyrene is used. In order to record the welding process, a camera (DMK 33UX287, The Imaging Source Europe GmbH, Bremen, Germany) is coaxially integrated inside the beam path. The following figure shows the experimental setup.

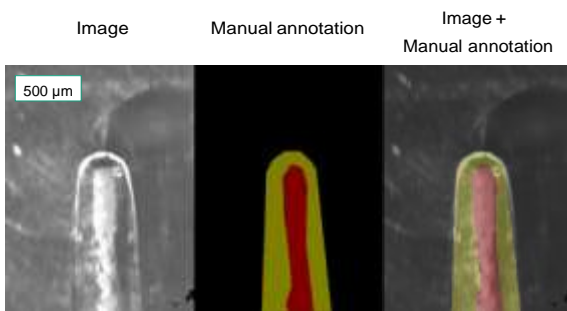


**Fig. 5** Schematic experimental setup for recording the welding process (left) and the defined classes (right) which will be used for the semantic segmentation process.

In total, 97 process videos are recorded and 3224 image frames are extracted. The following classes were defined:

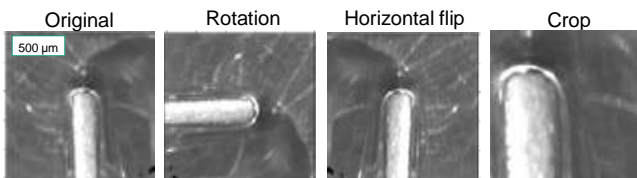
- background
- seam
- bubbles
- burnings

The clamping device as well as the non-molten material belong to the background. Bubbles are generated if the laser power is increased and exceeds a certain threshold. If the decomposition temperature inside the sample is exceeded, burnings occur. This leads to an overexposure of the camera sensor. The dataset is divided into a train, validation and test set with a ratio of 80%, 10% and 10%. For the training process, the whole data set has to be manually annotated. This is carried out using the open source program LabelMe. The following figure shows the annotation of an exemplary image frame.



**Fig. 6** Annotation of the data set. The annotation is then used for the training of the networks.

Since the data set size is quite low, there is a risk of an overfitting of the model on the data set. Therefore, the size is artificially increased using data augmentation. Here, the image frames are additionally cropped, rotated, or mirrored (see Fig. 7).



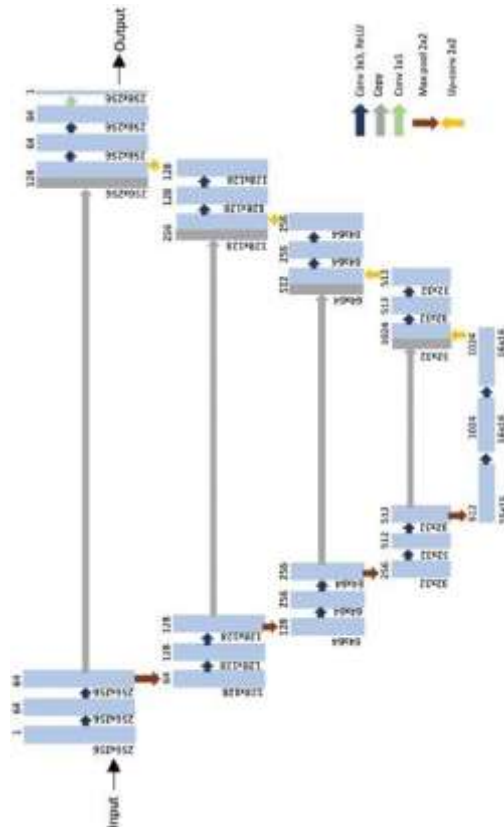
**Fig. 7** Increasing the data set using data augmentation in order to prevent overfitting.

### 3.2 Model architecture for semantic segmentation

In this work, a UNet is used which was developed by Ronneberger et al. for the detection of cells on microscopic images [13].

First, the input is processed in four convolutional blocks. Each block consists of two convolutional layers with a 3x3 filter, a ReLU activation function and a max-pooling layer. These elements are forming the downsampling path since the input is compressed. Furthermore, the number of filter channels is doubled with each downsampling step. The downsampling path is responsible for the classification of the input. In the upsampling path, the input is upscaled so that the output eventually has the same dimensions as the input. Whilst the downsampling is responsible for the classification, the upsampling ensures the localisation of the classes on the image frame. The upsam-

pling process is carried out with four upsampling blocks. Each block consists of a transposed convolutional layer with a 2x2 filter, two convolutional layers with a 3x3 filter and a ReLU activation function. Each downsampling process has a corresponding upsampling process which leads to a symmetric U-shape of the network. Additionally, the output of a downsampling process is directly connected to the corresponding upsampling process via a skip connection. All in all, the network consists of 23 convolutional layers. The architecture is presented in Fig. 8



**Fig. 8** For the semantic segmentation, the UNet architecture is used which was developed by Ronneberger et al. [13].

Apart from the classic U-Net, further models were U-Net used which were modified with a ResNet architecture.

