

Accepted Article Preview: Published ahead of advance online publication



First Steps towards Machine Learning for Prediction and Pre-Correction in Direct Laser Writing

Sven Enns, Julian Hering-Stratemeier and Georg von Freymann

Cite this article as: Sven Enns, Julian Hering-Stratemeier, Georg von Freymann. First Steps towards Machine Learning for Prediction and Pre-Correction in Direct Laser Writing. *Light: Advanced Manufacturing* accepted article preview 30 July, 2026; doi: 10.37188/lam.2026.128

This is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication. LAM are providing this early version of the manuscript as a service to our customers. The manuscript will undergo copyediting, typesetting and a proof review before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Received 28 November 2025; revised 28 July 2026; accepted 28 July 2026;
Accepted article preview online 30 July 2026

First Steps towards Machine Learning for Prediction and Pre-Correction in Direct Laser Writing

Sven Enns^{1 4 *}, Julian Hering-Stratemeier^{1 2}, and Georg von Freymann^{1 2 3 4}

¹*Physics Department and Research Center OPTIMAS, RPTU University Kaiserslautern-Landau, Erwin-Schrödinger-Straße 56, 67663 Kaiserslautern, Germany*

²*Opti-Cal GmbH, Erwin-Schrödinger-Straße 56, 67663 Kaiserslautern, Germany*

³*Fraunhofer Institute for Industrial Mathematics ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany*

⁴*Institute of Photonic Systems and Technologies, Leibniz University Hannover, 30167 Hannover, Germany*

*Correspondence to: Sven Enns: sven.enns@qtp.uni-hannover.de

Abstract

Additive manufacturing using 2-Photon Polymerization (2PP, aka direct laser writing DLW) enables the fabrication of almost arbitrary complex 3D structures from the meso to the sub-micron scale. However, deviations between the anticipated target structure and the actual print often occur due to physico-chemical processes, limiting the accuracy and reliability of this technology. To minimize these deviations, we hereby present our latest research in developing different neural networks, targeting the above-mentioned aspect. Our networks are trained on several experimental as well as theoretical datasets and show good results in predicting fabrication deviations and (pre-) correcting 2.5D μ -structures. Hence, we demonstrate, that besides conventional iterative correction methods, neural networks are a promising alternative to significantly improving the output quality in DLW. First directly pre-corrected structures show a reduction of the root mean square error up to 93% in the best case scenario. Furthermore, there are no fundamental limitations to transferring this machine learning approach to other 3D printing technologies, as they all face the same challenge in terms of fidelity. To our point of view, the use of neural networks has the potential to enhance the capabilities of this technology, enabling the creation of complex structures with increased accuracy and precision in the near future.

Keywords: 2-Photon Polymerization (2PP), Direct Laser Writing (DLW), neural networks, machine learning, artificial intelligence structure prediction, structure (pre-) compensation/correction

1 Introduction

One of the most high speed and high resolution variants of laser-based additive manufacturing is the 2-Photon Polymerization (2PP) technology, also known as direct laser writing (DLW) first introduced by Maruo *et al.* in 1997 [1]. There, photo sensitive materials, so-called photo resins, are locally exposed by a (tightly) focused femtosecond laser beam and subsequently cured. Depending on the laser power using an objective with high NA ($NA \approx 1.4$), feature sizes of less than 100 nm can be realized [2, 3]. On the other hand, the production of macroscopic components within the cm-regime, such as screws and nuts is also possible with this technology [4]. However, a major challenge in this field is the – mostly undesired – deviation between the envisaged target structure and its corresponding, experimentally 3D

printed counterpart. While these deviations are negligible for large structures (e.g. those in the cm-range), they become more significant as the size of the prints decreases. Unfortunately, this is fundamentally unavoidable with today's photo resins. Although the principles of DLW have been known for more than 25 years, a holistic calculation of these deviations, based on the physico-chemical processes involved, has been just as impossible as a correct prediction of the complete 3D printed structure. Nevertheless, rather simple straight forward calculations of the widths of single lines [5] or voxel and pillar dimensions [6] have been published in 2013 and 2019, respectively. Taking more physico-chemical processes correctly into account, however, already prevents the prediction of line heights or aspect ratios, as stated in 2022 by Pingali *et al.* [7].



A few more approaches focused really successfully on the estimation of printed line dimensions and structure characterization [8, 9], but, unfortunately still not taking into account all of the most relevant physico-chemical processes. The complexity and computational intensity of the processes to be modeled still exceed the capacities of today's computers. In contrast to such holistic simulation approaches, some research work exists on modeling and investigating partial aspects, such as the molecular diffusion [10], the so-called Schwarzschild- [11] and proximity-effects [10] or the shrinkage behavior [12]. Recently, some of the authors published a new algorithm, covering all the aforementioned physico-chemical processes in a very simplified, approximated way [13]. In principle, this approach allows the prediction of arbitrary 2.5D structures, based on seven physical parameters. Here, the term 2.5D structure refers to those structures that can be represented mathematically by a single 2D matrix M in which each index M_{ij} represents the height of the structure $z_{ij} = z(x, y)$ at the respective lateral coordinate (x, y) . Although the above mentioned prediction approach allows for a pre-correction for the experimentally expected structural deviations in DLW, the approach still depends on calibration structures and clearly has its limitations (see reference [13] for details). Due to these limitations and the general complexity of the physico-chemical processes and their dependence of many external factors, machine learning (ML) approaches could be a solution. Numerous studies have investigated the application of these approaches in additive manufacturing (AM) in general. Thereby, they are used in multiple stages of the AM workflow, including material design – to optimize geometries, topologies, and predict material properties – process optimization during fabrication, and post-processing tasks such as defect detection, quality control, and data analysis. A broad spectrum of ML techniques has been implemented for various fabrication processes [14, 15]. To further enhance the capabilities of ML, emerging approaches such as physics-informed machine learning (PIML) incorporate physical knowledge about the system into model architectures and loss functions, thereby improving model interpretability and predictive performance, for example as a counter part to resource demanding physical simulations [16]. Nevertheless, ML approaches need to be adapted for all the different applications. In the context of DLW, only a limited number of studies have explored the use of ML. Similar to other AM techniques, ML can support various stages of the DLW process. For instance, one study applied ML to calibrate data from a DLW-fabricated force sensor during post-processing, amongst other things also compensating for fabrication deviations [17]. Only very first publications point roughly into the direction of controlling printing design or DLW parameters to generate better printing results: Lee *et al.* developed an ML model for the automated

detection of quality characteristics during DLW printing [18]. Here, the model was trained to distinguish between 'cured', 'uncured', 'damaged', and 'illuminated' during the fabrication process, allowing for a practical coarse in-time light dosage adjustment. However, the deviation between anticipated structure and real 3D print can not be minimized by this approach, since accurate dimensions can not be identified. Moreover, the deforming processes (i.e. shrinkage) are not completed while fabricating. In a similar manner, Mourka *et al.* presented in a conference talk a ML model to predict the optimal exposure parameters for 2PP, based on scanning electron images [19]. Unfortunately, no paper followed yet. To the best of our knowledge, only Yang *et al.* tried to use ML to improve the geometric compliance in DLW [20]. They show an approach to find and correct for systematic deviations in simple structures, such as lenses and lines, by an ML framework. Although the average errors could be reduced by 50–80 %, the study used data-driven compensation algorithm for the structure correction and lacks of transferability to complex or even arbitrary DLW geometries.

To the best of our knowledge, no further publications have addressed our ultimate goal: providing a neural network with a target design that, in turn, generates a pre-compensated version of it, resulting in a 3D print that reproduces the original geometry as accurately as possible. Therefore, we report on the ongoing improvement in this context and introduce a very first machine learning based approach for direct structure prediction and pre-compensation in DLW. We developed, trained and tested various neural networks and demonstrate our most promising attempts so far.

In this work, we only refer to DLW as polymerization by two- or multi-photon absorption.

2 Materials & Methods

This section first describes the printing and measurement steps in detail. It then explains how the measurement data is processed, followed by the development of our neural network architectures. Finally, we briefly discuss the chosen loss function of the networks, which is a measure for the accuracy of the network and can be used to prioritize certain features during the training while ensuring error convergence of the network training.

2.1 Fabricating and Measuring

Within this study, all structures shown have been designed as 2D surface matrices M_{ij} . Converting the 2D surface matrices into the standard 3D printer data type '.stl' allows for a common translation into coordinates for the DLW machine: If not explicitly described otherwise, we have used the software DeScribe from Nanoscribe to discretize our structures into equidistant axial planes with a so-called slicing distance of $0.1\mu\text{m}$ and each plane into lateral lines along the x -axis separated by $0.1\mu\text{m}$ – called hatching distance. The thus generated data is interpreted

by the associated 3D printer Photonic Professional GT. Galvanometric mirrors have been used for lateral, and a piezo stage has been used for axial positioning, as well as a constant writing speed of 20 mm/s. The laser power has been kept constant at 45 mW whereas the beam always illuminates the complete entrance pupil of a 63x objective with a numerical aperture of 1.4 (Carl Zeiss Microscopy Deutschland GmbH). IP-S has been used as photo resin. The latter, the DLW machine, as well as the slicing software are products of Nanoscribe GmbH & Co. KG. The development steps after printing onto ultra-sonic cleaned and subsequently silanized glass substrates [21] followed the manufacturer's specifications [22]: first, resting in propylene glycol methyl ether acetate (PGMEA) for 20 minutes, afterwards resting in isopropanol for five minutes, and finally drying gently with nitrogen.

For measuring the structure's topography, a μ Surf confocal microscope (NanoFocus AG) equipped with a 100x objective (NA = 0.95) and a 60x objective (NA = 0.9) (Olympus Europa SE & Co. KG, both) have been used. To obtain the best possible results, the microscope's detection parameters,

such as exposure time and gain, have been optimized for each single measurement.

2.2 Data Preparation

The fabrication and measurement techniques described before are used to generate training and test data for our neural networks (see Sec. 2.3). Since the network training with 3D structures needs much more computational resources and is more time consuming and furthermore a suitable measurement technique to generate accurate training data is not available, different 2.5D structure designs are chosen to build a training data-pool from scratch. Besides including, e.g., simple blocks and step structures into the training data-pool, one focus here is on the fabrication of calibration geometries, such as AIR [23] or radially symmetric chirped CIN_r structures [24] according to ISO 25178–70 [25], where a sophisticated structural conformity is crucial, as they have been proven to be fabricable with DLW [26]. A full list of fabricated structures including exemplary illustrations is shown in the appendix 5.2. By measuring the height of the final structures, all processes that can lead to deviations are completed. Hence,

