



# 3DWS: reliable segmentation on intelligent welding systems with 3D convolutions

J. Fernández<sup>1,2</sup> · D. Valerueva<sup>1,2</sup> · L. Higuero<sup>1</sup> · B. Sahelices<sup>2</sup>

Received: 16 January 2023 / Accepted: 30 September 2023  
© The Author(s) 2023

## Abstract

Automated industrial welding processes depend on a large number of factors interacting with high complexity resulting in some sporadic and random variability of the manufactured product that may affect its quality. It is therefore very important to have an accurate and stable quality control. In this work, a deep learning (DL) model is developed for semantic segmentation of weld seams using 3D stereo images of the seam. The objective is to correctly identify the shape and volume of the weld seam as this is the basic problem of quality control. To achieve this, a model called UNet $\mathcal{L}++$  has been developed, based on the UNet and UNet++ architectures, with a more complex topology and a simple encoder to achieve a good adaptation to the specific characteristics of the 3D data. The proposed model receives as input a voxelized 3D point cloud of the freshly welded part where noise is abundantly visible, and generates as output another 3D voxel grid where each voxel is semantically labeled. The experiments performed with parts built by a real weld line show a correct identification of the weld seams, obtaining values between 0.935 and 0.941 for the Dice Similarity Coefficient (DSC). As far as the authors are aware, this is the first 3D analysis proposal capable of generating shape and volume information of weld seams with almost perfect noise filtering.

**Keywords** Automated manufacturing · Deep learning · 3D convolutions · Semantic segmentation · Industry 4.0

## Introduction

Industrial welding lines are currently highly developed processes involving high mechanical, electronic and computer technology and a high degree of optimization (Xia et al., 2020). In recent years, automated welding lines have benefited from the added value coming from new technologies such as the Industrial Internet of Things (IIoT), cloud com-

puting, big data or digital twins, among others, giving rise to the concept of Intelligent Welding Systems (IWS) (Wang et al., 2020a). Multiple works have been recently developed targeting monitoring and adaptive controlling techniques to optimize real time parameters and improve product quality (Cheng et al., 2020; Lu et al., 2020; Chen et al., 2019; Wang et al., 2020b; Cai et al., 2020; Targ et al., 1999; Melakhsou & Batton-Hubert, 2021; Cai et al., 2021; He & Li, 2016; Xiao et al., 2022). However, there is a gap in research regarding the possibilities that new technologies provide to integrate the welding process in an intelligent manufacturing environment capable of providing high value-added services such as an advanced quality control with strict safety requirements.

IWS frameworks are made up of two conceptual components, the physical system and the data flow processing (Wang et al., 2020a). The physical system of the welding line includes the human operators, the controller, the power source, the wire feeder, the robot, etc (see Fig. 1). The data flow enables intelligent management of control, monitoring, diagnosis, quality and maintenance systems (Bacioiu et al., 2019; Liu et al., 2022b; Miao et al., 2022; Cai et al., 2021). Within this data flow, data acquisition and processing are

---

✉ D. Valerueva  
dilyana.valerueva@estudiantes.uva.es

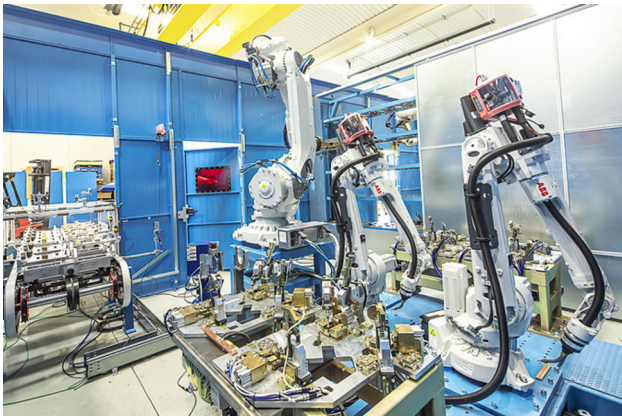
J. Fernández  
jesus\_fi@wip.es

L. Higuero  
luis\_hc@wip.es

B. Sahelices  
benjamin.sahelices@uva.es

<sup>1</sup> Research and Development Department, Amber Intelligence, Valladolid, Spain

<sup>2</sup> GCME Research Group, Department of Informatics, University of Valladolid, Plaza del Colegio de Santa Cruz 8, 47002 Valladolid, Spain



**Fig. 1** Photograph of the mechanical component of the weld line used in this work

integrated to generate a quality control subsystem. A strict, robust and reliable quality control is a high priority requirement for the manufacture of certain types of parts, among which are those involved in automotive safety. Quality control is a systematic and repetitive task, which must be carried out objectively, i.e., isolated from the relative perception of a human observer and based solely on the analysis of the data collected by the sensing system. These characteristics make quality control a very inappropriate task to be carried out by human beings, fundamentally because its repetitive nature generates tedium and favors subjective appreciation, giving rise to errors that are very difficult to detect and avoid.

The operation of the physical component of IWS depends on multiple variables that cannot always be controlled by the sensing and control system. Therefore, unexpected errors of a different nature may occur, such as problems with wire feeding, part positioning errors, gas pressure problems, human errors or other similar problems. For this reason, it may happen that defective final products may be periodically manufactured with very low frequency, which makes them more difficult to detect.

The development of reliable automated systems for quality control is one of the main challenges at IWS. Recent advances in machine learning (ML) and, specifically, in deep learning (DL) (Minaee et al., 2022), show that this line of research is very promising due to its great flexibility, robustness and scalability. Techniques based on 2D data from welding images (Bacoiu et al., 2019; Cheng et al., 2020; Singh & Desai, 2022) generate interesting results but are limited in their ability to identify defects related to the volume of material applied in the weld, an important element for achieving strict quality control. Recently, works have been developed in which 3D data is used to control the welding process (Miao et al., 2022; Liu et al., 2022b; Wang et al., 2022), but are based on either time-dependent analysis or 2D projections of the data collected.

In this work we propose to develop a complex data analysis model based on 3D data extracted from an existing gas metal arc welding welding line with high safety requirements. By having access to an existing manufacturing line we can use real data which generates additional added value to our results. The use of 3D point clouds obtained from real welding processes opens the door to the extraction of a new set of features linked to volumetric analysis of the weld material with great potential for improving strict quality control processes. In addition, a study of the ability of our DL models to eliminate the noise inherent in 3D data capture devices will be performed. The 3D data is processed using a UNet (Ronneberger et al., 2015) type DL architecture. The main advantage of using 3D data is the inclusion of new analyzable features in the weld that allow for more comprehensive quality control. As far as we know, this work is the first proposal for convolutional extraction of 3D information for analysis of the volume of weld seams in the IWS field. The results obtained validate this technique for quality control with strict requirements.

The remaining of the paper is organized as follows. In “**Related work**” section we describe similar works carried out to date, in “**Dataset**” section we explain the acquisition, processing and labeling of the data, in “**Models and training**” section we show the models used, the hyperparameter selection and the training techniques, in “**Results**” section we present and explain the results, in “**Manufacturing applications**” section we analyze the deployment of our solution in a real environment and, finally, in “**Conclusion**” section we summarize the conclusions of this work.

## Related work

In recent years, several scientific papers have been published on the use of DL technologies in various fields of industrial processes. In general, research is being carried out very actively in different areas of smart manufacturing such as procurement supply chain optimizations (Liu et al., 2022c) and complex data analysis for quality prediction in industrial applications (Liu et al., 2022a).

There are also multiple recent research papers in the area of control of the welding process and management of its data flow. Bacoiu et al. (2019) uses video images of the welding arc and processes them using a simple convolutional neural network (CNN) model to perform a simple classification process. It achieves more accurate results than those obtained with traditional models, but more importantly, it is more resilient and adaptable to variations in the input data. The results achieved show the good adaptability of the DL to this type of problems.

