



Deep learning for laser beam imprinting

J. CHALUPSKÝ,^{1,*}  V. VOZDA,¹  J. HERING,² J. KYBIC,² T. BURIAN,^{1,3} S. DZIARZHYTSKI,⁴ K. FRANTÁLOVÁ,¹ V. HÁJKOVÁ,¹ Š. JELÍNEK,^{1,3,5}  L. JUHA,¹ B. KEITEL,⁴ Z. KUGLEROVÁ,^{1,5} M. KUHLMANN,⁴ B. PETRYSHAK,² M. RUIZ-LOPEZ,⁴  L. VYŠÍN,¹ T. WODZINSKI,⁶  AND E. PLÖNJES⁴

¹*Institute of Physics, Czech Academy of Sciences, Na Slovance 2, 182 21 Prague 8, Czech Republic*

²*Faculty of Electrical Engineering, Czech Technical University in Prague, 166 27, Prague 6, Czech Republic*

³*Institute of Plasma Physics, Czech Academy of Sciences, Za Slovankou 3, 182 00 Prague 8, Czech Republic*

⁴*Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany*

⁵*Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, 121 16 Prague 2, Czech Republic*

⁶*GoLP/Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, 1049-001 Lisboa, Portugal*

*chal@fzu.cz

Abstract: Methods of ablation imprints in solid targets are widely used to characterize focused X-ray laser beams due to a remarkable dynamic range and resolving power. A detailed description of intense beam profiles is especially important in high-energy-density physics aiming at nonlinear phenomena. Complex interaction experiments require an enormous number of imprints to be created under all desired conditions making the analysis demanding and requiring a huge amount of human work. Here, for the first time, we present ablation imprinting methods assisted by deep learning approaches. Employing a multi-layer convolutional neural network (U-Net) trained on thousands of manually annotated ablation imprints in poly(methyl methacrylate), we characterize a focused beam of beamline FL24/FLASH2 at the Free-electron laser in Hamburg. The performance of the neural network is subject to a thorough benchmark test and comparison with experienced human analysts. Methods presented in this Paper pave the way towards a virtual analyst automatically processing experimental data from start to end.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

With the advent of intense X-ray free-electron lasers (XFELs), new unique experimental opportunities emerged. Most of the experiments benefit from the extraordinary beam parameters and versatility of these sources. Ultra-short pulses and high transverse coherence allow investigating ultra-fast processes with excellent spatio-temporal resolution utilizing, for example, methods of coherent diffractive imaging [1,2] or two-color pump-probe diffraction techniques [3]. Using precise focusing optics, very high intensities can be achieved making the nonlinear regime of laser-matter interactions accessible. This enabled to study nonlinear absorption phenomena in dilute environments [4] as well as to create and probe warm/hot dense matter [5–9]. X-ray free-electron lasers initiated a new era of laser-matter interaction research as documented, for example, in review papers by Rossbach [10] and Bostedt [11].

In parallel to user research, an active development of X-ray free-electron lasers is ongoing. New generation of XFELs, e.g. European XFEL [12] and LCLS II [13], targets at high repetition rates significantly increasing data acquisition speed. This enables more complex studies aiming at systematic variation of beam conditions, e.g. photon energy, intensity, pump-probe pulse delay, etc. However, such studies put immense demands on photon beam diagnostics playing a crucial role in experimental data evaluation. A proper focused beam characterization is an important prerequisite for high-energy-density experiments aimed at nonlinear phenomena in laser-matter

interactions. Since the studied nonlinearity is weighted by the intensity profile spanning over many orders of magnitude, gathered experimental data, whether it be spectroscopic, diffraction or otherwise, represent a mixture of contributions emerging from states of variable temperature. Discrimination between the hot and cold state thus relies on a good knowledge of the beam profile.

Unlike non-invasive online diagnostics [14,15], spatial beam characterization usually requires full beam usage. Characterization of the focus hits obvious difficulties specific for each particular method. Excessive radiation intensities and small focus dimensions are usually treated by increasing the focus-to-detector distance and/or by additional beam attenuation. The latter may be, however, accompanied by an undesirable relative increase of high-order harmonics content. Existing indirect and semidirect methods [16–21] exploit numerical detector-to-focus backpropagation and/or phase retrieval approaches employing Fresnel or Fraunhofer diffraction integrals. However, the implementation of rather unknown and variable partial coherence is required in order to avoid an underestimation of the focus [22]. A few semidirect and direct methods avoid numerical beam propagation by introducing an in-focus scanning mask [23], off-axis zone plate [24], or lithium fluoride crystal [25]; however, the proximity of the focus requires a strong beam attenuation preventing damage to the inserted mask. The direct method of ablation imprints [26–29] exploits damage to a target surface to characterize the beam, hence less beam attenuation is needed. However, the method is off-line and laborious, requiring a lot of human work, expertise, and data processing. All methods have specific advantages and drawbacks stimulating further developments.