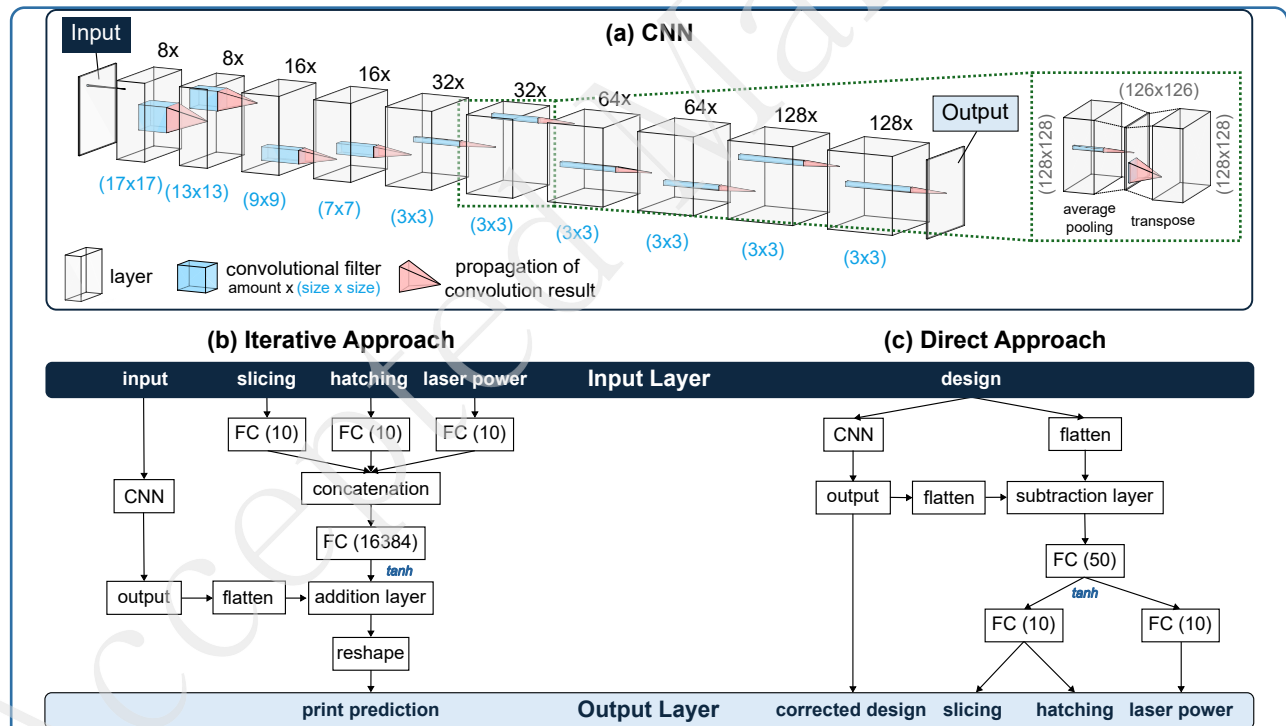


Fig. 1. Schematic visualization of the network architectures. (a) General architecture of the neural networks with only convolutional layers. Optionally, e.g., pooling layers were implemented as layer substitution by a convolutional, an average pooling, and a deconvolutional layer, as shown in the green inset. (b) **Iterative Approach** network for structure prediction including the three DLW parameters slicing, hatching, and laser power by a second branch. Therefore, also fully connected layers (FC) are used. Since the FC-outcome can be negative, a hyperbolic tangent is used as activation function, whereas all other layers still use rectified linear unit (ReLU) activation. (c) **Direct Approach** architecture, designed to correct for the printing deviations and predicting a corrected design with corresponding writing parameters. The field 'CNN' represents the network shown in (a).

the deviation mechanisms are fully included into the training data of the network. Since the fabrication and measurement is time consuming and a large data-pool is needed in order to perform a successful training of the network, additional data is created using data augmentation. Data augmentation is an often used technique to increase the amount of training data by creating new data on basis of already existing datasets [27]. In our work, data augmentation can be used by rotating or mirroring existing non-symmetric datasets. However, this process is limited not only to the degree of symmetry. Since the structure is additively fabricated by printing single lines, a writing direction is introduced which can lead to an anisotropy that has to be taken into account while using data augmentation. Hence, data augmentation by rotation is limited to the rotation of 180 degrees.

Starting from the measurement data obtained by confocal microscopy, the following data processing steps are performed, based on the steps by Eifler *et al.* [28]: First, the correct height of the structure has to be determined. Therefore, the derivation of the Abbott-Firestone curve [29] is calculated. Due to the height difference between substrate and structure, it always shows one very prominent jump, allowing for the anticipated height correction. Further, non-measured points are interpolated, methodically based on the type of structure, e.g., nearest neighbour for step height structures and linear for curved structures. Second, tilt and rotation errors due to positioning in the printer and microscope have to be corrected. Since the base size of the used structures is square, the rotation can be corrected by aligning the edges of the structures. Correcting the tilt is more complex, as a simple tilt correction of the microscope data based on the measured tilt of the substrate does not take into account the tilt of the substrate during the writing process. In order to correct for all tilting errors as precisely as possible, the measurement is compared to the designed structure by calculating a deviation matrix, representing the height difference between the design and the height corrected measurement data. For this difference matrix a regression plane can be calculated and used to finally correct the tilt within the measurement data. A schematic illustration of the described steps is shown in the appendix in Figure 12.

As normalization is used in neural networks to enhance numerical stability and convergence of the training, the height information is normalized to values between 0 and 1 given the largest structure in the training data. After the training is completed, the result of the network is multiplied by the normalization factor to regain absolute height information. The images are scaled to a size of 128×128 pixels. Given the structure base size of $(50 \times 50) \mu\text{m}^2$ and the confocal microscope's lateral resolution of $0.3125 \mu\text{m}$, this leads to a resolution of $0.391 \mu\text{m}$ for further treatment. Image sizes of 128×128 pixels are chosen as a trade-off to keep the computation time as low as

possible without losing too much information by reducing the resolution. Thereby an axial resolution of down to 1.5 nm [30] can be achieved. The combination of a structure design and its corresponding processed measurement result is a single dataset for our neural networks, where either the design or the measured print can be used as either input or label. Based on these procedures and tools, a data-pool of in total 5217 labeled datasets is generated. Different subsets of this data-pool are used depending on the trained network. The first subset, used to test the general suitability of the networks, consists of 628 datasets. Thereby, the size of the structures used to train the network is $50 \mu\text{m}$ by $50 \mu\text{m}$ with a height between $1 \mu\text{m}$ and $25 \mu\text{m}$. The base size is kept constant since the image size of our network has to be constant. A scaling of different structure sizes to the same image size would also scale and, hence, stretch or compress the occurring deviations, leading to a higher inaccuracy of the network. Later on, more structures and varying printing parameters are introduced to increase the data-pool. To get a consistent result, visibly, unusually distorted or defect structures were filtered and not used for the training, while measurement artefacts especially at the edges were kept for this study. The subsets used for the training are again split into so-called training (85%) and test data (15%). These are used to check the network's performance and to recognize possible over- or under-fitting. To compare the results of different networks and training realizations, a set of structures is defined as benchmark. In the result section we use blocks, CIN_r - and AIR -type structures, and a logo of the former University of Kaiserslautern to visualize different aspects of the network results. Thereby, blocks, CIN_r - and AIR -type are structures included into the training but with not trained parameters. The University's logo is only included into the test data, hence, allows for analyzing the performance of the network on not trained structures. To estimate a reachable accuracy of the network prediction, the repetition accuracy of the data generation is tested by multiple executions for different structure types, leading to a deviation of up to 300 nm . Since not every structure can be investigated, the accuracy is estimated with a puffer to be about 500 nm . This accuracy can also be regarded as the best possible accuracy that can be expected from the neural network.

2.3 Network Architecture

In the following, two different approaches are presented. The first approach aims at the target-print-deviation: integrating the design into a neural network, which then predicts the printed outcome. Having successfully learned this deviation, the network prediction can then be used to perform a design correction. Therefore the following correction mechanism is used analogous to Lang *et al.* [13]:

$$M_k = M_{\text{Design}} - \alpha \cdot ((M_{\text{Prediction}} - M_{\text{Design}}) * g_{\text{corr}}), \quad (1)$$

$$M_{k+1} = M_k - \alpha \cdot ((M_{\text{Prediction},k} - M_{\text{Design}}) * g_{\text{corr}}), \quad (2)$$

with M as surface matrix, k as iteration numerator, g_{corr} denoting a two-dimensional Gaussian distribution, and α as correction strength factor. Since this step can be performed multiple times we call it the **Iterative Correction (IC)** approach from now on. The second approach is called **Direct Correction (DC)**, since the goal is a direct prediction of a corrected design. Technically speaking, input and output are switched in comparison to the **IC**. In this case, the measured prints of the fabricated training datasets are given as input and the original designs as output. If now an anticipated design is given as input, the network will predict a corrected version of the design to get a printed structure as close as possible to the input design.

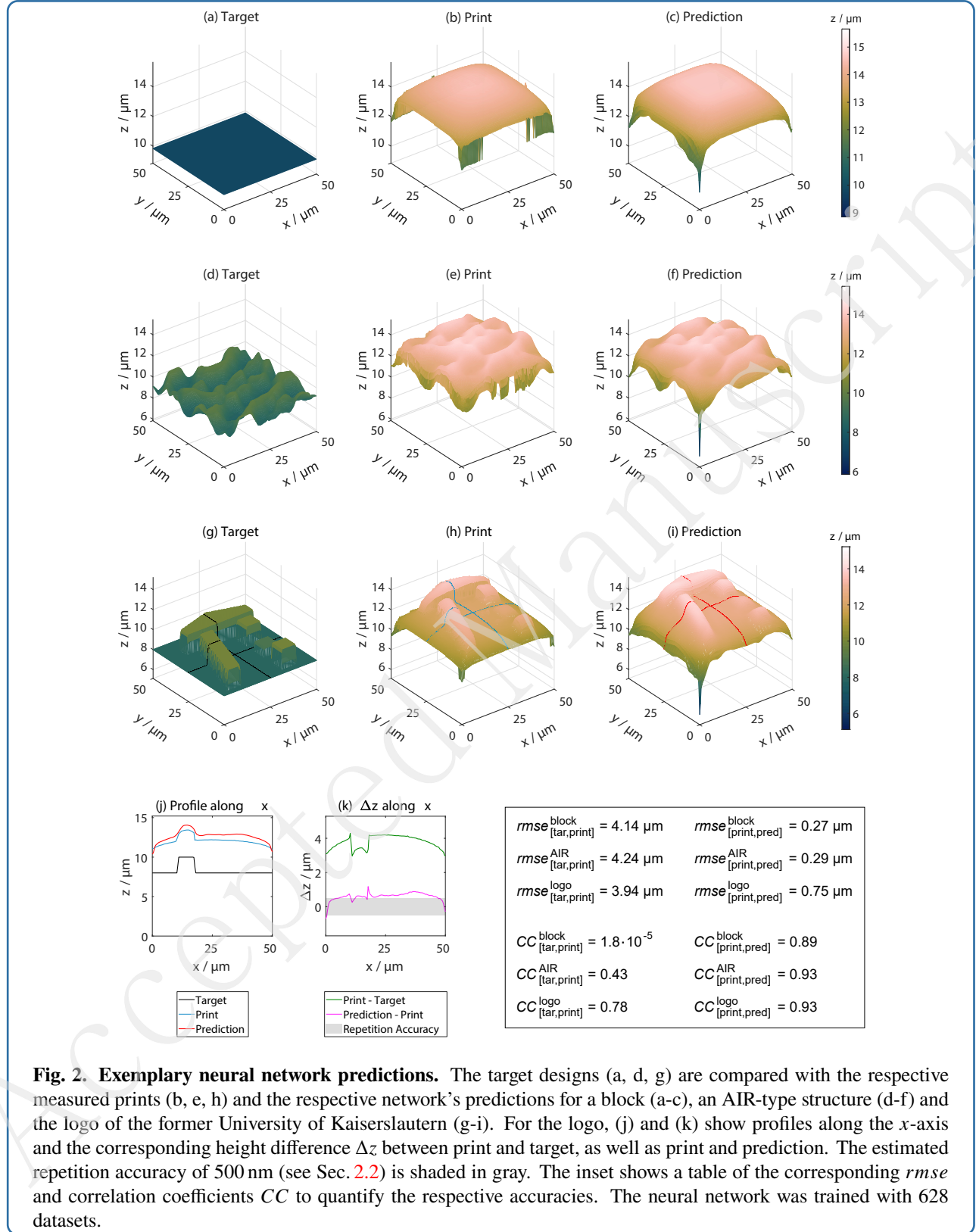
In order to analyze image data with neural networks in terms of image classification, object detection and so on, the most popular approach is the application of convolutional neural networks (CNNs) which will also be used in this work. We use supervised learning, where every input is labeled. In our case, the labeling is done by connecting the datasets of a design with the measurement of its printed result. Input and label can be switched as explained for the two different approaches. One of the crucial parts in using neural networks is to find a suitable network architecture for the problem at hand. To identify a starting point, we use the well-known CNN architectures AlexNet [31] as well as VGG16 and VGG19 [32] as inspiration. Hence, we present three different network architectures depending on used parameters and tasks. The first logical task is to predict the target-print-deviation. Based on this prediction, one can perform the above-mentioned **Iterative Correction** to get as close as possible to the target print. The task of the second architecture is the **Direct Correction** of a corrected design. In both cases, only different designs were initially used for training, while the DLW parameters remained constant for the beginning. Figure 1 (a) shows the basic network architecture for the training without DLW parameters consisting only of convolutional layers. Decreasing the filter sizes from 17×17 to 3×3 with a stride of 1 while increasing the layer numbers from 8 to 128, allows to capture larger as well as more detailed features. As visualized, input and output of the CNN network have to be of same size. In this case, this is true for every single layer by using zero-padding. However, this can lead to boundary effects. Therefore, we add average-pooling layers in between convolutional layers to reduce the strength of these effects. Since zero padding would lead to a huge average difference on the edges for every pooling layer, no padding is used, leading to a reduction of the output size of this layer. To regain the initial size, deconvolutional layers, also called transpose layers, are implemented as represented in the inset in Fig. 1 (a). While the network predictions are similar, especially when comparing the loss function values, there are characteristics for each network, as exemplary shown in the appendix 5.3 for the mentioned boundary effects. For all