Work in Cheng et al. (2020) focuses on image composition to detect excessive weld penetration into the base plate,

which compromises the quality of the welds but is directly unobservable. CNN-based DL technology is used to analyze the images and correlate with weld penetration. The results show that DL is a promising technique in this problem even though the study is done on a very specific characteristic and it is difficult to generalize.

In Yang et al. (2020) the authors perform transfer learning (TL) of a CNN-based DL model (VGG model) to classify weld seams defects. Their results show that TL reduces training times and improves the accuracy of their data model. It should be noted that the defect identification is performed with a limited number of weld defects.

The work in Wang et al. (2020b) monitors growth and penetration control in arc welding processes on a digital twin. For weld control, traditional image analysis methods are compared against CNN. The CNN models estimate the geometry parameters of the weld in progress, thus allowing the detection of the optimal welding point. In the results presented the DL method is shown to be superior to traditional image analysis techniques.

The research work in Dai et al. (2021) uses DL models similar to YoloV3 (Redmon & Farhadi, 2017) to detect the location and quality of spot weldings on the car body. Part of their work involves creating a network with lower computational requirements for use in a resource-constrained industrial manufacturing environment. Their results show substantial improvements over traditional techniques in the exact location of welding spots, even in models that have been lightened by reducing their number of parameters. However, the quality analysis performed is limited in terms of the features extracted from the weld data.

Very few works have been found about 3D visual data analysis of industrial welding processes. Among them, the work in Miao et al. (2022) presents a DL model to perform two-step welding control, a graphical analysis based on wavelets of eddy currents together with an analysis of welding images taken from a 3D sensor. In this second dataset the convolutional analysis is performed in 2D so no specific features of a volumetric analysis are generated. Its results indicate great effectiveness although the possibility of extracting features is limited by the type of 3D data processing performed. Finally in Liu et al. (2022b) a 3D convolutional analysis is performed to extract features from the spatiotemporal domain of the welding process in real time. The data are obtained by acquiring images in rapid sequence. Additionally, and to solve computational limitations, the authors develop techniques to lighten the models. The results are more accurate and robust compared to those obtained with other traditional techniques.

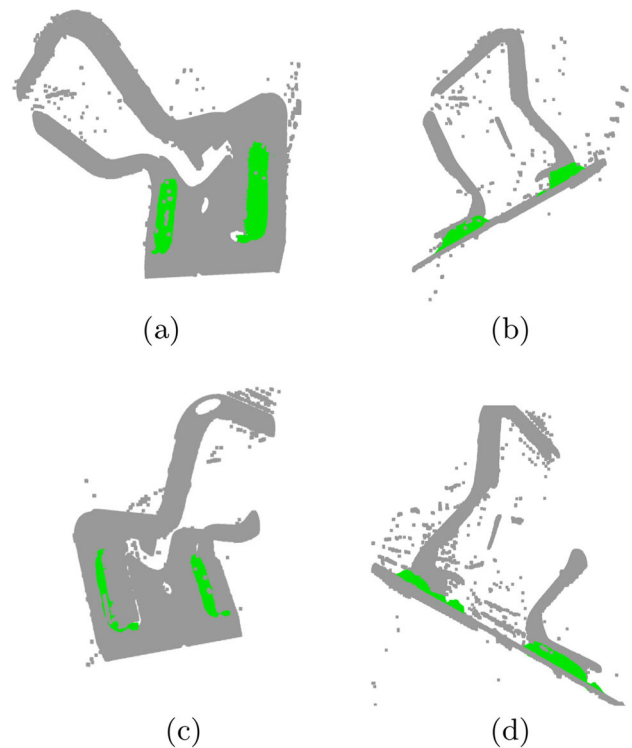
In the present work we go an step further, exploring the possibilities presented by the semantic segmentation of 3D point clouds of a real industrial welding process. The objective is to extract a new set of features related to 3D data to

improve the levels of reliability, stability and accuracy of IWS.

## Dataset

The physical context of this work is an existing automated welding line of metal parts (see Fig. 1), currently in production status. Although it is a highly automated machine, it is necessary to include a quality control system in the data flow, because defective parts are sporadically generated due to multi-source errors which are difficult to avoid, as explained in “Introduction” section. For this purpose, a data acquisition system is integrated, consisting of a 3D camera that generates point clouds for each manufactured part. The parts manufactured by our machine have strict quality requirements because they are components of an automotive safety subsystem. Each part consists of a metal base and a two-rods handle that is joined to the base by four weld seams, see Fig. 2.

We describe in this section how we generate the dataset to perform the training of our DL models and how we label this data to separate the weld seam from the part and from the void and noise.



**Fig. 2** Two 3D captures of the weld part. The first capture is viewed from two angles (a, b) and also the second capture (c, d). The parts are labeled and the weld seams are identified in green (Color figure online)

## Preprocessing

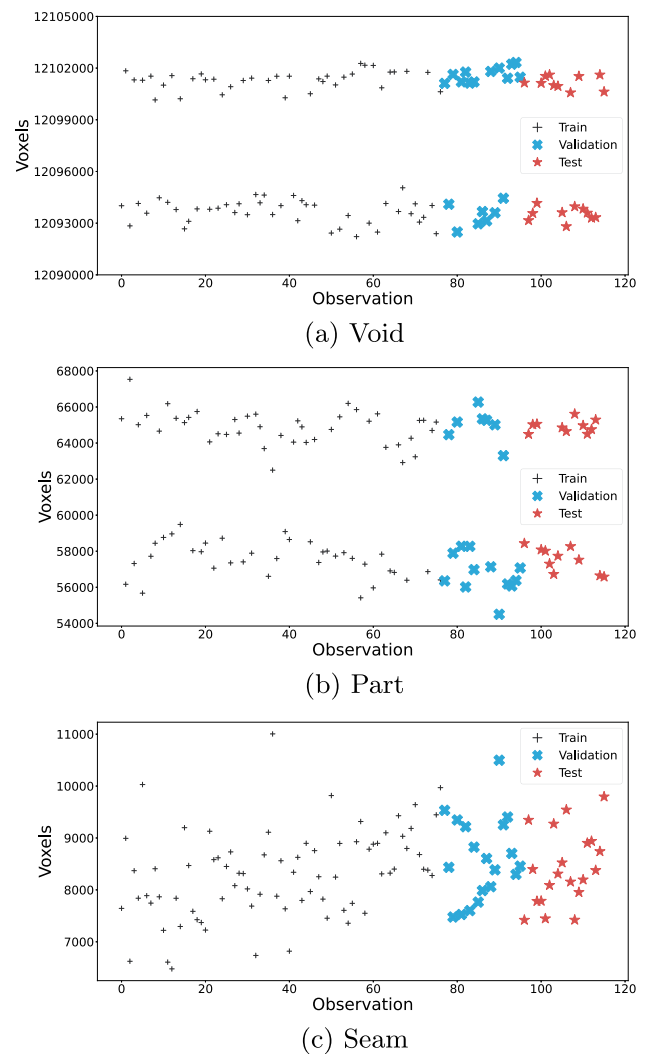
A 3D stereo image sensor is used to capture weld seam data immediately after fabrication. This device is able to capture the 3D point clouds but it is dependent on capture angles, glitter and shadow areas, so it is necessary to adjust the capture positions. Capture angles should ensure good visibility and the number of captures should ensure the mutual compensation for all shaded areas. In this case, due to the shape and geometry of the part, two different captures were necessary for each one in order to have complete visibility of all seams.