Systematic experimental studies obviously require variable beam conditions that must be characterized. Methods of ablation imprints are nowadays used to characterize focused beams employed in experiments [5,7,8]. However, the amount of created imprints can easily reach thousands depending on the variability of the experiment. All imprints are investigated *ex-situ* using microscopy techniques and ablation threshold contours are manually annotated with use of a graphical drawing tablet. Although some attempts to automatize the microscopy data acquisition have already been performed [30], autonomous image segmentation for this application remains unsolved. A specific nature of various target surfaces and presence of intrinsic and laser-induced surface artifacts lead to a very high variability of captured scenes. Furthermore, color schemes, exposure brightness, resolution, and other image features depend significantly on the microscope and camera used. Hence classical segmentation approaches often fail. On the other hand, convolutional deep neural networks represent useful and robust analytic tools applicable in problems with ambiguously defined initial conditions. Networks of various architectures find utilization, for example, in microscopy [31], photonics [32], phase-retrieval [33] and bioimaging [34–36].

Here we present the first application of a multi-layer convolutional neural network (CNN) U-Net to real ablation imprints in poly(methyl methacrylate) (PMMA). Section 2 introduces the theoretical basis of ablation imprints methods. Section 3 describes the beam characterization experiment carried out at beamline FL24 at FLASH2. Section 4 introduces the U-Net architecture and summarizes methods of network training/benchmarking. Section 5 provides a discussion of U-Net's performance in confrontation with human analysts and compares results of the fluence scan method [26,28] applied to retrieved imprint data. To maintain the clarity of the text, some details of the mathematical apparatus are summarized in Appendices.

2. Theory

The fundamental principle of ablation imprints methods resides in the fact that ablation induced by a femtosecond X-ray laser pulse occurs above a sharply defined ablation threshold. This leads to a formation of a sharp threshold contour that coincides with an iso-fluence beam contour located exactly at the ablation threshold fluence F_{th} . Among other threshold processes,

non-thermal melting [37] belongs to the most promising ablation processes applicable in laser beam imprinting. As the characteristic time scale of non-thermal phase transitions is only a few hundreds of femtoseconds [38], the ablation threshold contour is formed prior to onset of thermal effects. Furthermore, as the threshold fluence is rather low, the ablation threshold contour usually occurs at locations sufficiently distant from the hot beam center. These two important aspects of the interaction protect the contour from being compromised by hydrodynamic or thermomechanic effects and significantly extend the dynamic range of reliable application of imprinting methods.

A general laser beam freely propagating in z -direction can be described by its fluence profile $F(\boldsymbol{\rho}, z) = F_0(z)f(\boldsymbol{\rho}, z)$ where the vector $\boldsymbol{\rho} = (x, y)$ denotes transverse coordinates and $F_0(z)$ is the peak fluence at a given z -position. It follows from this definition that the maximum of the normalized fluence profile $f(\boldsymbol{\rho}, z)$ equals unity at all z -positions along the propagation axis. As the fluence $F(\boldsymbol{\rho}, z)$ tends to zero for $|\boldsymbol{\rho}|$ extending to infinity, iso-fluence curves form single or multiple closed contours of well-defined area S . This makes it possible to define the beam profile in a less general but more convenient way as [26,28]:

$$F(S, z) = F_0(z)f(S, z), \text{ where: } \forall z \in \mathbb{R} : f(0, z) = 1. \quad (1)$$