layers we choose the rectified linear unit (ReLU) as activation function. All parameters, such as filter size or stride, are chosen empirically to find the best possible combination. The shown architecture can be used for both, the **Iterative Correction** and the **Direct Correction** by switching the input and output. Since the fabrication parameters have a strong influence on the printing result, the network has to be adjusted in order to capture these effects. In the following, three crucial parameters – namely slicing distance, hatching distance and laser power – are taken into account. In general, not only the laser power but the exposure dose is a crucial parameter. Thereby, the exposure dose is determined by laser power, pulse duration, repetition rate and scan speed [33]. While for future models the exposure dose could be one of the network parameters, the present study focuses on the laser power to control the light intensity. Since we want to print as fast as possible for maximum efficiency, the scan speed is kept constant. Thereby, changing the laser power in a certain range leads approximately only to a change of the average structure height, changing hatching and slicing distances on the other hand lead to rather strong and non-symmetric changes. These stem especially from the direction dependent hatching. Further anisotropies can for example be introduced by the polarization of the light [34, 35, 36]. However, the effects are shown to be on the scale of nanometers, which is much smaller than the accuracy of the presented model and can hence be neglected at this stage of development. For this reason, circular polarization is chosen for symmetry reason, as is the standard configuration of the commercial system.

The updated network architectures to include the three parameters are shown in Fig. 1 (b) and (c). While for the deviation prediction the parameter information is included in the input layer (b), the **DC** (c) delivers a corrected design and the corresponding DLW parameters in the output layer. In both cases, the 'CNN-layers' represent the network architecture from (a). The underlying idea behind these architectures is to create a second branch to process the DLW parameters. For the prediction, the parameters are converted to a matrix with the size of the input image, representing an parameter dependent height offset, which is then added to the initial CNN prediction. Since the offset can contain negative values, a hyperbolic tangent is used as activation function for the fully connected layer before the addition step. In the case of **DC**, the approach is similar, except that this time the difference between the initial CNN prediction and the target design is calculated, and the appropriate DLW parameters are determined from this.

2.4 Loss Function

So-called loss functions are defined to evaluate the train and test performances of neural networks and allow for the comparison between different architectures. Furthermore, choosing a loss function is essential for the convergence of the network results and has an effect on which features are



more important along the training process. As illustrated in Appendix 5.1 using the MSE as an example, not every loss

function is suitable for the problem at hand. Therefore, we introduce a customized loss function (see Eq. 3) consisting of the following terms:

$$L = 1 - CC + \frac{|\bar{M}_{\text{pred}} - \bar{M}_{\text{lab}}| + \sum_{i,j=1}^n |M_{\text{pred}}^{i,j} - M_{\text{lab}}^{i,j}|}{1.5 \cdot \bar{M}_{\text{lab}}} \quad (3)$$

where CC is the Pearson correlation coefficient [37]:

$$CC = \frac{\sum_{i,j=1}^n (M_{\text{pred}}^{i,j} - \bar{M}_{\text{pred}}) \cdot (M_{\text{lab}}^{i,j} - \bar{M}_{\text{lab}})}{\sqrt{\sum_{i,j=1}^n (M_{\text{pred}}^{i,j} - \bar{M}_{\text{pred}})^2 \cdot \sum_{i,j=1}^n (M_{\text{lab}}^{i,j} - \bar{M}_{\text{lab}})^2}} \quad (4)$$

The correlation coefficient CC can take values between 1 for perfect correlation and -1 for perfect anti-correlation, and is implemented in such a way that even small surface features are given greater importance during the training process. Since the loss function will be minimized during the training process, we define the first part of the loss function as $L_1 = 1 - CC$ so that $L_1 \rightarrow 0^+$ in case of perfect correlation between prediction and label, and $L_1 \rightarrow 2$ for anti-correlation. Second, the absolute value of the difference between the mean height value of prediction and label is calculated and added so that the average height of the predicted structure matches the label. As last term of the loss function every difference between corresponding surface matrix elements of label and prediction are added. The latter two terms are then divided by the mean height of the label to keep the loss function dimensionless.

Finally, an empirically determined scaling factor of 1.5 is added to the denominator. As the goal of the network training is the minimization of the prediction error, several optimization methods exist to find the global minimum of the loss function. Thereby, stochastic gradient descent (SGD) methods are very common for CNNs. Several variations of the SGD have been developed in order to decrease convergence time and the probability of getting trapped in a local instead of the global minimum. A test of seven often used methods is shown in [38]. One way to increase convergence speed is the modification of the gradient for each step based on previously calculated gradients. Exemplary, the optimizer Adam (adaptive moment estimation) updates the parameters based on estimated first and second moments, taking into account

a moving average of previous gradients with exponentially decaying rates [39, 40]. We chose the optimizer algorithm Nadam for the training process, which is similar to Adam but includes the Nesterov moment, describing the calculation of the gradient not at the current parameter set, but one step ahead according to the currently accumulated moment to get a better correction for the next iteration step [41, 42, 43]. While the suggested default value for the learning rate, defining the step size for the gradient descent of the optimizer, is 0.001 for Adam and Nadam [39, 43], we implement a smaller learning rate of 0.0001, as it results in better convergence in our case. Thereby a large step size can lead to overshooting and a small step size increases the computation time. Resulting training duration and used hardware and programming libraries are also shown in appendix 5.5.

3 Results & Discussion

In this chapter we present some of the results that we obtained for the defined benchmark structures, giving an overview over achieved accuracy and prominent occurring effects. The first networks (Fig. 1) are designed to predict the occurring deviations for different structure types. Moreover, they can be used to perform the before mentioned **Iterative Correction** approach: three examples using a network, trained with 628 datasets without variation of the DLW parameters are shown in Fig. 2. As first example, the measurement of a printed block is compared to the corresponding network prediction (a-c). The height difference between target and print, as well as the bulging in the middle and drop off towards the edges of the structures are predicted well, as the correlation coefficient $CC_{[\text{print}, \text{pred}]}^{\text{block}} = 0.89$ and root mean squared error $rmse_{[\text{print}, \text{pred}]}^{\text{block}} = 0.27 \mu\text{m}$ (text-inset in Fig. 2) confirm. To evaluate structures, where shrinkage, proximity effect and steep edges are more present, Fig. 2 shows additionally an AIR-type structure (d-f) and the logo of the former University of Kaiserslautern (g-i). Thereby, the latter structure was not included in the training data. The correlation coefficients are consistently above 0.9 and underline the good agreement of the prediction with the actual 3D print. Since the University's logo is the only of those three structures which is not included in the training data, the highest $rmse$ -value of $0.79 \mu\text{m}$ can be observed

Table 1: **Quantities of the neural network's performance.** Comparison between target structure, printed structure, the network's prediction of a corrected print, and a corrected print for a block, CIN_r-type, and logo (illustrated in Fig. 3). The comparison is quantified by Pearson correlation coefficient CC (see Eq. 4) and root mean square error $rmse$.

	block		CIN _r		logo	
	$rmse / \mu\text{m}$	CC	$rmse / \mu\text{m}$	CC	$rmse / \mu\text{m}$	CC
[target, print]	4.14	1.8e-05	3.67	0.27	3.49	0.78
[target, predicted correction]	0.24	-4.6e-06	0.59	0.80	0.77	0.86
[target, printed correction]	0.27	-3.1e-05	0.58	0.88	0.90	0.87