One of the main drawbacks of this type of imaging capture devices is that they require complex techniques to minimize the adverse effects of noise. The noise present in the point clouds is due to two factors. The first one is the vapours present while the weld seam is still incandescent. The structured light pattern emitted by the capture device bounces off the vapour, causing it to appear in the point cloud. The second one is the presence of specular surfaces on the steel of the weld seam. This can cause, depending on the angle of incidence of the light, reflections that distort the point cloud. The combination of both situations causes the noise in the point clouds to present very irregular patterns that are difficult to filter out. Those noisy points are also called outliers. Multiple works can be found in the literature devoted to the detection of outliers in point clouds (see (Balta et al., 2018)) but in this work it has been decided not to apply noise reduction techniques in the capture phase, and instead to integrate this function within the DL model developed. The aim is thus not only to evaluate the ability of these DL models to correctly identify the weld seams and their characteristics, but also to isolate the noise.

Once the point cloud has been captured by the stereo 3D image sensor, the first task is to center it in a coordinate system, i.e. the midpoint of the cloud on each axis is located at the middle point on each axis of the coordinate system. The next step is voxelization, which consists on taking the 3D point cloud and adapt it to a normalized cube which dimensions are  $230 \times 230 \times 230$  voxels. These are the dimensions of the tensors used by the neural network for both the dependent and the independent variables. The values contained in the dependent variable correspond to the voxel labels as explained in “Dataset” section. The cube, or voxel grid, is used as the input for the DL model. As we are using a resolution of 4 voxels per millimeter, the cube captures a length of 57.5 mm on each axis.

## Labeling

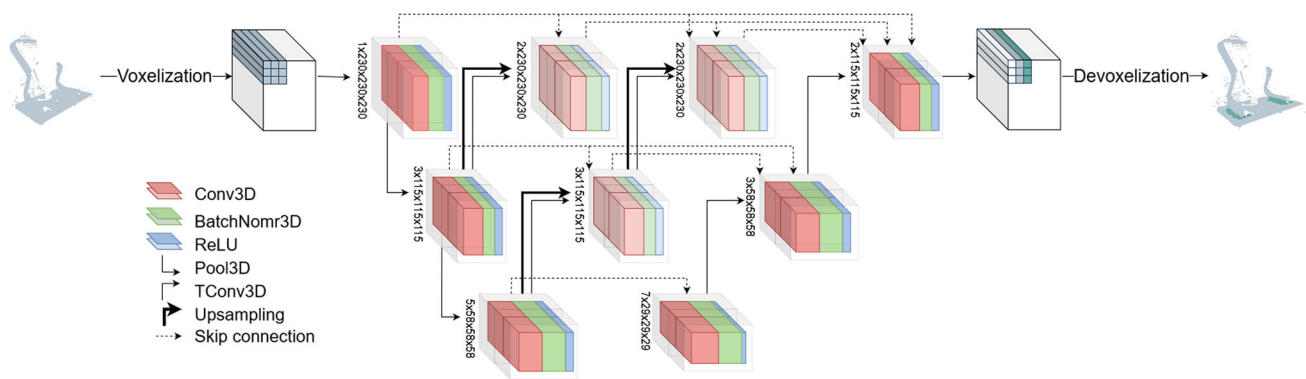
Labeling in 3D semantic segmentation problems is the task that classifies each point in the 3D cloud. In our problem, we identify three classes, i.e. part, void and weld seams. To



**Fig. 3** Number of void, part and seam voxels for each observation of the training, validation and test sets

achieve the labeling, it is necessary to distinguish the points that make up the part from those that make up the weld seam. The remainder is assigned by default to the void class. For the labeling purpose we use the Semantic Segmentation Editor (Automotive & Laboratory, 2021).

It is expected that most of the information corresponds to the area of the space without part, which we call void. In our case, on average, 99.43% of the voxels correspond to the void. Of the remaining voxels, 0.5% identify the part and the remaining 0.07% identify the weld seams. The precise distribution of the number of voxels representing the void, the part and the weld seam for all captures is shown in Fig. 3. As each part is represented by two captures from different angles, it can be observed that for the void and part labels there is a slight difference in the number of voxels of each of the two captures, as seen in Fig. 3a and b. However, for the seam plot there is no appreciable difference between the two



**Fig. 4** Schematic summary of the models used in this work. The basic UNet model is formed by the modules with a more vivid color. UNet++ is distinguished by a more subdued color in its added modules. Finally, our UNetL++ proposal includes new links that appear in the figure marked with thick strokes

captures (Fig. 3c) because the number of voxels that make up the seam is approximately the same for both captures. The most relevant aspect of Fig. 3 is that the voxel distribution pattern remains stable between the training, validation and test sets. Overall, the distribution is homogeneous between the 3 types of labels as well as between the training, validation and test sets. As can be seen in Fig. 3, in order to achieve a correct training of our models one of the main problems we have to solve is the clear imbalance of the classes. We will explain in the “Models and training” section how we have tackled with this problem.

As previously explained, to minimize shadow areas two data captures are taken for each part. Figure 2 shows an example of the two captures from a single part with two different views of each one. In Fig. 2a appears the part captured from one of the angles with two of the weld seams identified and labeled and the other two seams in the shadow, so unidentified. Figure 2c depicts the same part, captured from another angle. The weld seams that previously appeared in shadow are now identified and the seams that were previously identifiable now appear in shadow. Figure 2b and d respectively correspond to the same captures but are shown in different angles in order to better perceive the noise and the distribution of material in the weld seams. In the part represented by Fig. 2a, certain points on the left seam can be seen that are apparently not part of the seam. But, when the angle of vision is changed, see Fig. 2b, it is discovered that actually these spots are noise that had coincided in the line of sight. Analogously the same effect can be seen in Fig. 2c and d in the left and right seams. In addition, in Fig. 2d can be clearly appreciated the profile of the left seam in which is visible an apparent lack of welding material in some areas. All this information in 3D will be used by our models to efficiently filter noise and identify the weld seam, as will be shown in “Results” section.

## Models and training

### Models

The models we propose in this work are based on the UNet (Ronneberger et al., 2015) and UNet++ (Zhou et al., 2018) architectures which are deep neural networks based on encoders/decoders used for image generation, see Fig. 4. The encoder performs convolution and sub-sampling operations in order to reduce the resolution of the inputs and extract their most important spatial characteristics. The decoder operates symmetrically to the encoder, increasing the dimensionality of the data residing in the latent space in order to reconstruct an image with the original resolution. Between the encoder and the decoder there are several links which mitigate the information leakage caused by the encoder and enhance the reconstruction. These models have shown good results in the literature for 2D data. The change to 3D data implies an exponential increase in the amount of information to be processed and it is therefore necessary to adapt the model to the new requirements. In particular, 3D convolutions require extracting features not only linked to each  $x$ - $y$ -plane but also to the  $z$ -axis. In other words, the number of features to be managed is greater while maintaining the same basic convolution operation.

The proposed model, UNetL++, where  $\mathcal{L}$  stands for the set of new links introduced, is an evolution of UNet and UNet++ whose purpose is to adapt to 3D data processing with a twofold objective:

- Mitigate the loss of information in the reconstruction and increase the linkage paths between the contraction and expansion phases due to the exponential increase in data. For this purpose, the UNet++ model has been evolved by increasing the connections between the encoder and the decoder (see Fig. 4). This evolution is characterized by an increase in the inner blocks reconstruction paths,

where we call inner blocks to those that are not present in a classical UNet architecture. Transposed convolutions and upsample operations are coupled in order to increase the quality and richness of the reconstruction and to enhance the transmission of information between the aforementioned stages.