A connection between the definition in Eq. (1) and $F(\boldsymbol{\rho}, z)$ can be explained by introducing a contour map function $S(\boldsymbol{\rho}, z)$ assigning each iso-fluence contour of $F(\boldsymbol{\rho}, z)$ a scalar value of the corresponding area. Substituting the map function into Eq. (1), we get the original 2D definition $F(S(\boldsymbol{\rho}, z), z) = F(\boldsymbol{\rho}, z)$. Integrating the fluence profile over the entire transverse plane, we get the pulse energy:

$$E_{\text{pulse}} = F_0(z) \iint_{\mathbb{R}^2} f(\boldsymbol{\rho}, z) d^2\boldsymbol{\rho} = F_0(z) \int_0^\infty f(S, z) dS = F_0(z)A_{\text{eff}}(z), \quad (2)$$

being independent on the z -position due to the energy conservation law. From this identity the definition of the effective beam area $A_{\text{eff}}(z)$ follows as the "volume" below the normalized 2D profile $f(\boldsymbol{\rho}, z)$ or as the area below the normalized f -scan curve $f(S, z)$ at a given z -position:

$$A_{\text{eff}}(z) = \iint_{\mathbb{R}^2} f(\boldsymbol{\rho}, z) d^2\boldsymbol{\rho} = \int_0^\infty f(S, z) dS. \quad (3)$$

The boundary condition, the solution of which determines the locus of the threshold contour and its area at a given z -position, can be respectively expressed as: $F_{\text{th}} = F_0(z)f(\boldsymbol{\rho}, z) = F_0(z)f(S, z)$. Therefore, an imprint (indexed by i, j) of contour area S_{ij} created by a pulse of energy E_{ij} at a position z_i fulfills: $F_{\text{th}}A_{\text{eff}}(z_i) = E_{ij}f(S_{ij}, z_i)$. Using Eq. (2), we introduce the threshold pulse energy $E_{\text{th}}(z_i) = F_{\text{th}}A_{\text{eff}}(z_i)$ and express a normalized $f(z)$ -scan sequence:

$$f(S_{ij}, z_i) = \frac{E_{\text{th}}(z_i)}{E_{ij}}. \quad (4)$$

The sequence in Eq. (4) is constructed from raw as measured $F(z)$ -scan imprint data $\{S_{ij}, E_{ij}, z_i\}$ where the beam is imprinted while varying both the pulse energy and z -position. Contrary to constant F_{th} , the threshold pulse energy $E_{\text{th}}(z)$ is variable and must be evaluated at each z_i -position by extrapolating the measured sequence to zero contour area; hence a sufficient number of imprints (typically hundreds) must be created at pulse energies varied between the threshold and maximum output pulse energy of the laser source. By fixing the z -position, we recover the so called fluence scan (f -scan) curve.

Equation (4) can be expressed in a very useful non-normalized form using a relative function:

$$\Phi(S_{ij}, z_i) = \frac{1}{E_{ij}} = \frac{1}{F_{\text{th}}A_{\text{eff}}(z_i)}f(S_{ij}, z_i), \quad (5)$$

which, contrary to Eq. (4), avoids the necessity of $E_{\text{th}}(z_i)$ evaluation. While the left-hand side of the equation can be constructed solely from raw ablation imprint data $\{S_{ij}, E_{ij}, z_i\}$, the right-hand

side is proportional to the fluence profile defined in Eq. (1). A beam profile of the peak fluence $F_0(z) = E_{\text{pulse}}/A_{\text{eff}}(z)$ can be conveniently expressed as $F(S, z) = F_{\text{th}}E_{\text{pulse}}\Phi(S, z)$. This allows to directly compare imprint data with a modeled or otherwise measured fluence profile $F(S, z)$ while the scaling factor F_{th} , if unknown, can be treated as a fitting parameter. The applicability of the $F(z)$ -scan method is limited to a z -range in which the maximum output pulse energy E_{max} exceeds the ablation threshold such that $E_{\text{max}} > F_{\text{th}}A_{\text{eff}}(z)$; otherwise the imprints are no longer visible.

3. Experimental

The experiment was performed at beamline FL24 of the FLASH2 (Free-electron Laser in Hamburg, Germany) facility [39]. During the experimental campaign, laser beams of several wavelengths, particularly, 18 nm, 13.5 nm, and 8 nm, were studied. The beam was focused with use of an adaptive Kirkpatrick-Baez (KB) optical system [40] consisting of two independently bendable elliptical mirrors which allow positioning of the focus at a desired location. Astigmatism and coma of the focusing optics was minimized with the aid of the wavefront sensor located approx. 4.3 m downstream from KB mirrors. The pulse energy was varied using thin-foil solid attenuators, namely, niobium, zirconium, aluminum, and silicon. Filter combinations were always chosen such that the relative content of the 3rd harmonic frequency stayed at reasonably low level below 10% of the fundamental harmonic. Energy of each individual pulse was monitored with use of a gas monitor detector [15].