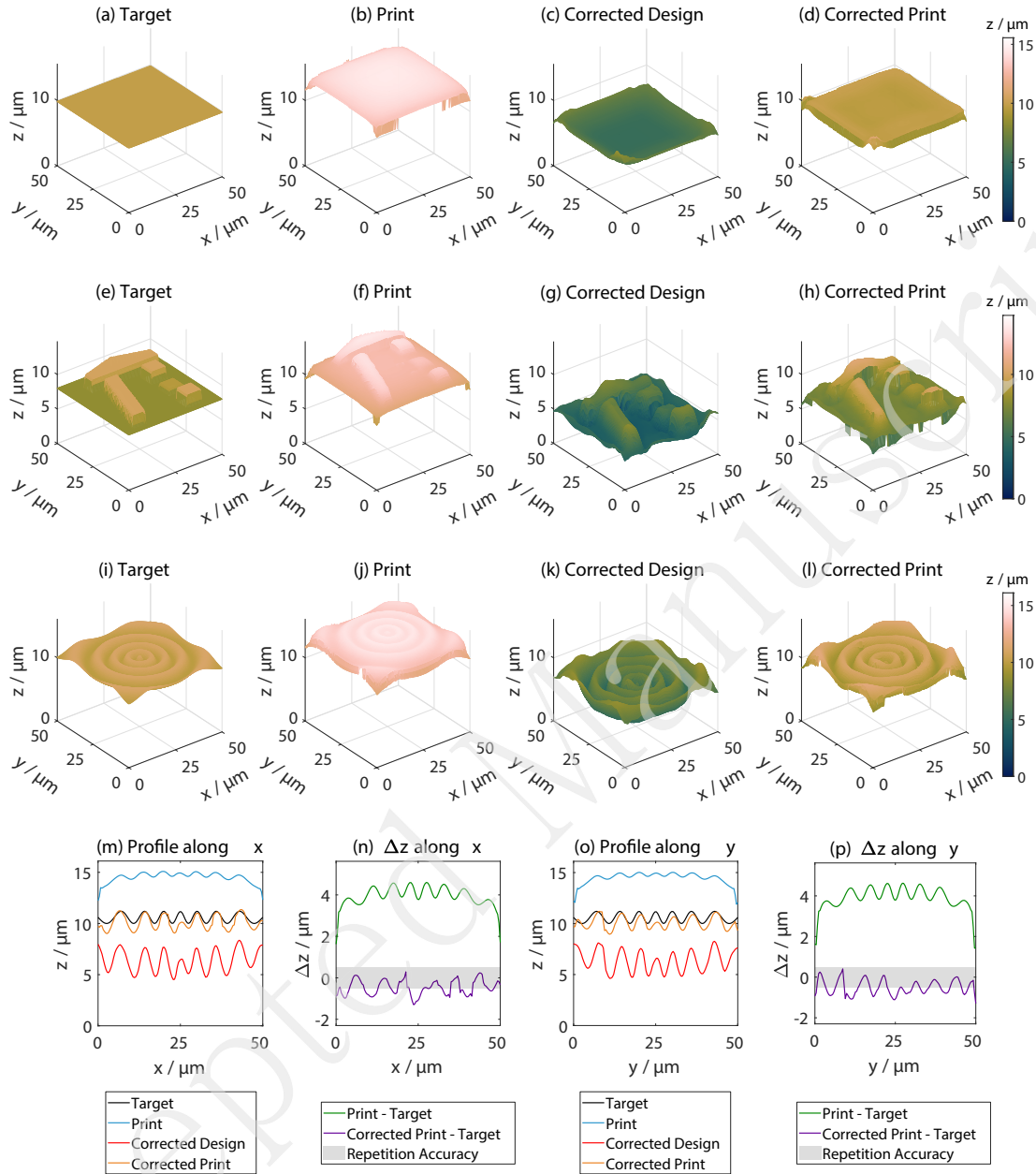
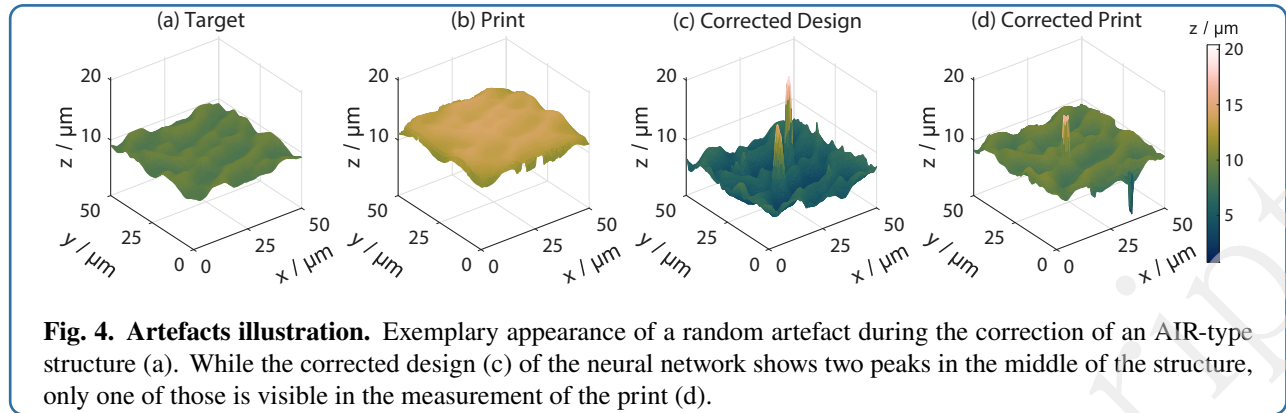


Fig. 3. Results of the Direct Correction (DC) approach. Examples for the direct design correction of a network trained with 628 datasets. Shown are the corrections of a block (a)-(d), the logo (e)-(h), which was not included in the training data, and a CIN_r -type structure (i)-(l) including profile information (m)-(p).

here. However, the profile plots in (j-k) show an almost constant offset between print and network's prediction, being only slightly out of the estimated range of repetition accuracy. Highest deviations can always be seen in areas of steep edges.

While the shown results prove the general suitability of neural networks to predict the printing outcome, a further goal is to pre-correct the target design, hence, being

able to print as close to the target structure as possible. Therefore, the aforementioned methods of **Direct** and **Iterative Correction (DC, IC, respectively)** are applied. The DC-approach is the quicker and simpler method, since the input of the target design directly results in a corrected design. Again, the network's performance is tested by three benchmark structures: a block, a circular chirped CIN_r -type structure – which are both part of the training data



but with untrained parameters – and the logo. The results are visualized in Fig. 3. While the improvement is already clearly visible, Tab. 1 quantifies once again the CC and $rmse$ values. Thereby, the corrected print of the block reaches an $rmse$ similar to the prediction accuracy of the network, shown before. For the CIN_r-type structure, especially the increase of the correlation coefficient from 0.27 between target and uncorrected print to 0.88 between target and corrected print, as well as the $rmse$ -reduction by about 84 % is remarkable. As the profile plots in Fig. 3 (m-p) show, the difference between target and corrected print is partly in the range of repetition accuracy. However, the frequency and especially the amplitude of the CIN_r prints' surface features do still not match the target design sufficiently. Furthermore, the correction does not necessarily preserve the symmetry of the structure. In particular, regarding the pre-correction of such calibration structures, this is a disadvantage in comparison to the algorithmic approach presented by Lang *et al.* [13]. However, also for the non-trained logo, the correction reduces the initial $rmse$ value from 3.67 μm to 0.9 μm while increasing the correlation from 0.78 to 0.87. As visible in the surface plots, the main deviations are caused by artefacts at the edges and the less steep flanks in comparison to the target. Furthermore, the flat surface areas of the target design stay bulged in the corrected print. The combination of bulging and edge artefacts can lead to locally strong deformations. While an overall improvement is clearly visible and quantitatively confirmed, the feature deformation still can cause problems, depending on the application of the printed structure.

During several tests and training processes, the appearance of random artefacts was noticeable, although not quantifiable. One example is shown in Fig. 4, where a random peak in the middle of the structure appeared. While the rest of the corrected print shows good agreement with the target, this random artefact leads to increased error values and can render the structure unusable.

In contrast to the **Direct Correction**, the **Iterative Correction**, using the deviation prediction according to Eq. 1 and 2, is illustrated in Fig. 5. Again, based on the examples

of an AIR-type structure (a-g) and the logo structure (h-n). Since for this IC-approach only the structure's design and the DLW parameters are needed, all correction iterations can be prepared at once and printed within one single job, saving time and ensure same printing conditions throughout all iteration steps for good comparability. As shown in (g) for the AIR-type structure and in (n) for the logo structure, the network's prediction of the corrected designs show convergent behavior towards higher correlation and lower $rmse$ values for ongoing iteration steps, just as expected. However, the measurement data show a different behavior. While in both cases, a minimum of the $rmse$ is obtained at the second iteration step, the correlation coefficient undergoes a steady decrease. This effect and the again increasing $rmse$ after the second iteration are both results of artefacts. These artefacts get more dominant for every iteration as visualized in appendix 5.4, showing the corrected prints of the AIR-type structure for the first six iteration steps. The artefacts mainly occur at the edges of the structure and at steeper flanks, as has already been identified as problematic for the DC-approach, before. Since the artefacts cannot be properly mapped by the network, they can be enhanced by every iteration step. While in the shown cases the best corrected print is achieved for the second iteration, this does not have to be the case for every structure. Hence, using the IC-approach still demands for the printing of several corrected structures to find the best possible result. Anyhow, since these iterations do not rely on the measurements of each other, they can be printed all in one single job, making this approach very efficient. Fig. 5 presents the second iteration for both structures. A good overall agreement between the target and the corrected print is obvious with only few artefacts. As the profile plots reveal, the difference lies almost completely in the repetition accuracy range. A comparison between the uncorrected print, the DC-approach, and the best result of the IC-approach for the CIN_r-type and AIR-type structures is shown in Tab. 2. Here, two different error characteristics are evaluated. First, the average height difference relative to the design structure is calculated. Second, to further

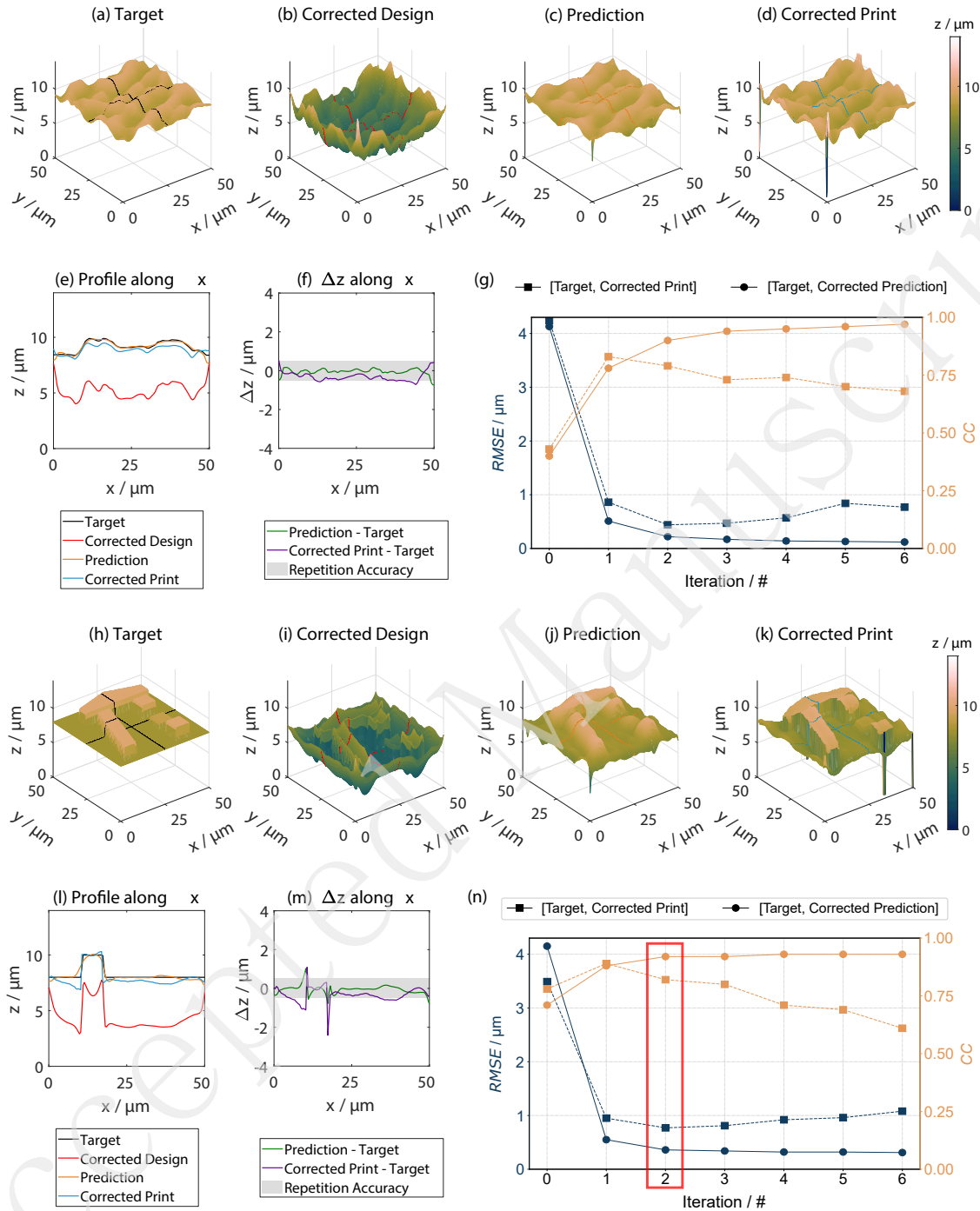


Fig. 5. Iterative Correction (IC). The best result of an AIR-type structure (a-f) is reached after two iteration steps during the IC-approach. In (g), the *rmse* and *CC* values are plotted against the iterations. Similarly, (h)-(n) show the results for the logo structure. The uncorrected result is denoted as iteration 0, respectively. The neural network was trained with 628 datasets.

analyse the quality of the surface features additionally to the correlation coefficient, the mean height is subtracted for each structure, leaving the surface features. The *rmse* is then

calculated accordingly. Finally, metrological parameters, such as the axial amplification coefficient α_z , and the arithmetic and quadratic surface roughness S_a and S_q are

Table 2: **Quantitative comparison of Iterative (IC) and Direct Correction approach (DC).** Both correction approaches are quantified by mean height difference $\Delta\bar{M}$, height corrected root mean square error $rmse$, Pearson correlation coefficient CC (of the inner 80 % of the surface CC_{80}), arithmetic and squared surface roughness parameters S_a and S_q and axial amplification coefficient α_z (of the inner 80 % of the surface $\alpha_{z,80}$). The respective better value is written in green.