- Limit the computational and storage complexity required for training and inference. We hypothesize that the basic convolutional operation is enough to extract the necessary features in the 3D data and therefore to carry out efficient training. In this way, the increase in resources implicit in the increased complexity of the architecture is offset by the reduction in resources by simplifying the encoder. For this purpose, we propose to use a simple CNN encoder which we call Basic-CNN.

Figure 4 shows the scheme of the proposed models previously explained. The main difference between UNet++ and our proposal UNet $\mathcal{L}++$  lies in the new internal propagation paths added, see upsample links in Fig. 4. This means that for each transposed convolution a new upsampling operation is added. Therefore the amount of information obtained in the voxel grid regeneration procedure is not only duplicated but also contains different features, because convolution and upsample operations generate information of different characteristics. That is, our proposal UNet $\mathcal{L}++$  increases the connection paths between the encoder and the decoder, enhancing the richness and quality of the information transmitted and processed in the internal blocks of the architecture.

The encoders used in this design, Basic-CNN family, have lower computational and storage requirements than state-of-the-art encoders such as ResNet (He et al., 2015) and DenseNet (Huang et al., 2017), with which a comparative analysis is performed in “Results” section. It is observed, compared to the UNet++ model, that our proposal includes a more sophisticated link topology, duplicating inner block reconstruction paths with a different reconstruction operation.

## Training

The supervised training phase iteratively adjusts the parameters of the model to accurately achieve segmentation of each of the weld seams that appear in the 3D point clouds. The details of our labeled dataset have been described in “Dataset” section. Labeling 3D point clouds is a complex and time-consuming task because it must be performed manually. It is therefore very expensive to get a large enough amount of labeled data to guarantee the sufficient diversity needed by the DL models for a correct learning. A total of 116 parts were labeled, 20 of which were reserved to the test set and of the remaining parts, 20% and 80% to the validation and training set respectively. The test set is not used in the training

phase and therefore our models have never seen it, allowing for an objective validation of our results.

The initial trainings showed excessive overfitting, precisely because of the insufficient number of available parts. To solve the overfitting problem, two basic data augmentation operations, rotation and translation, were implemented. Since we are dealing with 3D data, both operations can be performed on any of the three axes. The choice of the rotation or translation axis is made randomly. Using an experimental methodology, the operating parameters of the two transformations were identified. Specifically, the translation has random values between 10 and 40 voxels, depending on the axis. Rotation is performed on random values of angles ranging from  $\pi/20$  to  $\pi/12$  radians. The data augmentation is implemented dynamically on the GPU during the training phase. Once the data is loaded into the GPU memory, data augmentation is applied with random parameters within the previously described range. Once the data has been augmented, it is used for the training of the neural network in the current epoch. This operation is carried out in each epoch so its data is different from the data of the rest of the epochs. The probability of data repetition is practically nil. The number of epochs used for training each model is 250, which implies that approximately 250 different versions of the original dataset have been used. This way, the low variability due to the small number of available parts is mitigated.

The use of 3D point clouds implies that the number of voxels representing empty space is much larger than the number of voxels representing parts and this, in turn, is much larger than the number of voxels identifying the weld seam. In particular, as described in “Dataset” section, more than 99% of the dots correspond to the void. The problem is known as class imbalance and causes results during training to be biased by the majority class. This reduces the sensitivity of our results, i.e., the ability to correctly identify the weld seam spots. To solve this problem, we have used a variant of the cross-entropy loss function which applies weights to the different classes and allows their parameterization. Its mathematical expression can be seen in the Eq. (1) where  $y_c$  is the true value of the label,  $x_c$  is the predicted value and  $w_c$  is the weight which penalize the class  $c$  (Naceur et al., 2020). The values of  $w_c$  are selected based on the average ratio of the voxels belonging to the part relative to the voxels belonging to the weld seam, which are 0.77 and 0.23 respectively. To offset the imbalance between these two classes the inverse of the above ratio is assigned, i.e., a value of 0.23 for the voxels belonging to the part (majority class) and a value of 0.77 to the voxels belonging to the weld seam (minority class). A zero value is assigned to the voxels corresponding to the void to prevent them from interfering with the loss function since these voxels are easily identifiable at the network input (value zero) and are directly copied to the output. Thus, the network is forced to minimize its loss function by assigning

a greater weight to the voxels belonging to the weld seam class. This makes the parameters resulting from the training phase very sensitive to this minority class and therefore the network is also sensitive to this class. This new loss function prevents the model from directing its learning towards the identification of the simplest and most common class, the void.

$$WCE = - \sum_{c=1}^C w_c \cdot y_c \cdot \log \frac{\exp(x_c)}{\sum_{i=1}^C \exp(x_i)} \quad (1)$$

The learning rate has been dynamically varied during training using the plateau reduction technique. This algorithm reduces the learning rate when a metric has stopped improving, in our case the Dice Similarity Coefficient (DSC). Learning rate is reduced by a factor of 2 if no improvement is observed for 20 epochs. For the same purpose batch normalization is used which leads to a significantly smoother optimization and a more predictive and stable behavior of the gradients during training (Ioffe & Szegedy, 2015; Santurkar et al., 2018). To accelerate the learning of the network we use the ReLU activation function and adaptive moment estimation or ADAM as a function optimizer. Finally, early stop training has been used to select the best model.

The time required for training the proposed model in Fig. 4 was, approximately, 10h on a single NVIDIA RTX A5000 GPU with 24-GB of memory which shows that the model proposed in this work is lightweight and capable of evolving to more complex datasets.

## Results

In this section we show the results of our models when applied to the test set, that is, 3D point clouds from parts fabricated on an automated weld line currently in operation which have not been used in the training phase. We start with a comparative analysis between our proposal and other equivalent proposals in the state of the art. In particular, we evaluate the ability of our encoder to generate results comparable to more complex encoders. Next, we evaluate the sensitivity of the model when varying its topological complexity in order to propose a usable model that requires a minimum amount of computational and memory resources. We show the ability of our model to replicate a human in the task of distinguish the weld seam from the rest, i.e., the part, the noise, and the void and, finally, we show a quantitative analysis of the noise filtering capability of our model.

We use the DSC as an index to evaluate the results, see Eq. (2), where TP or true positive refers to correctly identified weld seam voxels, FP or false positive refers to voxels identified as weld seam when they are not, and FN or false

negative refers to unidentified weld seam voxels. The most significant feature of the DSC index is that it measures the ability of the model to identify the weld seam, regardless of the size of the seam relative to the 3D cube representing the part. As can be seen in Eq. (2), the TN or true negative voxels, i.e. correctly classified voxels that are not weld seam, are not taken into consideration thus avoiding the negative effects of the class imbalance problem. The DSC index is equivalent to the F1 score metric and is a harmonic mean between the Precision, or quality of the identified weld seam voxels, and the Recall, or the proportion of identified weld seam voxels. The range of values of the DSC is [0,1] and the closer the value is to one, the better the performance of the model.