Imprinting targets were composed of 5- μm thin layer of poly(methyl methacrylate) spin-coated on silicon wafers (dimensions: 20×10 mm). Samples were mounted on an in-vacuum 4-axis translational-rotational stage that allowed translation in directions normal to and parallel with the laser beam and rotation around the vertical axis [30]. The interaction chamber was equipped with a long working distance microscope intended for *in-situ* inspection and precise target alignment. The center of the ultrahigh vacuum interaction chamber was located approx. 2 m downstream from the KB optics at the expected focus position. The target holder was controlled using a semi-automated script executing predefined exposure protocols. During the experiment, two types of imprint scans were performed. The through-focus $F(z)$ -scan was carried out to precisely determine the focus position. The $F(z)$ -scan protocol was set to scan along the beam axis through the expected focus in 1-mm steps (40 positions in total). The z -scan was repeated at 12 attenuation levels between the estimated PMMA threshold in focus and the maximum FEL output power. Subsequently, the fluence scan was performed at the focus and 20 mm downstream of the focus. Fine-resolved f -scans were composed of up to 42 attenuation levels, each numbering 20 single-shot imprints. Several thousands of ablation imprints were created, representing a sufficiently large basis for training, validation, and testing of the neural network. Irradiated samples were inspected *ex-situ* using a Nomarski microscope Olympus BX51M equipped with a Canon EOS800D digital camera. Ablation threshold contours were manually annotated using an XP-Pen Artist 24 graphical drawing tablet. Training and testing of the U-Net script was carried out using a desktop computer equipped with an Intel Core i7-11700 processor and 128-GB DDR4 RAM. Parallelized computations were performed with the aid of the GeForce RTX 3080 Ti 12GB graphical processing unit (GPU).

4. Methods

4.1. U-Net architecture

Segmentation of the ablation imprint interior in microscopy images is done using a convolutional neural network employing the deep learning U-Net architecture [35] originally developed for biological microscopy images. The network as shown in Fig. 1 has two major parts: the encoder (left part of the U-shape) successively aggregates semantic information, increasing the number of channels while reducing spatial resolution in a series of steps. In each layer, a spatial convolution

with a small learnable kernel and a ReLU (rectified linear unit) nonlinearity is applied as indicated by blue arrows. The spatial resolution is reduced by max-pooling (red arrows), taking maximum over 2×2 pixel blocks. The decoder (right part of the U-shape) then progressively interpolates back to the original resolution by convolving with another filter (green arrow). It combines information from previous levels with information obtained directly from the encoder at the current level through the so-called skip connections (grey arrows). In this way, the segmentation network can take into account both local information and the larger context. The U-Net has been shown to produce pixel-accurate segmentations even when trained on relatively modest size datasets.