	AIR				CIN _r			
	target	print	DC	IC	target	print	DC	IC
$\Delta\bar{M} / \mu\text{m}$	0	4.198	-0.775	-0.274	0	3.614	-0.389	-0.367
$rmse / \mu\text{m}$	0	0.605	0.380	0.344	0	0.653	0.433	0.321
CC	1	0.432	0.854	0.793	1	0.266	0.878	0.812
CC_{80}	1	0.774	0.935	0.911	1	0.666	0.911	0.887
$S_a / \mu\text{m}$	0.411	0.472	0.580	0.440	0.374	0.461	0.660	0.381
$S_q / \mu\text{m}$	0.474	0.633	0.694	0.564	0.418	0.625	0.751	0.490
α_z	1	1,229	1.439	1.124	1	1.288	1.779	1.050
$\alpha_{z,80}$	1	0.735	1.376	0.987	1	0.672	1.683	0.931

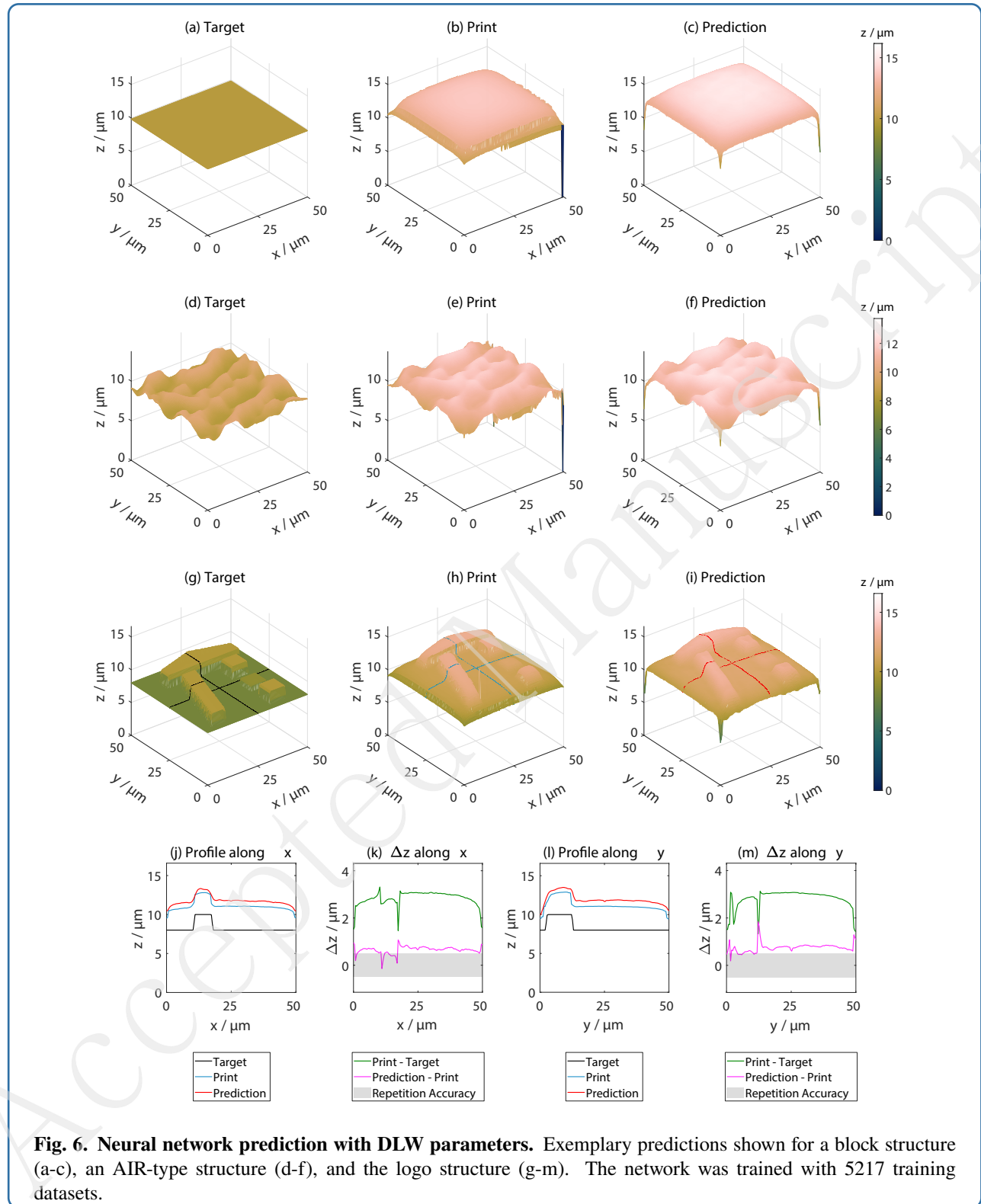
compared. Thereby the index '80' refers to the percentage of the evaluated surface: only the inner 80 % of the surface was evaluated to get rid of edge effects. As the data show, the **Iterative Correction** method delivers the best results regarding the design. Only for the correlation coefficient, the **DC**-approach shows slightly higher values. However, as visualized in Fig. 5 and appendix 5.4, the increasing influence of artefacts in the **IC**-approach is assumed to cause this lower correlation. This assumption can be supported by comparing the correlation coefficient over the complete structure CC and the reduced one CC_{80} . Although the **DC**-approach still shows a higher correlation value, the effect of the edge region is stronger for the **IC**-approach. Since the increasing artefacts appear mostly on edges, this can also be the case in steeper regions in the middle of the structure, additionally lowering the correlation. Noteworthy is however, that despite of the lower correlation, the axial amplification coefficient α_z for the **IC**-approach is closer to the ideal value of 1 and shows additionally the lowest surface roughnesses S_a and S_q , even lower than the non-corrected print. These results show, that the **IC**-approach leads to the best possible results, but needs more effort than the **DC**-approach, since the real error of the corrected structure does not converge with higher iteration numbers.

Since the general functionality and suitability of the neural network has been shown, it now has to be investigated, if the network can also perform well in case of added DLW

parameters. The updated network architecture to include these parameters into the learning and prediction process is shown in Fig. 1 (b) and (c). The respective prediction results are illustrated in Fig. 6 and Tab. 3. While the correlation coefficient between the prints and their associated network predictions show values above 0.94 and are similar to the results without DLW parameters, the $rmse$ tend to be higher. Hence, the general ability of the network to predict the exemplary shown structures is still assured, however, larger deviations can be expected. Thereby, these deviations depend also on the chosen parameter set. Understandably, this can be attributed to the general amount of training datasets and also their distribution among the different parameter values. While the predictions of the network for different DLW parameters show only few artefacts with the largest deviation caused by a difference of the average height, the corrections by the network introduce further effects. This is noticeable by comparing the mean height differences and height corrected $rmse$ s of both approaches in Tab. 3 and Tab. 4. While the mean height difference varies in a range of about 1 μm , the height corrected $rmse$ between the corrected print and the target is about twice as large as between the print and the prediction. As shown in Fig. 7, even for a simple block structure the deviation from the design is much larger than for all previously shown approaches. For all corrected prints, both, the image illustration and the roughness values show an

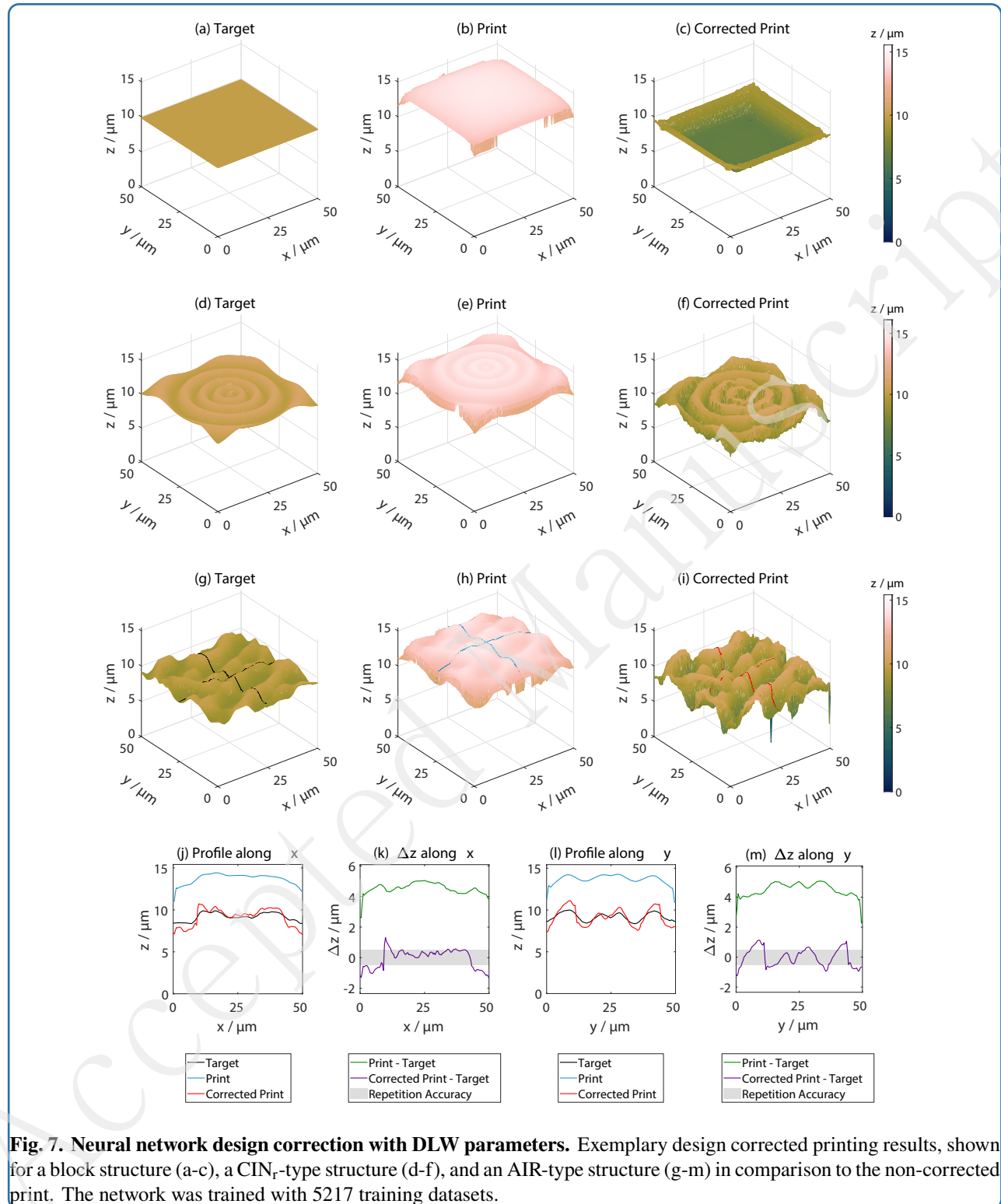
Table 3: **Quantities of the neural network's performance with DLW parameters.** Comparison of target structure with measurement and network prediction of the corrected print for a block, CIN_r and logo shown in Fig. 6. The network was trained with 5217 datasets including parameter variations.

	block			AIR			logo		
	$\Delta\bar{M} / \mu\text{m}$	$rmse / \mu\text{m}$	CC	$\Delta\bar{M} / \mu\text{m}$	$rmse / \mu\text{m}$	CC	$\Delta\bar{M} / \mu\text{m}$	$rmse / \mu\text{m}$	CC
[target, print]	3.22	0.79	1.6e-5	1.86	0.49	0.61	2.59	0.49	0.84
[target, predicted]	1.06	0.27	0.94	0.56	0.18	0.96	0.65	0.25	0.96



increased roughness of the structure surface. Moreover, the profile plots not only show a good agreement of the average structure height of the corrected print, but also

an overshooting correction for the features. This results in a stronger variation, observable in the difference plots (k) and (m). In contrast to the results before, the edges



do not show a worsening effect as correlation coefficients and amplification coefficients stay similar when the edges are not included in the calculation. Some of the increased errors, such as the higher roughness of the structures, can be

partially caused by the choice of printing parameters, e.g. a too large hatching distance. While the prediction approach for a design and a corresponding DLW parameter set should only lead to a prediction inside of a repetition accuracy

Table 4: **Quantitative comparison of Iterative (IC) and Direct Correction approach (DC), with DLW parameters.** Analysis of printing results shown in Fig. 7. Both correction approaches are quantified by mean height difference $\Delta\bar{M}$, height corrected root mean square error $rmse$, Pearson correlation coefficient CC (of the inner 80 % of the surface CC_{80}), arithmetic and squared surface roughness parameters S_a and S_q and axial amplification coefficient α_z (of the inner 80 % of the surface $\alpha_{z,80}$). The red and green cells indicate the additional comparison to the results of the **DC**-approach in Tab. 2, where no DLW parameters have been included. Green shows a better and red a worse result, respectively.