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (2)$$

## Encoder/decoder evaluation

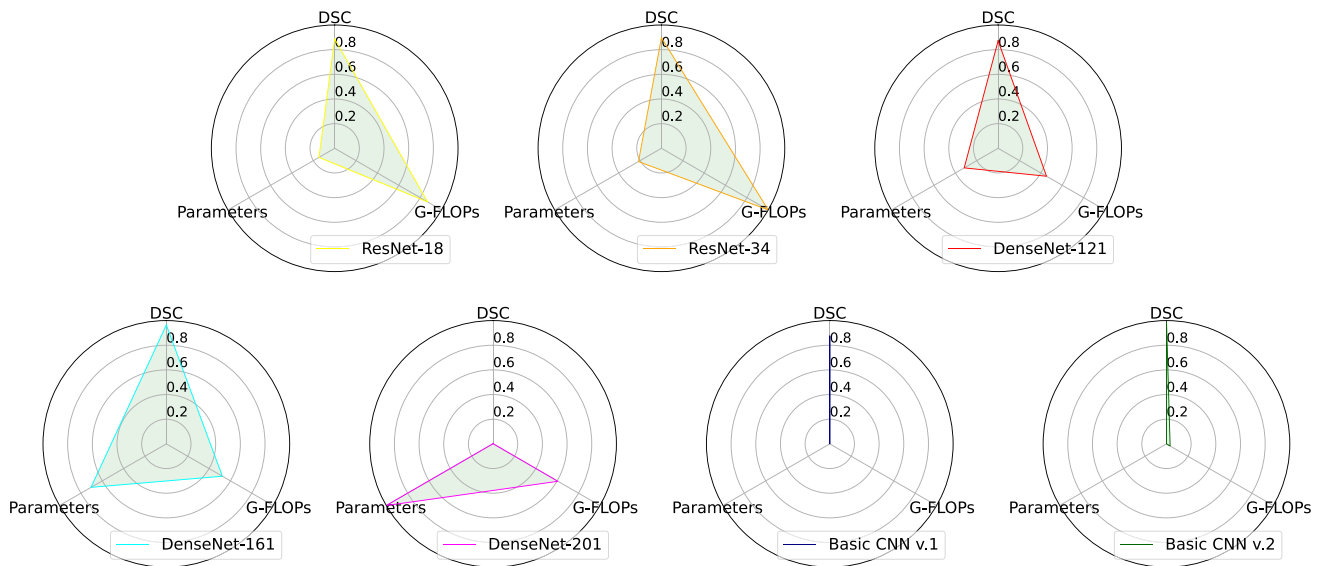
The target of this subsection is to evaluate the feasibility of our proposal and analyze its complexity. For this purpose we evaluate three types of encoders, two of them complex and high performance (ResNet and DenseNet) and the third our proposed simplified encoder (Basic CNN), using a 3D model based on a 3 level UNet architecture (see Fig. 4). The input of the model is a tensor of size  $230 \times 230 \times 230$  voxels in which each voxel represents the void or the existence of part/noise, and whose output is an identical tensor in which each voxel houses the void, part/noise or weld seam. As can be seen in Figs. 2 and 6 the data contain a large amount of noise which was explained in “Dataset” section. No specific noise suppression technique has been used since the correct identification of the weld seam implicitly performs this task. The DSC index used to identify the weld seam is an accurate measure of the noise suppression capability of our model.

Table 1 shows the evaluation results for the encoders ResNet and DenseNet as they are state-of-the-art proposals that generate very good results in semantic image segmentation. We also show the results of the encoder proposed in this work (Basic CNN). The quality of the results is measured with the DSC index and the associated Precision and Recall metrics. The complexity is assessed by two indicators: the size complexity, measured by the number of parameters, and the computational complexity, measured by the total number of floating point operations to perform on each inference. This value is expressed in Giga Floating-Point Operations (GFLOP) where each fused multiply-accumulate (MAC) operation is counted as 2 FLOP.

Table 1 shows the results of the Resnet-18, Resnet-34, DenseNet-121, DenseNet-169 and DenseNet-201 models, compared with the models proposed in this work, Basic CNN v.1 and v.2. For the state-of-the-art models (first five rows of Table 1) a DSC between 0.894 and 0.939 can be observed,

**Table 1** DSC, Precision, Recall, GFLOP and number of Parameters of the encoders used in the UNet model

Encoder	DSC	Precision	Recall	GFLOP	Parameters
ResNet-18	0.9357	0.9252	0.9487	1922.4	66,188,041
ResNet-34	0.9362	0.9161	0.9597	2205.3	96,501,385
DenseNet-121	0.9350	0.9289	0.9441	1024.0	142,657,217
DenseNet-169	0.9394	0.9317	0.9493	1178.2	316,735,681
DenseNet-201	0.8935	0.9266	0.8656	1356.2	449,160,385
Basic CNN v.1	0.9351	0.9221	0.9504	45.1	10968
Basic CNN v.2	0.9411	0.9301	0.9541	119.12	34586

**Fig. 5** Radar charts with min-max scaling showing scaled values of DSC, GFLOP and Parameters for each model in Table 1

i.e., values close to 1 that show a good trade-off between Precision and Recall in the identification of the weld seam. The same table shows the computational and storage complexity of each encoder. In this work we propose the use of a simpler encoder called Basic-CNN, which performs the basic convolutional operations with lower requirements than the ResNet and DenseNet encoders. We show results for two variants of Basic-CNN in the last two rows of Table 1, differing only in that the second version performs a width scaling by increasing the number of channels in the encoder layers, with its subsequent impact on the remainder of the architecture. The last two rows of Table 1 show that our proposal obtains similar results, even better in some cases, with much lower memory (0.01%) and computational complexity (5% to 11%).

Figure 5 shows the radar charts comparing DSC, GFLOP and the number of Parameters for the models in Table 1. It can be seen that our proposals based on Basic-CNN achieve similar or better results than state-of-the-art encoders, but at a fraction of the size and computational cost.

## Topology evaluation

In this subsection we evaluate the influence of the topological complexity on the quality of our model. For this purpose, we use the Basic-CNN encoder with the two versions described in “Results” section, and incrementally vary the topology and interconnections. We evaluate a UNet, a UNet++ and our proposal UNet $\mathcal{L}++$  described in “Models and training” section. The results can be seen in Table 2. The iGFLOP value represents the increase in computational complexity with respect to the basic UNet model, i.e. the computation derived from the insertion of the inner blocks, as described in “Models and training” section. The increase in computational complexity of our proposal is very slight compared to the state-of-the-art encoders shown in Table 1. Both the number of parameters and the GFLOP are similar in the three levels of topological complexity.

Our proposal UNet $\mathcal{L}++$  gets the best DSC result among all the proposed topologies. However, this improvement is marginal for the dataset used. As explained in “Models and training” section, our proposal increases the internal connections and also the information processing capacity in the



**Table 2** DSC, Precision, Recall, GFLOP, iGFLOP and number of Parameters of the proposed models as they become topologically more complex

Topology	DSC	Precision	Recall	GFLOP	iGFLOP	Parameters
Basic CNN v.1	0.9351	0.9221	0.9504	45.1	0	10968
Basic CNN ++ v.1	0.9412	0.9500	0.9344	38.6	11.28	9577
Basic CNN $\mathcal{L}++$ v.1	0.9414	0.9333	0.9513	68.4	36.99	11513
Basic CNN v.2	0.9411	0.9301	0.9541	119.12	0	34586
Basic CNN ++ v.2	0.9415	0.9367	0.9481	100.0	34.61	29147
Basic CNN $\mathcal{L}++$ v.2	0.9416	0.9360	0.9493	164.07	94.26	34033

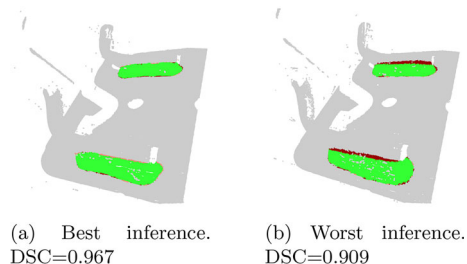
The encoder Basic-CNN (v.1 and v.2) is used for all the topologies

intermediate layers, thus significantly reducing the amount of information lost between the encoding and decoding stages. So, our proposal increases the number and complexity of the features extracted from the dataset. Since the weld seams of the current dataset have a regular volume and few complex morphological features, the results of our model do not differ much from those of the state of the art. However, the ability to identify and process complex features is critical when weld seams are more irregular. For this reason, we believe that our proposal will significantly improve the results in new welding lines in which more irregular parts are produced, as will be shown later in this section.