- UNet with ResNet-18 architecture
- UNet with ResNet-34 architecture
- UNet with ResNet-101 architecture

For further information about ResNet architectures refer to [14].

### 3.3 Result of semantic segmentation

The four models are trained with the training set. The hyperparameter which were varied are shown in the following table.

**Table 1** Hyperparameter tuning parameters

Hyperparameter	Value
Batch size	8, 16, 32, 64
Learning rate	$10^{-2}$ - $10^{-5}$
Optimiser	SGD, ADAM

The training is executed on GPU Nvidia Quadro and Tesla (Nvidia Corp, Santa Clara, California) for about 100 – 150 epochs depending on the model architecture. To further improve the performance of the models, an evaluation is carried in finetuning the hyperparameters optimiser, batch size and learning rate. Fig. 9 shows some classification results for the classes background, seam, bubbles and burnings which are generated using the UNet model. A good agreement between the ground truth and the classification result can be observed. It can be seen that an intact weld seam remains transparent similar to the background. The clamping devices shows some scratches which reflect the light. If the laser power is increased, bubbles occur in the middle of the seam. In this area, the material loses its transparency. Burnings on the sample lead to an overexposure of the camera sensor so that only the burnings can be observed. The surrounding material is underexposed and remains dark. Apart from burnings, bubbles can also be observed.

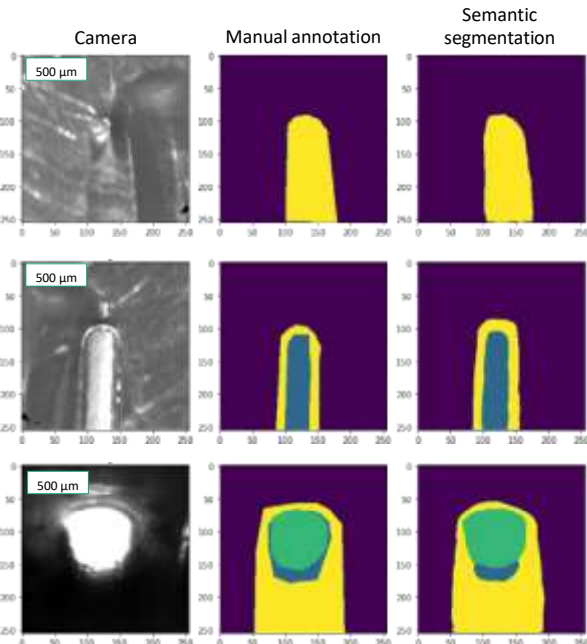


Fig. 9 Three exemplary classification results. Each row shows a different example. Left: image frame, middle: ground truth, right: semantic segmentation.

The classification results can be evaluated using the Intersection over Union value (IoU). As it can be seen in Fig. 10, the IoU value is defined as the area of overlap contained between the ground truth and the predicted (segmented) mask divided by the union between them [7].

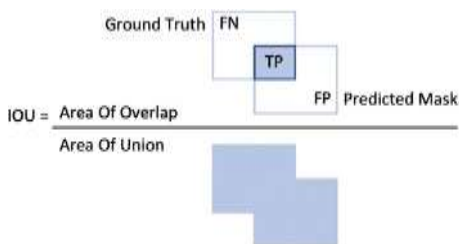


Fig. 10 Definition of the intersection over union metric.

The IoU values are calculated for each model (see Tab. 2). All models show a high accuracy with an IoU value of 0.92 – 0.93. This high agreement between the ground truth and the classification output can be explained by the high proportion of the background which covers the image frames. Here, the IoU value of the background is 0.96 – 0.97. The classes seam and bubbles have IoU values of 0.79 – 0.82 and 0.88 – 0.89 respectively. The lowest accuracy is achieved for the burnings with a value of 0.74 – 0.79.

Table 2 Summary of IoU values.

Model	UNet	UNet_	UNet_	UNet_
		ResNet18	ResNet34	Resnet101
IoU Total	0.93	0.93	0.92	0.93
IoU Background	0.97	0.97	0.96	0.97
IoU Seam	0.82	0.82	0.79	0.81
IoU Bubbles	0.88	0.88	0.84	0.89
IoU Burnings	0.79	0.79	0.74	0.78

In order to better identify the reasons for the classification errors, some exemplary outputs are presented in Fig. 11. It can be seen in the upper picture, that unfavourable light reflection causes the model to classify the seam as bubbles. Furthermore, scratches on the aluminium plate of the clamping device are identified as seams. Problems also occur when there are burnings on the sample as the sensor is overexposed.