	AIR			CIN _r		
	target	print	DC	target	print	DC
$\Delta\bar{M} / \mu\text{m}$	0	4.198	-0.096	0	3.614	1.065
$rmse / \mu\text{m}$	0	0.605	0.620	0	0.653	0.691
CC	1	0.432	0.935	1	0.266	0.886
CC_{80}	1	0.774	0.947	1	0.666	0.893
$S_a / \mu\text{m}$	0.411	0.472	0.902	0.374	0.461	0.919
$S_q / \mu\text{m}$	0.474	0.633	1.041	0.418	0.625	1.033
α_z	1	1,229	2.179	1	1.288	2.444
$\alpha_{z,80}$	1	0.735	2.175	1	0.672	2.540

range, the **Direct Correction** approach has a theoretically infinite amount of possible design-parameter-combinations. This results in a higher chance of including error sources which could not be mapped during the training process of the network. However, while the red marked values in Tab. 4 show worse results in comparison to the **DC**-results without DLW parameters (Tab. 2), especially for the correlation coefficient a few improved results are noticeable. This is actually surprising, since the amount of training datasets per parameter-set is lower, hence, underlining the potential of the chosen approach. A better prediction implies a better suitability of the **Iterative Correction**-approach in the future, as seen in the studies without including the DLW parameter. This requires the development of a suitable algorithm to update both, design and parameters at the same time without overshooting.

4 Summary & Outlook

In this work, different neural networks were developed to predict and correct for deviations between the anticipated structures and their respective 2PP printed counterparts. The networks were trained on both, experimental and theoretical datasets and show generally convincing results for 2.5D structures. Our approaches demonstrate that using neural networks is a promising alternative to conventional iterative correction methods. Once the network is trained, directly pre-corrected designs can be printed. Therefore, time consuming iterations can be reduced saving experimental effort. Thereby, the studies of exemplary benchmark structures illustrate a reduction of the root mean square error between 74% to 93% in the case without varied DLW parameters. The approach including DLW parameters show reduction of the $rmse$ between 46% and 85%, emphasizing the importance of increasing training data sets. Across both cases an increase of the correlation coefficient to the range between 0.79 and 0.93 is achieved. Depending on

many factors such as the amount of training data, used structure types and parameters, the result for every structure type might differ strongly. However, the numbers show the potential of this approach, making further development worthwhile.

Since the performance of the network prediction is crucially influenced by the amount of training data, the study is currently restricted to a single device and resin. Thereby, the utilization of different laser sources and devices can yield divergent printing results [44, 45] and resins show characteristic behavior regarding feature sizes, surface roughness and shrinkage. Nonetheless, there is no general physical limitation to extending this approach to a larger number of materials and devices. However, for each added parameter, the amount of necessary training data increases significantly, requiring new studies for proper evaluation and further step by step development of the model. For future work, more investigations are necessary to further improve the performance of neural networks and to extend its applicability to various structure types and sizes, such as real 3D and stitched structures. Additionally, more fabrication parameters, like (varying) hatching directions and polarization, fabrication speed or the numerical aperture of the objective lens are to be implemented to cover a significantly broadened application area in the field of DLW. Furthermore, calibration structures for transferring the neural network outputs to different DLW machines must be developed, to make this machine learning approach beneficial for all DLW working groups.

Acknowledgement

The authors gratefully acknowledge the institute for measurement and sensor technology (MTS) at the RPTU University Kaiserslautern-Landau for the opportunity of taking confocal measurements and the measuring know how.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 172116086 – SFB 926.

Author contributions

Sven Enns developed all neural networks, carried out all the experimental work and played the key role in writing this paper. He was supervised in terms of content and concept by Julian Hering-Stratemeier and Georg von Freymann. The latter two also contributed to the writing of this manuscript.

Conflict of interest

The authors declare no conflicts of interest.

References

- [1] Maruo, S., Nakamura, O. & Kawata, S. Three-dimensional microfabrication with two-photon-absorbed photopolymerization. *Optics letters* **22**, 132–134 (1997).
- [2] Fischer, J. & Wegener, M. Three-dimensional optical laser lithography beyond the diffraction limit. *Laser & Photonics Reviews* **7**, 22–44 (2013).
- [3] Wollhofen, R. et al. Functional photoresists for sub-diffraction stimulated emission depletion lithography. *Optical Materials Express* **7**, 2538–2559 (2017).
- [4] Haria, R. Nanoscribe Photonic Professional GT 3D printer produces microscopic molds in series. (2017). At <https://3dprintingindustry.com/news/3d-printer-nanoscribe-photonic-polymer-molds-125041/>.
- [5] Fischer, J. et al. Three-dimensional multi-photon direct laser writing with variable repetition rate. *Optics Express* **21**, 26244–26260 (2013).
- [6] Purto, J. et al. Nanopillar diffraction gratings by two-photon lithography. *Nanomaterials* **9**, 1495 (2019).
- [7] Pingali, R. & Saha, S. K. Reaction-diffusion modeling of photopolymerization during femtosecond projection two-photon lithography. *Journal of Manufacturing Science and Engineering* **144**, 021011 (2022).
- [8] Guney, M. G. & Fedder, G. K. Estimation of line dimensions in 3d direct laser writing lithography. *Journal of Micromechanics and Microengineering* **26**, 105011 (2016).
- [9] Adão, R. M. R. et al. Two-photon polymerization simulation and fabrication of 3d microprinted suspended waveguides for on-chip optical interconnects. *Optics Express* **30**, 9623–9642 (2022).
- [10] Waller, E. H. & von Freymann, G. Spatio-temporal proximity characteristics in 3d μ -printing via multi-photon absorption. *Polymers* **8**, 297 (2016).
- [11] Yang, L. et al. On the schwarzschild effect in 3d two-photon laser lithography. *Advanced Optical Materials* **7**, 1901040 (2019).
- [12] Ovsianikov, A. et al. Shrinkage of microstructures produced by two-photon polymerization of zr-based hybrid photosensitive materials. *Optics Express* **17**, 2143–2148 (2009).
- [13] Lang, N. et al. Towards efficient structure prediction and pre-compensation in multi-photon lithography. *Optics Express* **30**, 45302 (2022).
- [14] Wang, C. et al. Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing* **36**, 101538 (2020).
- [15] Qin, J. et al. Research and application of machine learning for additive manufacturing. *Additive Manufacturing* **52**, 102691 (2022).
- [16] Faegh, M. et al. A review on physics-informed machine learning for process-structure-property modeling in additive manufacturing. *Journal of Manufacturing Processes* **133**, 524–555 (2025).
- [17] Thompson, A. J., Power, M. & Yang, G.-Z. Micro-scale fiber-optic force sensor fabricated using direct laser writing and calibrated using machine learning. *Optics Express* **26**, 14186–14200 (2018).
- [18] Lee, X. Y. et al. Automated detection of part quality during two-photon lithography via deep learning. *Additive Manufacturing* **36**, 101444 (2020).
- [19] Mourka, A. et al. Machine learning predicts printing parameters for multi-photon polymerization three-dimensional direct laser writing (3d-dlw) (conference presentation). (Proceedings of SPIE 11271, Laser 3D Manufacturing VII. San Francisco: SPIE, 2020).
- [20] Yang, Y. H. et al. Machine-learning-enabled geometric compliance improvement in two-photon lithography without hardware modifications. *Journal of Manufacturing Processes* **76**, 841–849 (2022).
- [21] Liu, X. J. et al. 3d printing of bioinspired liquid superrepellent structures. *Advanced Materials* **30**, 1800103 (2018).
- [22] Nanoscribe GmbH & Co. KG. NanoGuide – IP-S (2023). At <https://support.nanoscribe.com/hc/en-gb>.

- [23] Eifler, M. et al. Calibration of z-axis linearity for arbitrary optical topography measuring instruments. *Proceedings of SPIE 9525, Optical Measurement Systems for Industrial Inspection IX*. Munich: (SPIE, 2015).
- [24] Dai, G. L. et al. Define and measure the dimensional accuracy of two-photon laser lithography based on its instrument transfer function. *Journal of Physics: Photonics* **3**, 034002 (2021).
- [25] ISO. Geometrical product specification (GPS) – surface texture: Areal – part 70: Material measures: Iso 25178-70:2014 (2014).
- [26] Eifler, M. et al. Calibration sample for arbitrary metrological characteristics of optical topography measuring instruments. *Optics express* **26**, 16609–16623 (2018).
- [27] Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data* **6**, 60 (2019).
- [28] Eifler, M. et al. Towards a continuous frequency band chirp material measure for surface topography measuring instrument calibration. *Proceedings of SPIE 11782, Optical Measurement Systems for Industrial Inspection XII* (SPIE, 2021).
- [29] Abbott, E. J. & Firestone, F. A. Specifying surface quality - a method based on accurate measurement and comparison. *Journal of Mechanical Engineering* **55**, 569–572 (1933).
- [30] NanoFocus AG. NanoFocus – μ Surf Operating Manual (June 2003). (Delivered with the device. Not available online.).
- [31] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe: (Curran Associates, Inc., 2012).
- [32] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition (2015). Print at <https://doi.org/10.48550/arXiv.1409.1556>.
- [33] Skliutas, E. et al. Multiphoton 3d lithography. *Nature Reviews Methods Primers* **5**, 15 (2025).
- [34] Török, P. et al. Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: an integral representation. *Journal of the Optical Society of America A* **12**, 325–332 (1995).
- [35] Török, P. et al. Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: an integral representation: errata. *Journal of the Optical Society of America A* **12**, 1605–1605 (1995).
- [36] Rekštytė, S. et al. Nanoscale precision of 3d polymerization via polarization control. *Advanced Optical Materials* **4**, 1209–1214 (2016).
- [37] Pearson, K. VII. note on regression and inheritance in the case of two parents. *Proceedings of the royal society of London* **58**, 240–242 (1895).
- [38] Dogo, E. M. et al. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. *Proceedings of 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems*. Belgium: (IEEE, 2018).
- [39] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization (2017). Print at <https://doi.org/10.48550/arXiv.1412.6980>.
- [40] Keras - adam. <https://keras.io/api/optimizers/adam/>.
- [41] Nesterov, Y. E. A method for solving a convex programming problem with convergence rate $O(1/K^2)$. *Soviet Mathematics Doklady* **27**, 372–376 (1983).
- [42] Sutskever, I. et al. On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning*. Atlanta: (JMLR.org, 2013).
- [43] Keras - Nadam. At <https://keras.io/api/optimizers/Nadam/>.
- [44] Cantoni, F. et al. Round-robin testing of commercial two-photon polymerization 3d printers. *Additive Manufacturing* **76**, 103761 (2023).
- [45] Ladika, D. et al. X-photon 3d lithography by fs-oscillators: wavelength-independent and photoinitiator-free. *Light: Advanced Manufacturing* **5**, 48 (2024).
- [46] Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems (2016). Print at <https://doi.org/10.48550/arXiv.1603.04467>.
- [47] A superpower for ml developers. At <https://keras.io> (2015).