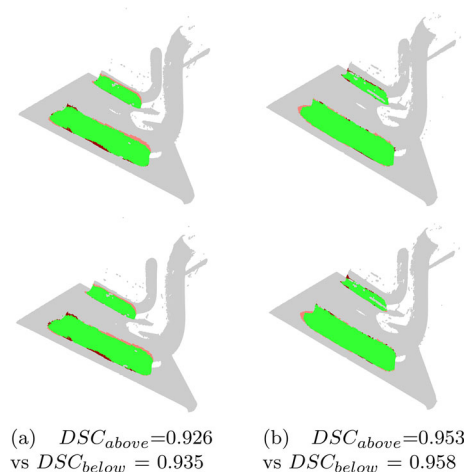
### Ability to replicate human model

The results in this subsection correspond to the model UNet $\mathcal{L}++$  with Basic-CNN encoder v.2. Figure 6 shows the best and the worst inference of the test set, i.e., parts that have not been used in the training phase. Incorrectly identified areas are represented in red tones and correspond mainly to the weld seam boundary which is difficult to classify even by a human. Dots in coral red represent regions erroneously categorized as no weld seam. Dots in dark red depict regions incorrectly classified as weld seam. The gaps seen in the weld seam are not inference faults but noise appearing in the line of sight. It can be seen graphically that the network identifies accurately the weld seams, with a DSC range between 0.9093 and 0.9667. These numbers show that our proposal filters the noise very efficiently. Although in Fig. 6 can be seen multiple points of noise apparently in the weld seam, it is just an effect of the viewing angle since almost the entire weld seam volume is identified. That is, if the angle of view were changed, the noise would disappear of the line of sight.

Using the same color scheme, Fig. 7 shows the performance of the model UNet++ against our model UNet $\mathcal{L}++$  using in both cases the same encoder (v.2). We have chosen two parts from the test set in which differences between the two methods can be seen. Although the differences are small, it can be seen that numerically the quality of the segmentation of the weld seam is better for our model. This small improvement can be important to provide a better quality control system, since the weld seam isolation may precede other pro-



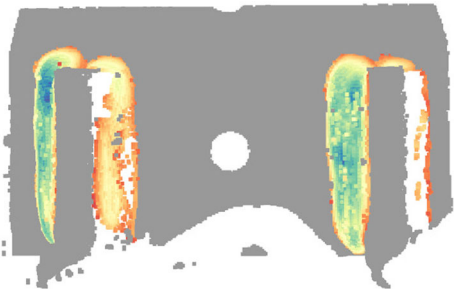
**Fig. 6** Best and worst inference in the test set. Green and red dots show respectively correctly and incorrectly identified areas (Color figure online)



**Fig. 7** Inferences of the UNet++ (above) and the UNet $\mathcal{L}++$  (below) for two parts (a, b) of the test set (Color figure online)

cedures to study and measure more advanced characteristics of the seam, as will be explained in “Manufacturing applications” section. In addition, as already argued, we expect that the differences between models may increase when the weld seam morphology becomes more complex.

Figure 8 represents each voxel as the probability of belonging to the seam class which is obtained by inference of a part from the test set. Cold colors represent higher probability and warm colors represent lower probability. This is a representation of how confident the neural network is in its inference results. The capture used in this example correspond to the view of the left seam of each rod. It can be seen that the left seams have a higher probability than right



**Fig. 8** Example of inference representing the probability of belonging to the weld seam class

ones, i.e. the neural network is clearly more confident in its inference.

Finally, an attempt has been made to compare our 3D semantic segmentation proposal with similar proposals. As far as we know, in the state of the art there is no equivalent proposal of 3D semantic segmentation. For this reason we have compared our model with a weld seam 2D semantic segmentation work (Wang & Mei, 2022). First of all, it must be clarified that it is a 2D convolutional system in which both the application environment and the data acquisition methodology are different. Our proposal generates a DSC of 0.942 while that of Wang and Mei (2022) is 0.989. These are good results in both cases. The difference between both methods can be explained by the greater complexity and amount of information involved in 3D segmentation tasks, which makes the segmentation process more challenging.

### Noise filtering performance

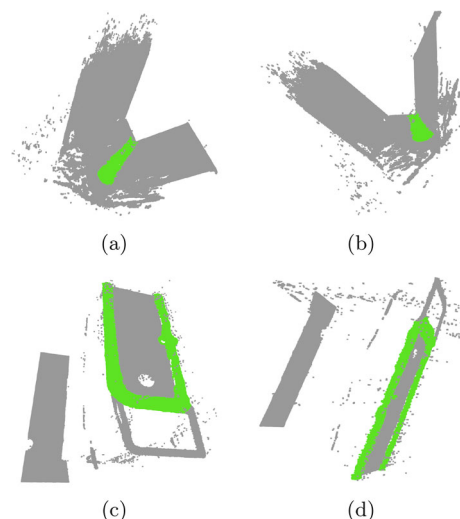
In this subsection we demonstrate the ability of the proposed model (UNet $\mathcal{L}++$  with Basic-CNN encoder v.2) to filter the noise. Table 3 shows, for each part of the test set, the number of noise voxels in the point cloud in the second column, in the third column how many of them have been assigned to the part or void class, i.e. have not been classified as weld seam, and in the fourth column the percentage. The proportion of noise voxels that have not been identified as weld seam, on average, is 99.23%. Such good results ensure that our model is able to isolate weld seams even in situations where reflections and smoke are present, which guarantees good performance in welding cell-dominated industrial plants deployments.

### Generalization capacity

In this subsection we prove the better generalization capacity of our model when applied to weld seams with more irregular and complex features. For this purpose, we build two new datasets. The first one,  $\mathcal{D}_1$ , is made up of 112 pieces with a T-joint weld seam (Fig. 9a, b). The second,  $\mathcal{D}_2$ , is

**Table 3** Number of observed noise voxels, number of those voxels predicted as part/void and noise filtering ratio for each of the 20 observations in the test set

Observation	Observed noise voxels	Predicted part/void voxels	Filtering ratio (%)
1	1249	1235	98.88
2	1307	1307	100.0
3	1201	1196	99.58
4	1671	1622	97.07
5	921	920	99.89
6	1304	1293	99.16
7	1532	1532	100.0
8	1228	1221	99.43
9	1243	1243	100.0
10	1173	1168	99.57
11	945	944	99.89
12	1062	1033	97.27
13	1140	1103	96.75
14	845	845	100.0
15	1462	1457	99.66
16	998	984	98.60
17	744	738	99.19
18	1097	1097	100.0
19	1095	1095	100.0
20	1061	1057	99.62
Average	1163.90	1154.50	99.23