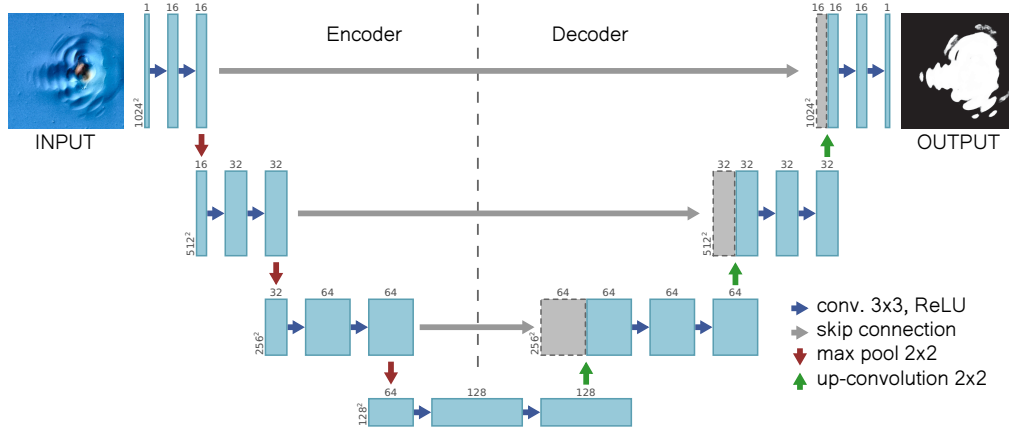


Fig. 1. Schematic of the U-Net architecture with an input image size $N = 1024$, $l_U = 4$ descent levels, $e_U = 2$ encoding layers per level, and $k_U = 16$ convolution kernels in the first level. Light blue and gray boxes represent three-dimensional matrices (multi-channel feature maps) with dimensions indicated on the left and above the box. The left and right part of the image depicts the encoder (contracting path) and decoder (expansive path), respectively. Convolutions with learnable 3×3 kernels combined with the ReLU operation are indicated by blue arrows. Red downward arrows depict the max-pooling operation halving the image resolution. Green upward arrows represent 2×2 up-convolutions (upsampling) doubling the image resolution. In each descent level, skip connections (grey arrows) transfer information from encoder to decoder to better localize fine features.

There are now many variants of the U-Net architecture. As input we use square images of size $N \times N$. Other parameters include the number of descent levels l_U , the number of encoding layers at each level e_U , and the number of convolution kernels in the first dimension k_U . Higher values of l_U , e_U , and k_U lead to a larger and more powerful network with more parameters which is usually capable of better segmentation results at the cost of increased memory and computational time. It also requires more training data to avoid overfitting.

Once the shape of the network is fixed, the network is parameterized by a weight vector θ . Given a set $\mathcal{T} = \{X_j, M_j\}$ of pairs of training images (X_j) and manually annotated (ground-truth) binary masks (M_j), the network U_θ is trained by minimizing the total loss function $\sum_j L(Q_j, M_j)$ quantifying the sum of differences between network predictions $Q_j = U_\theta(X_j)$ (probability maps) and ground-truth binary masks (M_j). The difference is expressed by means of the standard binary cross-entropy loss function [41]:

$$L(Q, M) = - \sum_{i \in \Omega} [M(i) \ln Q(i) + (1 - M(i)) \ln(1 - Q(i))]. \quad (6)$$

The summation is done over the set Ω of all elements (pixels indexed by i) of the mask image $M(i)$ and predicted probability map $Q(i)$. The cross-entropy loss function L lends itself very

well to the optimization task, since it is differentiable with respect to Q . The loss L achieves its minimum $L = 0$ if $Q(i) = M(i)$ for all pixels $i \in \Omega$ and it grows over all bounds the more Q differs from M . It is the most popular cost function for image segmentation; it is convex and its derivative is easy to calculate. The minimization is done using the Stochastic Gradient Descent (SGD) optimizer with Nesterov's momentum m_U [42]. The learning rate γ_U is reduced when the validation loss stops decreasing. The resulting weight vector:

$$\theta = \arg \min_{\theta'} \sum_{(X_j, M_j) \in \mathcal{T}} L(U_{\theta'}(X_j), M_j), \quad (7)$$

represents a parametrization of the trained network.

4.2. Network training

To create a sufficiently large training dataset, four human ground-truth analysts, designated as $GTA_{0,2,4,5}$, participated in manual annotation of imprint contours. In total, 2221 Nomarski imprint images of dimensions 6000×4000 pixels were manually processed and used for training of the U-Net network. Training imprints were created at all three wavelengths. As the size of full-power and moderately attenuated PMMA imprints may exceed the neural network input dimensions limited by the available GPU memory, image preprocessing is necessary. One option is to downsample the image which, however, leads to a loss of information. Instead, we have divided the oversized imprint into small 1024×1024 -pixel square tiles and processed them independently, as illustrated in Fig. 2. We use the fact that the most difficult image part to segment is the object boundary. The first tile is centered at one point of the imprint (object) contour extracted from the training mask. The center of the neighbouring tile lies at the intersection of the imprint contour and the edge of the previous square. The imprint contour is followed in this way until it is fully covered.

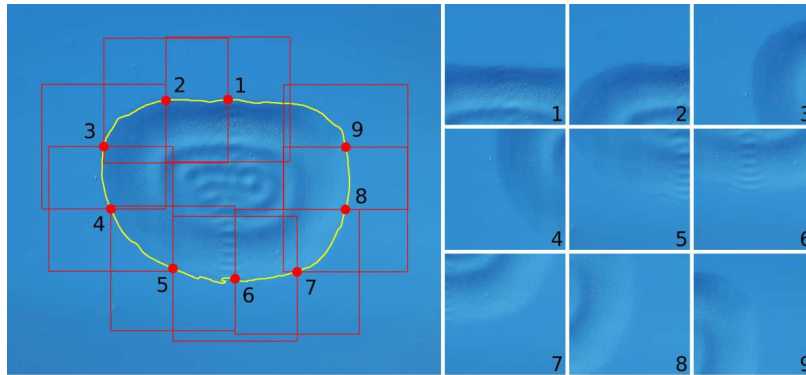


Fig. 2. Illustration of used tiling algorithm applied to large input images. Original imprint with indicated contour and cutout squares is shown on the left side. Final cutouts are on the right side.