5 Appendix

5.1 MSE as Loss Function

As mentioned earlier, the loss function is crucial for the training process of a neural network. Not only the convergence of the training but also the learned features and hence the network performance is strongly affected. Here, the widely used mean square error MSE as loss function is evaluated. While the MSE is a simple loss function and in many cases well suited, the complexity of our approach cannot be mapped. As exemplary visualized in Figure 8, the MSE function is able to predict the average structure height. However, especially small features vanish due to averaging over the whole structure surface. Therefore, a more complex loss function is needed. The illustration emphasizes the importance of the loss function and reaffirms our choice of the loss function as described in chapter 2.4.

5.2 Training Data

As listed in Tab. 5, several different types of structures are used for training the neural networks. Thereby, each structure has been generated with varying parameters, such as its height or the dimension of sub-features. Hence, different amounts of each structure type have been included into the training datasets. In order to get an initial pool

of training data, the 628 structures listed in Tab. 5 are fabricated. To keep the amount of varied parameters low for the first steps, DLW parameters are not varied in this case. Building on these 628 structures, the amount of training data is increased to implement the DLW parameters hatching & slicing distance and laser power into the machine learning approaches. Therefore, 2137 additional structures are fabricated. While the structure types stay similar to the first 628, this time DLW parameters are varied. Hence, for similar structure types, the influence of the DLW parameters can be extracted from the new data sets. To further increase the amount of training data, 2452 more sets are generated using data augmentation. In total 5217 data sets are available, listed in Tab. 6.

Some of those structures are exemplary illustrated as CAD images in Fig. 9.

5.3 Padding Influence

As described in the main manuscript, padding does have an influencing factor on the neural networks' outputs. Figure 10 illustrates these influences onto the results of different neural network architectures and padding options in comparison to the actual measurement.

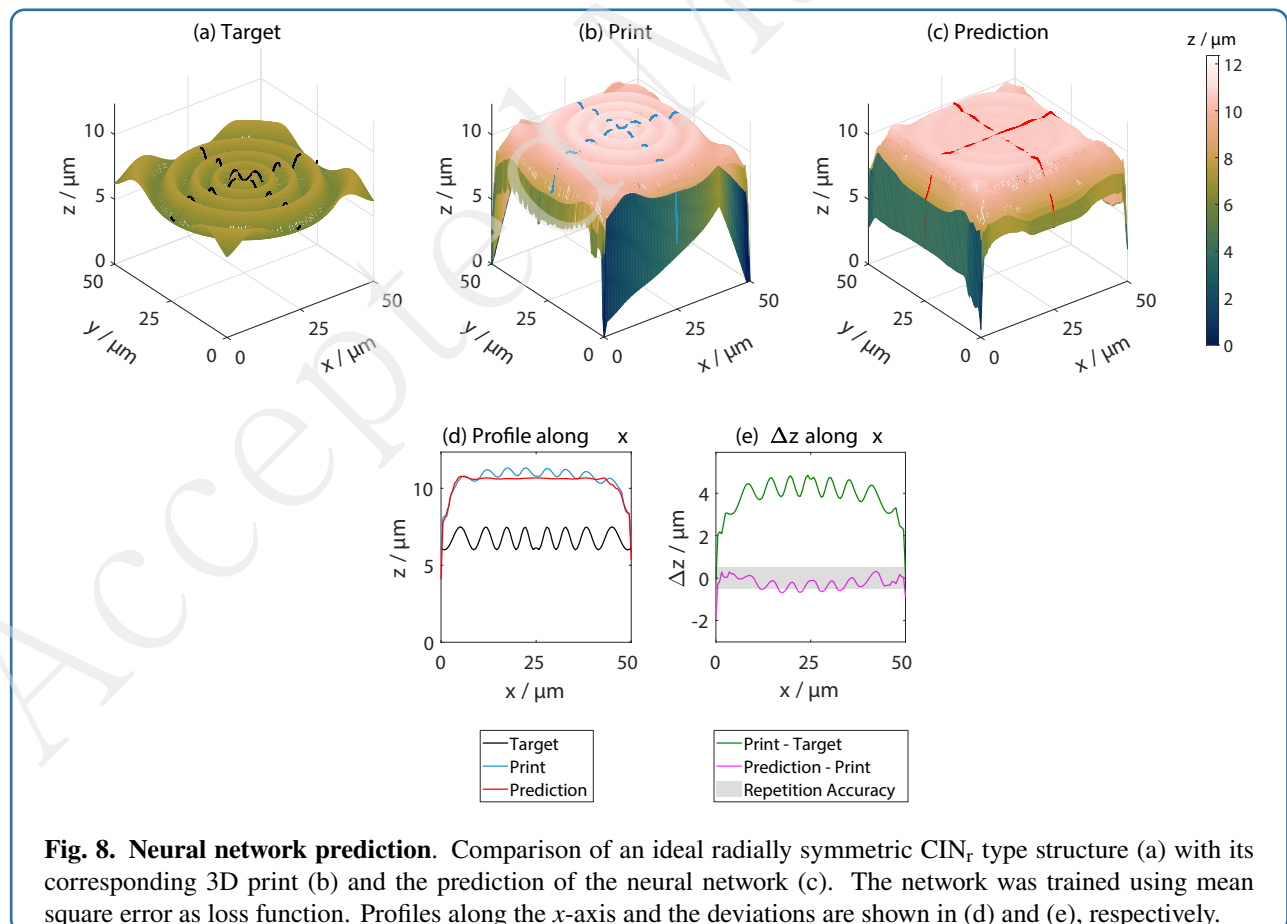


Fig. 8. Neural network prediction. Comparison of an ideal radially symmetric CIN_r type structure (a) with its corresponding 3D print (b) and the prediction of the neural network (c). The network was trained using mean square error as loss function. Profiles along the x -axis and the deviations are shown in (d) and (e), respectively.

Table 5: **Training structures without DLW parameters.** This dataset serves as basic training data without variation of DLW parameters. Slicing and hatching distances of $0.1 \mu\text{m}$ as well as the laser power at roughly 45 mW are kept constant. All 628 structures are physically fabricated and no data augmentation is used.

structure type	# of samples	varied parameters	value range
block – corrected block	61 – 48	minimum height	$(2 - 17) \mu\text{m}$
step – corrected step	70 – 48	minimum height	$(5 - 10) \mu\text{m}$
		step height	$(0.5 - 4) \mu\text{m}$
		step position	$x = (16 - 34) \mu\text{m}$
CIN _r -type – corrected CIN _r -type (radial)	227 – 43	minimum height	$(2 - 10.5) \mu\text{m}$
		amplitude	$(0.4 - 1) \mu\text{m}$
		minimum frequency	$(0.005 - 0.06) \mu\text{m}^{-1}$
		maximum frequency	$(0.1 - 0.4) \mu\text{m}^{-1}$
cosine structure (x -direction)	23	minimum height	$(2 - 8) \mu\text{m}$
		amplitude	$(0.5 - 3) \mu\text{m}$
		frequency	$(0.02 - 0.075) \mu\text{m}^{-1}$
CIN (x -direction)	15	amplitude	$(0.5 - 0.85) \mu\text{m}$
radially symmetric step	20	step height	$(-4 - 4) \mu\text{m}$
		step diameter	$(10 - 20) \mu\text{m}$
array of small blocks on a large block	13	# of blocks	$1 \times 1 - 3 \times 3$
		height	$(3 - 7) \mu\text{m}$
		edge length	$(2 - 20) \mu\text{m}$
various shapes on a large block	30	minimum height of large block	$(2 - 12) \mu\text{m}$
		height of additional structures	$(0.25 - 3) \mu\text{m}$
AIR-type	30	minimum height	$(2 - 10) \mu\text{m}$
		height scaling	$(1 - 5) \mu\text{m}$

Table 6: **Additional training data.** This data is implemented to cover for the DLW parameters, added as input or output of the neural networks. For all structures, the slicing and hatching distance was varied between $(0.05 - 0.7) \mu\text{m}$ and the laser power was varied between approximately $(24 - 60) \text{ mW}$. Here, in total 2765 structures are physically fabricated, including the 628 data sets shown in Tab. 5. The remaining 2452 structures are generated using data augmentation.

structure type	# of samples
block – corrected block	219 – 23
step – corrected step	224 – 19
staircase structure	167
CIN-type – corrected CIN-type	291 – 10
AIR-type	257
cosine structure (x -direction)	290
radially symmetric step	168
wedge-shaped structure	168
paraboloid structure	165
various shapes on a large block	168
random structures	28
data augmentation	2452

5.4 Iterative Correction

Since the **Iterative Correction (IC)** approach leads to enhancement of random artefacts, this is exemplary shown in Fig. 11 for the first six iteration steps. The artefacts are most prominent at the edges of a structure.

5.5 Computation

The neural networks were implemented in Python using TensorFlow [46], an open-source machine learning platform, in combination with the API Keras [47]. For further data handling NumPy, SciPy, and scikit-learn packages were additionally used. In Tab. 7 the computation times to train a network are listed in dependence on the amount of datasets, training iterations and utilized hardware.

Table 7: **Computation resources.** Overview over the resources used for training the different amount of datasets.