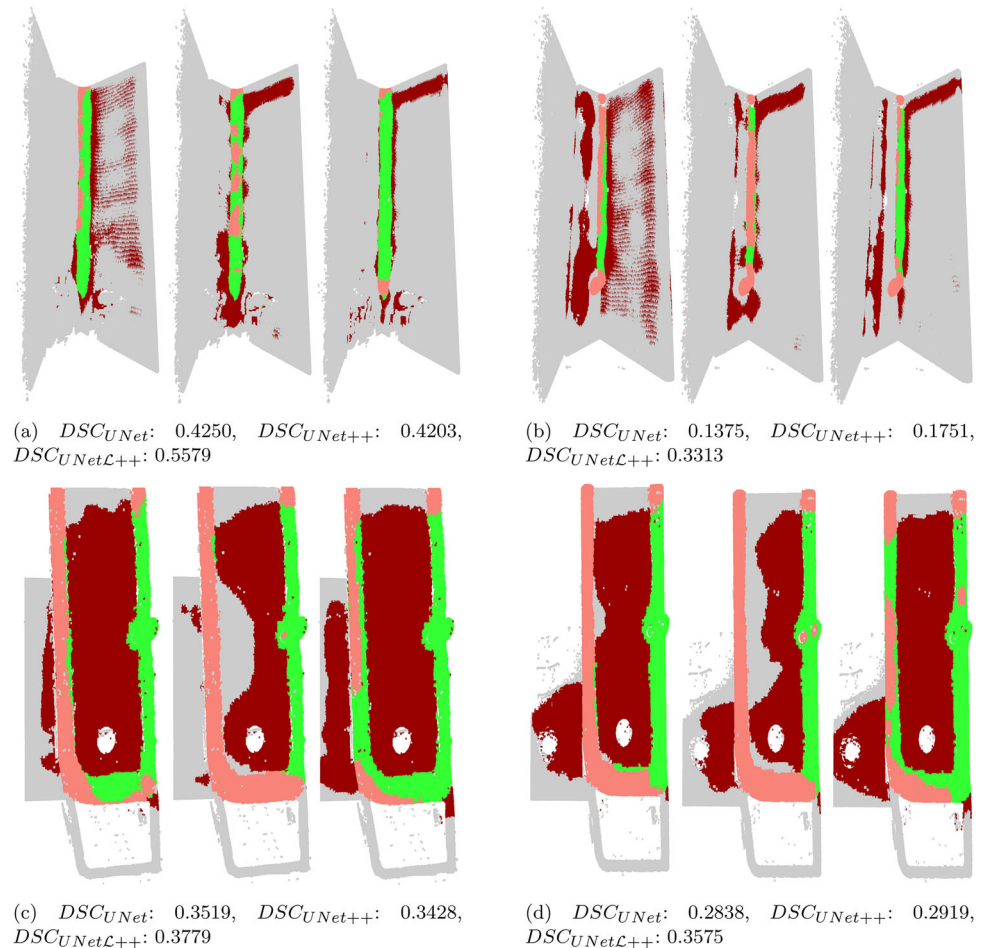


**Fig. 9** Parts with T-joint weld seams (a, b) and parts with rectangular overlap weld seams (c, d) (Color figure online)

composed of 22 pieces with a rectangular overlap type weld seam (Fig. 9c, d). We make inferences on these datasets with the models UNet, UNet++ and UNet $\mathcal{L}++$ , all of them with the v.2 encoder. It is important to emphasize that no training

**Table 4** Average DSC, Precision and Recall of the models when applied to two new datasets

Architecture	$\mathcal{D}_1$ : T-joint			$\mathcal{D}_2$ : rectangular overlap		
	DSC	Precision	Recall	DSC	Precision	Recall
UNet	0.2055	0.1378	0.4605	0.2306	0.1602	0.4296
UNet++	0.2724	0.2004	0.4682	0.2542	0.1969	0.3753
UNet $\mathcal{L}++$	0.3116	0.2226	0.5620	0.2654	0.1735	0.5877

**Fig. 10** Inferences examples for T-joint (a, b) and overlap (c, d) weld seams with UNet (left), UNet++ (middle) and UNet $\mathcal{L}++$  (right) (Color figure online)

or fine tuning has been carried out. The inferences on the new data are performed using the latent knowledge learned with the original parts, causing the DSC values to be lower than for the original dataset. The shape of the new parts and the length and morphology of the new weld seams are completely different from the original ones, so the performance of the models will be a useful measure of their ability to generalize when faced with completely different scenarios. The results can be found in Table 4. It can be seen that, for the two datasets, our model outperforms both the UNet and the UNet++. Regarding dataset  $\mathcal{D}_1$ , UNet $\mathcal{L}++$  improves the DSC of UNet and UNet++ by 51.63 % and 14.39 % respectively. In the case of dataset  $\mathcal{D}_2$ , the improvement is 15.09% and 4.41% respectively. In addition, the Recall of our model is recurrently much better, indicating the ability of UNet $\mathcal{L}++$

to find complex weld seams in completely different environments than those used to train it. The Precision does not differ much from that of the UNet++ model, but since the Recall is considerably higher this means that the proportion of voxels correctly identified as weld seam remains the same but with a much higher area of the weld seam identified. Again, using the same color coding as in the previous subsection, Fig. 10 illustrates a significant example of the performance of the three models. We have chosen two representative parts from each dataset in which differences between the three models can be seen. Numerically and graphically, it can be noticed that UNet $\mathcal{L}++$  offers a much more accurate segmentation than the other models. For the parts of both datasets, UNet $\mathcal{L}++$  successfully identifies a larger area of the weld seam (green dots) than the rest of the models, while the area

incorrectly classified as weld seam (dark red) is not significantly greater than in the rest of the cases. These results demonstrate the ability of UNet $\mathcal{L}++$  to generalize its knowledge and offer better performance in novel situations where more complex analysis is required.

## Manufacturing applications

This section aims to show the value added of applying our model in an industrial quality control inspection. The proposed model in this work allows to obtain an accurate 3D representation of the isolated weld seam, which in turn makes it possible to calculate very useful information for diagnosing the quality of the seam, such as its presence, length, width, volume and location. Moreover, having the 3D isolated weld seam allows for more complex analysis such as volumetric measurements that traditionally have been almost impossible to calculate without the use of destructive macrographic tests (Wang & Mei, 2022). In addition, our segmentation model can be embedded in a processing pipeline where the segmented seam is fed into other DL algorithms specialized in other fine-grained defects such as porosity, projections and irregularities. The quality of the segmentation obtained through our proposal allows for a precise identification of these defects.

The performance of our proposal could be a limiting factor for its application in an automatic quality control of robotic welding cells. The process from the capture of a raw point cloud until the segmentation is done, involves several stages including the centering of the point cloud, the voxelization, the model inference, and the selection of the voxels categorized as weld seam. Using the hardware described in “Models and training” section and a high-performance compiler for voxelization (the heaviest process), the complete process can be performed at around 27 parts per second. Typically, the time elapsed between robot movements and data capture is at least two seconds. Hence, a single computing unit could simultaneously cater to multiple independent inspection stations without slowing down the cycle times of the manufactured parts. Furthermore, resource usage by the system hosting the solution increases linearly as the workload increases, thus constituting a scalable system.

## Conclusion

In this paper we propose a DL model that analyzes stereo images using 3D convolutions for the semantic segmentation of weld seams in a real intelligent welding system environment. The DL model receives as input a voxelized 3D point cloud of the part that has just been manufactured and generates as output a 3D voxel grid in which each voxel is labeled. The proposed model, UNet $\mathcal{L}++$ , is a topological

enhancement of UNet++ using a simple CNN encoder. This model uses less than 0.01% of the number of parameters and between 5 and 11% of the computational complexity of comparable proposals, but achieves very high quality results with DSC similar or superior to those systems. Specifically, by applying UNet $\mathcal{L}++$  to the test set, a Precision between 0.93 and 0.94, a Recall of around 0.95 and, therefore, a DSC of around 0.94 are achieved. These values reflect that the geometry and volume identification of the weld seams is very accurate, which opens the door to the development of precise defect detection systems. Nearly complete noise filtering is also achieved, which is an important advance for the stability of quality control systems. Additionally, a study of the ability to generalize of our proposal when applied to more complex and irregular pieces has been carried out. This study shows that our UNet $\mathcal{L}++$  proposal improves the architectures of the state of the art.

To the author’s knowledge, this is the first proposal for 3D DL analysis of weld seams at the instant of fabrication that achieves very high quality identification of their shape and volume characteristics, as well as almost complete noise elimination.