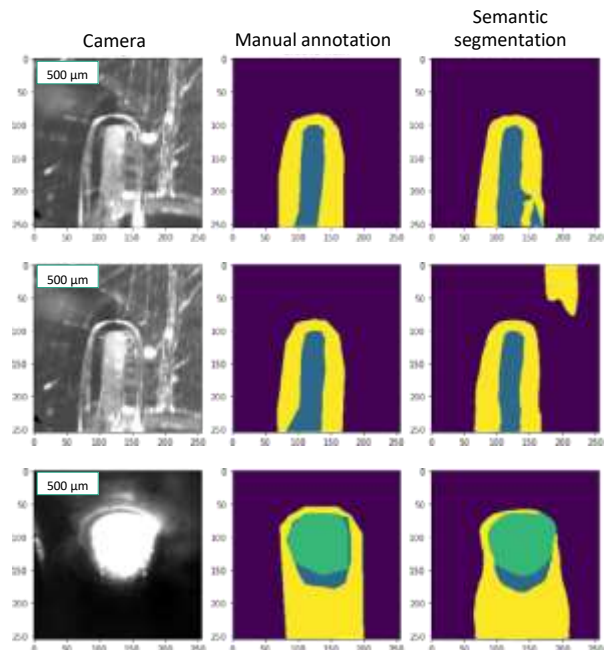


Fig. 11 Three exemplary defective classification results. Each row shows a different sample. Left: image frame, middle: ground truth, right: semantic segmentation.

In the next step, a regression analysis is carried out to predict the laser power on the basis of the segmentation results. The data set for this task is created by extracting the number of pixels for each class. The following table shows an example of the structure of the regression analysis dataset.

**Table 3** Structure of the regression analysis dataset.

Sample No.	Background [px]	Bubbles [px]	Burnings [px]	Seam [px]	Laser power [W]
1	302111	13385	0	73304	14.1
2	216229	6440	92907	73224	47.67
3	319247	32280	0	37273	7.48
4	369279	13704	0	5817	8.3

The number of pixels for the classes background, bubbles, burnings and seam are passed into the model. The output will then be compared to the laser power.

### 3.4 Prediction of process parameters using regression

In this chapter, the process parameters are predicted using different NN architectures. The dataset (see example in Tab. 3) is split up into a training and test set with a ratio of 90:10. Additionally, the dataset is pre-processed using scaling and normalisation to improve the model’s performance. The optimal parameters for the batch size and optimiser are identified using hyperparameter tuning and the networks are trained for 500 epochs. As evaluation metric, the root mean squared error (RMSE) and the coefficient of determination  $R^2$  are used.  $R^2$  is defined as the proportion of the variance in the dependent variable (laser power) that is predictable from the independent variables (number of pixels of the classes). It can be calculated as follows [5]:

$$R^2(y_{true}, y_{pred}) = 1 - \frac{\sum_{i=1}^n (y_{true} - y_{pred})^2}{\sum_{i=1}^n (y_{true} - \bar{y}_{true})^2} \quad (1)$$

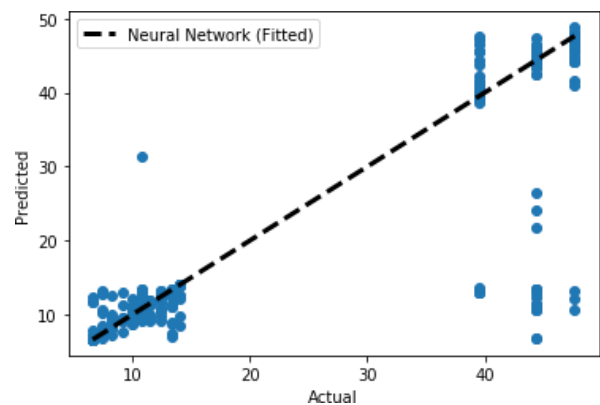
Here,  $y_{true}$  is the true laser power,  $y_{pred}$  is the predicted value and  $\bar{y}_{true}$  is the mean of all true laser powers. If  $R^2 = 1$ , all values can be perfectly depicted with the regression model. Therefore, the higher the number is, the better the model fits the dependent variables. The following table shows the results of three models where the highest accuracy was achieved.