Along-contour tiling makes it possible to extract highly resolved images composed of approximately 50% of the imprint interior area and 50% of the unaltered area. Contrary to uniform tiling, this approach is better tailored to our problem since the majority of the training images contains the threshold contour, i.e. the feature of interest. The tiling procedure extends the training dataset (up to 7652 images) and significantly improves the performance of the training process while maintaining the native image resolution. The reverse operation of aggregation of individually processed (segmented) tiles is performed by simple averaging. For small imprints that fit into a single tile, no aggregation is necessary.

The contribution of GTA_{0,2,4,5} to the training dataset was 50%, 37%, 8%, 5% of tiled images, respectively. The training dataset was split into training (89% = 6802 images) and validation (11% = 850 images) subsets. The training process consisted of 50 epochs, each epoch numbering 6802 steps. In each epoch all training images and the corresponding ground-truth binary masks were loaded with randomly varied augmentation significantly extending the training data variability. Employing the Albumentation library [43], we imposed the following set of augmentation operations: random 90° rotation, horizontal and vertical flip; random change of brightness and contrast (both up to ±15%); random change of the hue, saturation, and values (HSV) of color components (up to ±15%); elastic deformation of images. The elastic deformation was applied with probability of 25% whereas all other operations were applied with probability of 50%. To compensate the effects of color changes due to illumination and material differences, we applied contrast limited adaptive histogram equalization (CLAHE) [44].

4.3. Network benchmarking

To provide a comprehensive benchmark test of U-Net capabilities, a testing dataset $\mathcal{R}$ consisting of $r = 473$ in-focus fluence scan imprints was independently analyzed by four ground-truth analysts GTA_{0,1,2,3} and served as testing data not only for the trained network, but also to compare the annotators among themselves. GTA₁ and GTA₃ did not participate in network training. None of the testing images was used in the training process and thus remained unseen by the neural network. Since the CNN performance may depend on its parameters, 65 network models of varied numbers of descent levels $l_U \in \{3 \dots 7\}$ and encoding layers $e_U \in \{4 \dots 16\}$ were trained on the same input data and studied. The models will be further designated as UNET $\{l_U, e_U\}$. Based on recommendations in the literature [35,42] and initial experiments, we have set the input image size $N \times N = 1024 \times 1024$ (pixels), number of convolutional kernels $k_U = 16$, and Nesterov's momentum $m_U = 0.98$. Ranges of varied parameters l_U and e_U were mostly limited by the available memory capacity of the GPU.

4.4. Evaluation metrics

To quantitatively assess the performance of trained U-Net models, the measurement by GTA₀ was chosen as a reference (REF) to which all other analysts were related. Since all GTAs are almost equally experienced, the analyst who generated the most of the training data was chosen as the reference GTA. Given a trained CNN model U_θ and an input image $X_j \in \mathcal{R}$, we compute the prediction probability map $Q_j = U_\theta(X_j) \in [0, 1]$ which is of the same size as the image X_j and contains for each pixel the probability of being part of an imprint. We compare the prediction map with the corresponding reference mask M_j^{REF} using the relative area deviation (RAD):

$$\delta_j = \delta(M_j^{\text{REF}}, Q_j) = \frac{\|Q_j\| - \|M_j^{\text{REF}}\|}{\|M_j^{\text{REF}}\|}. \quad (8)$$

and Dice score:

$$D_j = D(M_j^{\text{REF}}, Q_j) = 2 \frac{\|M_j^{\text{REF}} \circ Q_j\|}{\|M_j^{\text{REF}}\| + \|Q_j\|}, \quad (9)$$

where the Hadamard operator ($\circ$) represents an element-by-element product of image matrices and the norm $\|M\| = \sum_i M(i)$ is a sum over all N^2 matrix elements (image pixels). To compare the masks M_j^{GTA} , as measured by other ground-truth analysts, we evaluate the RAD and Dice score as $\delta(M_j^{\text{REF}}, M_j^{\text{GTA}})$ and $D(M_j^{\text{REF}}, M_j^{\text{GTA}})$, respectively.