**Acknowledgements** Acknowledgments to Fernando Buitrago (Department of Theoretical, Atomic and Optical Physics, Universidad de Valladolid) for the material support provided and his help in the development of the project.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The research leading to these results received funding from the program “Subvenciones para la realización de proyectos de I+D+i en el ámbito de Castilla y León cofinanciadas con FEDER” under Grant Agreement No. FUNGE 061-217731.

**Data availability** The datasets generated and analysed during the current study are not publicly available due the fact that they constitute an excerpt of research in progress but are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

## References

- Automotive, H., & Laboratory, I. (2021). Semantic segmentation editor. <https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor>
- Bacioiu, D., Melton, G., Papaalias, M., et al. (2019). Automated defect classification of ss304 tig welding process using visible spectrum camera and machine learning. *NDT & E International*, 107(102), 139. <https://doi.org/10.1016/j.ndteint.2019.102139>
- Balta, H., Velagic, J., Bosschaerts, W., et al. (2018). Fast statistical outlier removal based method for large 3d point clouds of outdoor environments. *IFAC-Papers Online*, 51, 348–353. <https://doi.org/10.1016/j.ifacol.2018.11.566>
- Cai, W., Wang, J., Jiang, P., et al. (2020). Application of sensing techniques and artificial intelligence-based methods to laser welding real-time monitoring: A critical review of recent literature. *Journal of Manufacturing Systems*, 57, 1–18. <https://doi.org/10.1016/j.jmsy.2020.07.021>
- Cai, W., Jiang, P., Shu, L. S., et al. (2021). Real-time laser keyhole welding penetration state monitoring based on adaptive fusion images using convolutional neural networks. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/S10845-021-01848-2>
- Chen, H., Guo, N., Huang, L., et al. (2019). Effects of arc bubble behaviors and characteristics on droplet transfer in underwater wet welding using in-situ imaging method. *Materials and Design*, 170(107), 696. <https://doi.org/10.1016/j.matdes.2019.107696>
- Cheng, Y., Wang, Q., Jiao, W., et al. (2020). Detecting dynamic development of weld pool using machine learning from innovative composite images for adaptive welding. *Journal of Manufacturing Processes*, 56, 908–915. <https://doi.org/10.1016/j.jmapro.2020.04.059>
- Dai, W., Li, D., Tang, D., et al. (2021). Deep learning assisted vision inspection of resistance spot welds. *Journal of Manufacturing Processes*, 62, 262–274. <https://doi.org/10.1016/j.jmapro.2020.12.015>
- He, K., & Li, X. (2016). A quantitative estimation technique for welding quality using local mean decomposition and support vector machine. *Journal of Intelligent Manufacturing*, 27, 525–533. <https://doi.org/10.1007/S10845-014-0885-8>
- He, K., Zhang, X., Ren, S., et al. (2015). Deep residual learning for image recognition. [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
- Huang, G., Liu, Z., Maaten, LVD., et al. (2017) Densely connected convolutional networks. In *Proceedings—30th IEEE conference on computer vision and pattern recognition, CVPR 2017 2017-January*. <https://doi.org/10.1109/CVPR.2017.243>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd international conference on machine learning, ICML 2015* (Vol. 1, pp. 448–456). <https://doi.org/10.48550/arxiv.1502.03167>
- Liu, C., Wang, K., Wang, Y., et al. (2022a). Learning deep multimani-fold structure feature representation for quality prediction with an industrial application. *IEEE Transactions on Industrial Informatics*. <https://doi.org/10.1109/TII.2021.3130411>
- Liu, T., Wang, J., Huang, X., et al. (2022b). 3dsmda-net: An improved 3dcnn with separable structure and multi-dimensional attention for welding status recognition. *Journal of Manufacturing Systems*, 62, 811–822. <https://doi.org/10.1016/j.jmsy.2021.01.017>
- Liu, Y., Yang, C., Huang, K., et al. (2022c). A systematic procurement supply chain optimization technique based on industrial internet of thing and application. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/JIOT.2022.3228736>
- Lu, R., Wei, H., Li, F., et al. (2020). In-situ monitoring of the penetration status of keyhole laser welding by using a support vector machine with interaction time conditioned keyhole behaviors. *Optics and Lasers in Engineering*, 130(106), 099. <https://doi.org/10.1016/j.optlaseng.2020.106099>
- Melakhsou, A. A., & Batton-Hubert, M. (2021). Welding monitoring and defect detection using probability density distribution and functional nonparametric kernel classifier. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/S10845-021-01871-3>
- Miao, R., Shan, Z., Zhou, Q., et al. (2022). Real-time defect identification of narrow overlap welds and application based on convolutional neural networks. *Journal of Manufacturing Systems*, 62, 800–810. <https://doi.org/10.1016/j.jmsy.2021.01.012>
- Minaee, S., Boykov, Y., Porikli, F., et al. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44, 3523–3542. <https://doi.org/10.1109/TPAMI.2021.3059968>
- Naceur, M. B., Akil, M., Saouli, R., et al. (2020). Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Medical Image Analysis*, 63(101), 692. <https://doi.org/10.1016/J.MEDIA.2020.101692>
- Redmon, J., & Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In *Proceedings—30th IEEE conference on computer vision and pattern recognition, CVPR 2017 2017-January* (pp. 6517–6525). <https://doi.org/10.1109/CVPR.2017.690>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 234–241). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Santurkar, S., Tsipras, D., Ilyas, A., et al. (2018). How does batch normalization help optimization? *Advances in Neural Information Processing Systems*, 31, 2483–2493.
- Singh, S. A., & Desai, K. A. (2022). Automated surface defect detection framework using machine vision and convolutional neural networks. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/S10845-021-01878-W>
- Tang, Y. S., Wu, J. L., Yeh, S. S., et al. (1999). Intelligent modelling and optimization of the gas tungsten arc welding process. *Journal of Intelligent Manufacturing*, 10, 73–79. <https://doi.org/10.1023/A:1008920631259>
- Wang, B., Hu, S. J., Sun, L., et al. (2020a). Intelligent welding system technologies: State-of-the-art review and perspectives. *Journal of Manufacturing Systems*, 56, 373–391. <https://doi.org/10.1016/j.jmsy.2020.06.020>
- Wang, Q., Jiao, W., & Zhang, Y. (2020b). Deep learning-empowered digital twin for visualized weld joint growth monitoring and penetration control. *Journal of Manufacturing Systems*, 57, 429–439. <https://doi.org/10.1016/j.jmsy.2020.10.002>
- Wang, Q., & Mei, J. (2022). Shdm-net: Heat map detail guidance with image matting for industrial weld semantic segmentation network. [arXiv:2207.04297](https://arxiv.org/abs/2207.04297)
- Wang, X., Chen, T., Wang, Y., et al. (2022). The 3d narrow butt weld seam detection system based on the binocular consistency correction. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/S10845-022-01927-Y>
- Xia, C., Pan, Z., Polden, J., et al. (2020). A review on wire arc additive manufacturing: Monitoring, control and a framework of automated system. *Journal of Manufacturing Systems*, 57, 31–45. <https://doi.org/10.1016/j.jmsy.2020.08.008>
- Xiao, M., Yang, B., Wang, S., et al. (2022). Research on recognition methods of spot-welding surface appearances based on transfer learning and a lightweight high-precision convolutional neural network. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/S10845-022-01909-0>
- Yang, Y., Pan, L., Ma, J., et al. (2020). A high-performance deep learning algorithm for the automated optical inspection of laser welding. *Applied Sciences*. <https://doi.org/10.3390/app10030933>

Zhou, Z., Siddiquee, MMR., Tajbakhsh, N., et al. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 3–11). [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)

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