**Table 4** The used models and the resulting  $R^2$  and RMSE values

No.	Model structure	Batch size / Optimiser	$R^2$ / RMSE
1	512, 256, 128	64 / ADAM	0.59 / 10.06 W
2	512, 512, 256, 256, 128, 128	64 / SGD	0.58 / 10.09 W
3	512, 512, 256, 256, 128, 128	32 / ADAM	0.6 / 9.92 W

For example, a model structure of 512, 256, 128 means that the NN consists of three layers with 512 neurons in the first, 256 in the second and 128 in the third layer. The results show that a  $RMSE > 9$  W and  $R^2$  value  $> 0.58$  are achieved which are calculated based on the predicted and real laser power. It can also be seen that a simple model with 3 layers achieves a similar performance than models with more layers. In order to better estimate the distribution of the error, the predicted and the actual laser power are

plotted in a fitted graph for model no. 3 as an example (see Fig. 12).



**Fig. 12** Fitted graph for model no.3. The graph shows the comparison between the predicted (vertical axis) and the actual (horizontal axis) laser power.

The x-axis shows the actual laser power values and the y-axis shows the corresponding prediction of the neural network. Ideally, the prediction should be identical to the actual laser power values so that all points are located on the dashed line. It can be seen that there is a good agreement with most predicted values in the lower laser power range between 1 – 15 W. Therefore, all points are located near the dashed line.

However, at laser powers  $> 40$  W, higher deviations can be detected. In some cases, the neural network outputs a laser power between 5 – 10 W which leads to a high RMSE. The reason behind this is that at higher laser powers, burnings occur in the sample. As already mentioned in the previous chapter, this leads to an overexposure of the sensor which makes subsequent prediction of the laser power more difficult. Based on the shown results in Fig. 12, the  $R^2$  and RMSE values can be calculated on the basis of the test set.

### 4. Conclusion and outlook

In this work, the absorber free LTW process was analysed using computer vision and deep learning algorithms. The welding process was recorded at different laser powers. The extracted image frames were manually annotated and then used to train different neural networks in order to perform semantic segmentation. The models were able to detect thermal damages in the welding seam with a good agreement between the ground truth and the predicted mask.

The geometric pattern of the feature areas was provided as input to the regression models to predict the laser power. The analysis was carried out using a NN regressor with a  $R^2$  score between 0.5 – 0.6. A good agreement can be found at lower laser powers from 1 – 15 W where at higher deviations occur at laser powers  $> 40$  W. The reason for this observation is that burnings occur at higher laser powers which leads to an overexposure of the camera sensor. This on the other hand makes a prediction more difficult.

In future works, the data set should be increased to further improve the classification performance of the model. This can be done by recording more videos or by further applying data augmentation techniques. Another step in

improving the classification performance is to improve the quality of the image frames. As it could be seen, unfavourable light reflections which were caused by scratches on the clamping device are influencing the results.

In order to establish this method as a new process monitoring technique, the future work should focus on an in-process analysis where the welding process is analysed in real-time. It might be necessary to facilitate the model architecture in order to achieve a low latency. Furthermore, a closed-loop control of the laser power should be established.

## References

- [1] E. Haberstroh and W. M. Hoffmann: Proc. 4M2007, (2007) 1.
- [2] N. Nguyen, M. Brosda, A. Olowinsky, and A. Gillner: J. Laser Micro Nanoeng., 14, (2019) 3.
- [3] V. Mamuschkin, A. Haeusler, C. Engelmann, A. Olowinsky, and H. Aehling: J. Las. Appl., 29, (2017) 2.
- [4] F. Maiwald, S. Englmaier, and S. Hierl: J. Laser Micro Nanoeng., 16, (2021) 1.
- [5] C. Knaak, G. Kolter, F. Schulze, M. Kröger, and P. Abels: Proc. SPIE, Vol. 11139, (2019) 1113905.
- [6] C. Knaak, U. Thombansen, P. Abels, and M. Kröger: Procedia CIRP, 74, (2018) 623.
- [7] A. Géron: “Hands-on machine learning with Scikit-Learn, Keras and Tensorflow: Concepts, tools and techniques to build intelligent systems” (O’Reilly, Sebastopol, CA, 2019).
- [8] M. Brosda, A. Olowinsky, and A. Pelzer: Physics Procedia, 83, (2016) 218.
- [9] A. Schmaizl, S. Steger, M. Dostalek, and S. Hierl: Proc. SPIE, Vol. 9736, (2016) 97361J
- [10] Y. LeCun and Y. Bengio: “Convolutional networks for images, speech and time series” (MIT Press, Cambridge, MA, 1995).
- [11] J. Long, E. Shelhamer, and T. Darrell: Proc. CVPR, (2015) 3431.
- [12] I. Goodfellow, Y. Bengio, and A. Courville: Deep Learning, 1, (2016) 99.
- [13] O. Ronneberger, P. Fischer, and T. Brox: Proc. MICCAI, (2015) 1.
- [14] K. He, X., Zhang, S. Ren, and J. Sun: Proc. CVPR, (2016) 770.

(Received: June 10, 2021, Accepted: December 1, 2021)