# of training datasets	iterations	used RAM / GB	used CPUs	used GPUs	training duration / d:h:m
497	1000	1.5	8	-	0:11:59
568	1000	1.6	8	-	0:14:41
628	1000	1.8	4	-	1:14:15
1886	500	6.1	8	1x Volta V100	1:14:44
2779	500	8.8	12	1x Volta V100	2:00:42
5217	1000	18.5	16	1x Volta V100	6:13:44

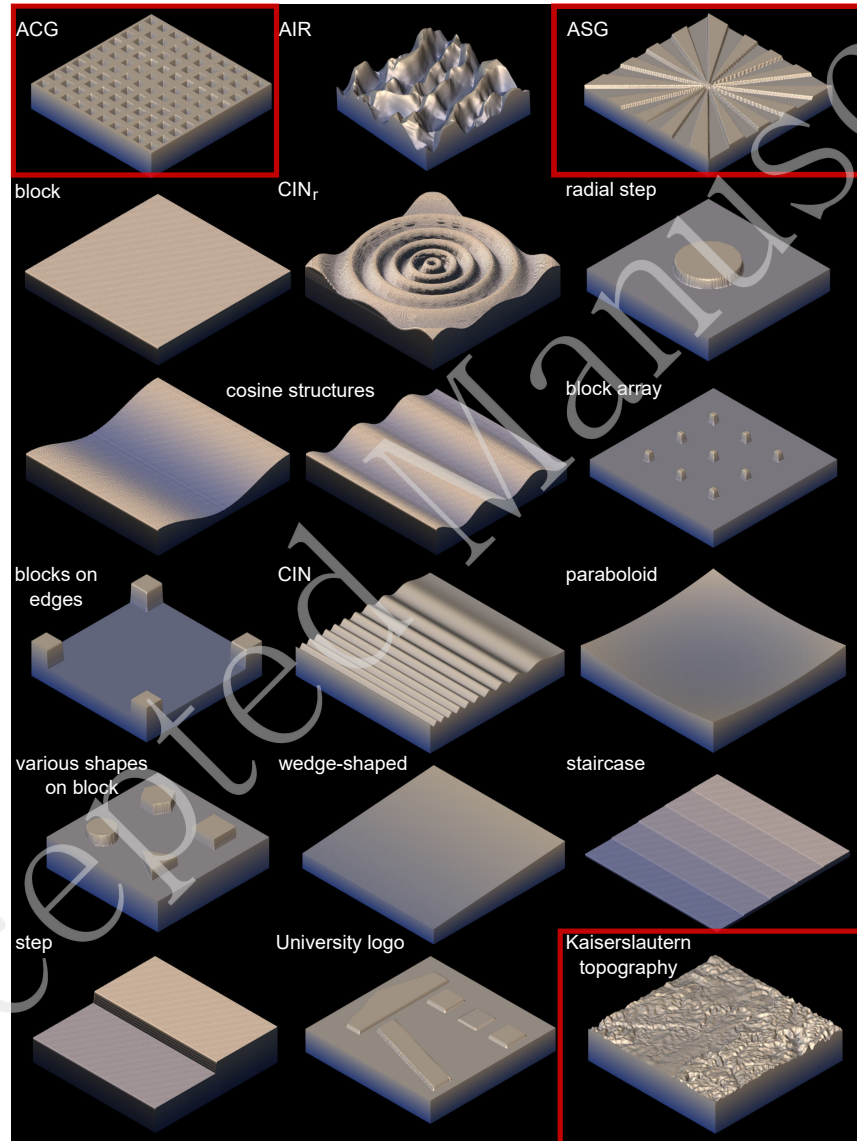


Fig. 9. CAD schemes. Exemplary 3D illustrations of the structures, used for training and testing and developing of the neural networks. The ACG, ASG and topography structures marked in red were used to develop and test the models and the measurement and data processing workflow. Because of more frequent occurrence of errors due to lack of resolution or measurement artefacts, those structures were removed from training data. As measurement artefacts lead to a distorted analysis of the network performance, the results are not published as benchmark.

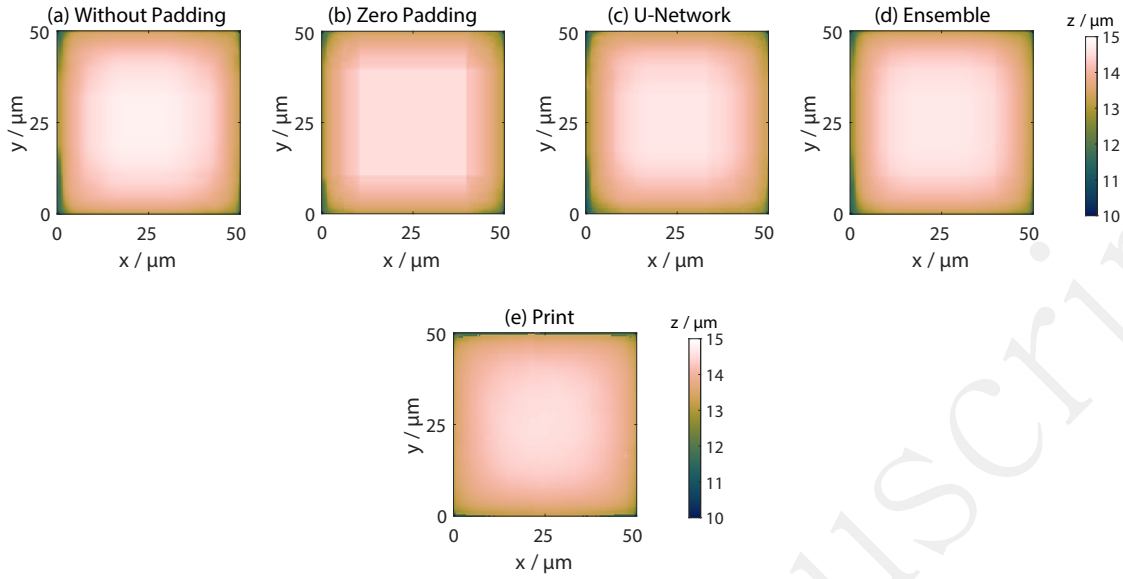


Fig. 10. Comparison of neural network architectures. Prediction of a block by different architectures regarding padding and image sizes throughout the neural networks (a-c), where each architecture shows specific characteristics in its prediction result. Combined result of the four different approaches via ensemble learning is shown in (d). The actual printing result is shown in (e).

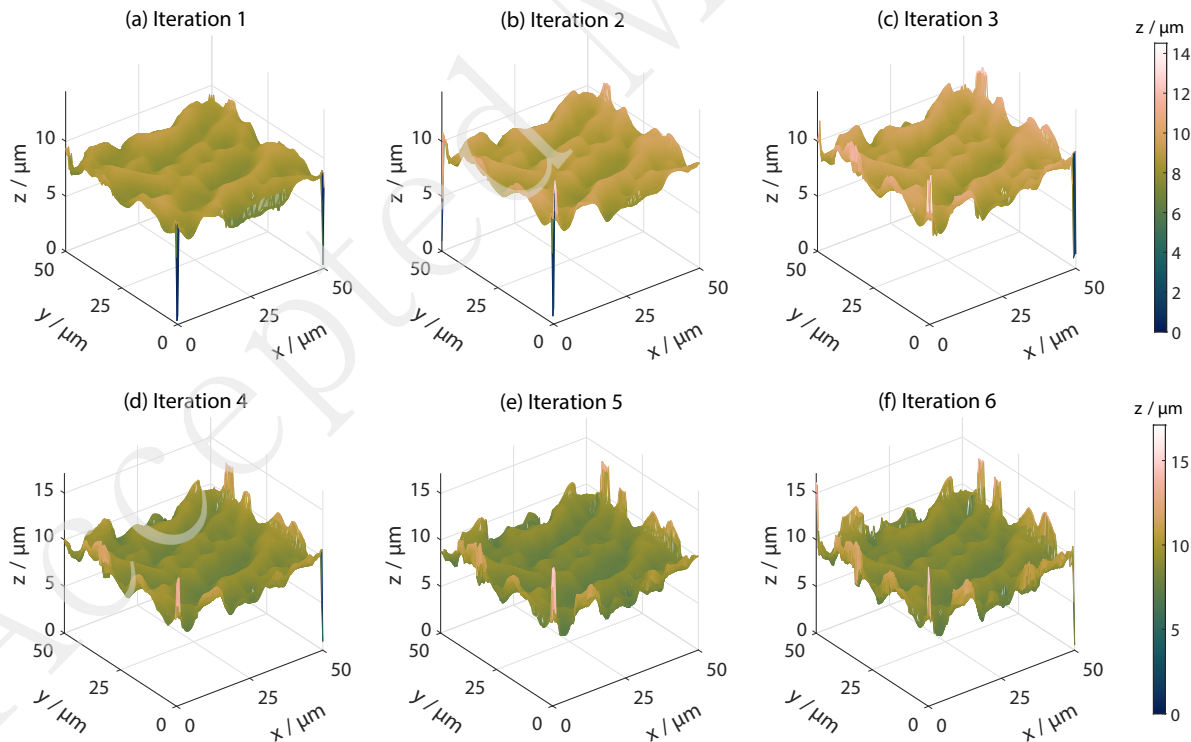


Fig. 11. Iterative Correction Approach. Shown are the neural network's corrected printing results for the first six iterations of an AIR-type structure. For higher iteration steps the increasing influence of artefacts, especially at the structure's edges becomes visible.

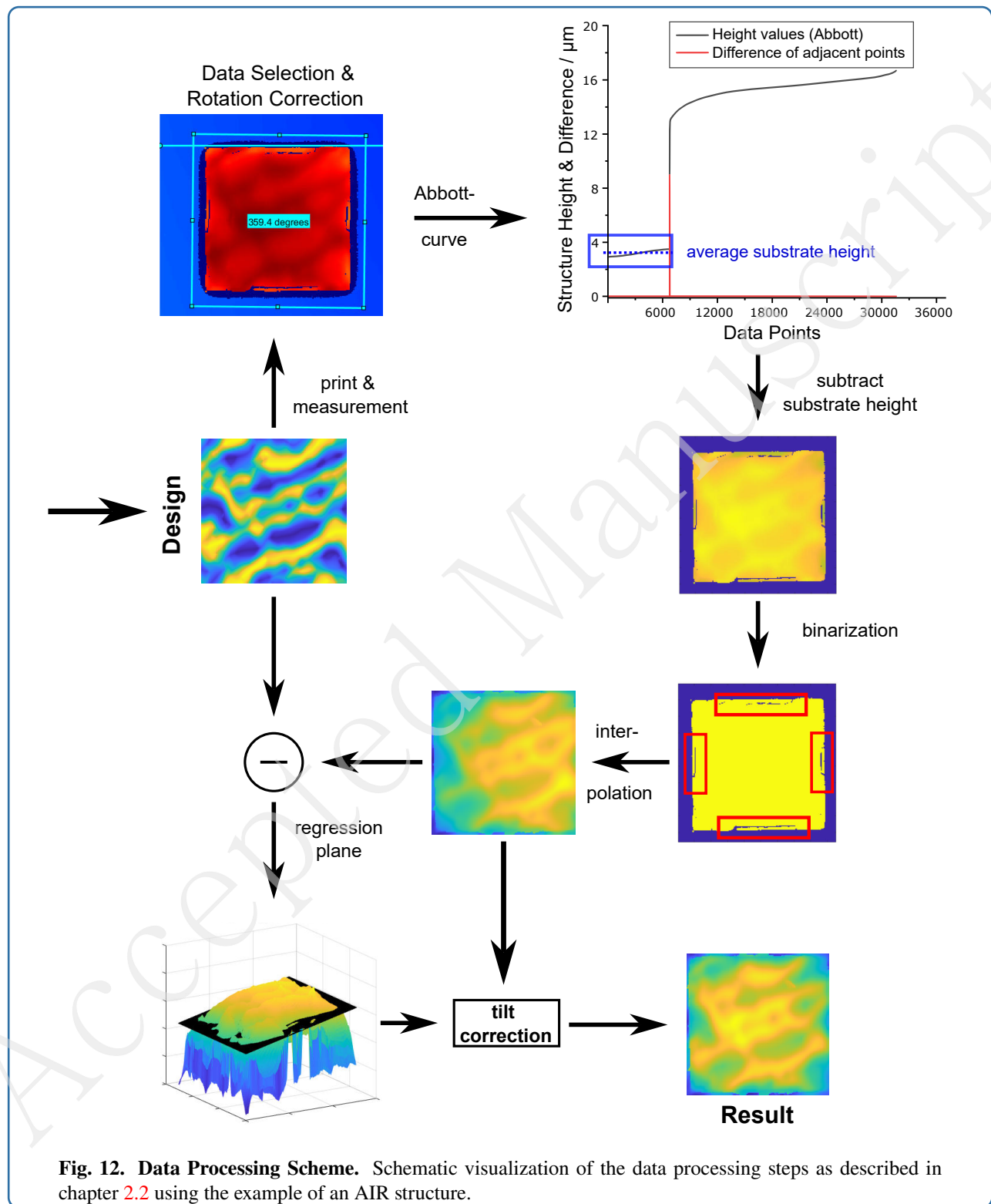


Fig. 12. Data Processing Scheme. Schematic visualization of the data processing steps as described in chapter 2.2 using the example of an AIR structure.