The relative area deviation (RAD) quantifies the difference between the compared and reference mask area relative to the reference area. Therefore, zero deviation indicates a perfect match of areas. A more general Dice score metric evaluates an overlap area normalized to a half of the

sum of both the compared and reference mask areas and thus expresses a similarity of mask shapes. Unity-valued Dice score thus means an identical shape of both masks implying zero RAD but the reverse statement is not necessarily true.

Averaging over the $\mathcal{R}$ dataset, we obtain the mean relative area deviation (MRAD) as: $\langle \delta \rangle = \frac{1}{r} \sum_j \delta_j$. Evaluating the mean squared RAD $\langle \delta^2 \rangle = \frac{1}{r} \sum_j \delta_j^2$, we calculate the standard deviation of the mean: $\sigma_\delta = (\langle \delta^2 \rangle - \langle \delta \rangle^2)^{\frac{1}{2}}$. In a similar way, the mean Dice score $\langle D \rangle$ and its standard deviation σ_D is evaluated.

5. Results and discussion

5.1. Network training

As described in Methods subsections 4.2 and 4.3, 65 network models of varied parameters were trained using the same data. In Fig. 3(a), three validation Nomarski images (X) of PMMA ablation imprints created at the three particular wavelengths are shown. Figure 3(b) depicts corresponding manually annotated binary masks (M). While the black color has a value of unity and represents the imprint interior, the white color equals zero and corresponds to an unaltered part of the surface. Some non-Gaussian laser beams may exhibit local minima in the fluence profile predominantly due to diffraction effects and typically out of focus. If below the threshold, imprints may contain islands of unaltered surface imposing holes in the mask interior, as demonstrated in Fig. 3 at the wavelength of 18 nm. Figure 3(c) displays grayscale maps ($Q = U_\theta(X) \in [0, 1]$) recovered by UNET{7,8} model where each pixel is assigned a probability of occurrence in the imprint interior. The resulting prediction maps are in an excellent agreement with ground-truth validation masks indicating a good accuracy of this model. As shown in Fig. 3(a), the intensity, color, and shape of the imprint is quite variable depending on the microscope setting, photon and pulse energy. For example, as the photon energy approaches the carbon K-edge, the ablated surface in the imprint created at wavelength of 8 nm becomes considerably rougher due to an increased PMMA absorption length. On the contrary, the imprint created at wavelength of 13.5 nm is smoother but thermally molten inside. With respect to this variability, the U-Net network is capable of performing quite robustly.

5.2. Network benchmarking

To benchmark the performance of all studied U-Net models as well as other GTAs quantitatively, we used an independent testing dataset $\mathcal{R}$ consisting of $r = 473$ fluence scan imprints created at fixed in-focus position and wavelength of 18 nm. Testing images were used neither as training nor as validation data. Since the majority of training data was provided by GTA₀, the testing data processed by this analyst served as a reference measurement to which all other analysts were related via similarity metrics introduced in Methods subsection 4.4.

Figures 4(a) and 4(b) display mean relative area deviations and Dice scores related to GTA₀ reference for all ground-truth analysts GTA_{0,1,2,3} and U-Net models UNET{ l_U, e_U }. For the purposes of completeness, zero MRAD and the unity-valued mean Dice score of GTA₀ to himself is also shown. Mean values $\langle \delta \rangle$ and $\langle D \rangle$ are indicated as color boxes assigned to each particular ground-truth analyst or U-Net model. As evident from inset bar plots, there is a clear trend in both metrics indicating an improving performance of U-Net predictions with the number of descent levels l_U . While the MRAD tends to zero, the mean Dice score gradually approaches unity as l_U increases. On the other hand, the dependence on the number of encoding layers e_U does not show any significant trend. Furthermore, MRAD closer to zero does not necessarily imply a better mean Dice score neither for U-Net models nor for GTAs. Hence the choice of the best model is primarily done with regard to a more general Dice score. MRAD is applied as a secondary criterion. Values of both metrics evaluated for all GTAs and the best three U-Net models are sorted by Dice score in Table 1. Within standard deviations, all values are consistently similar for

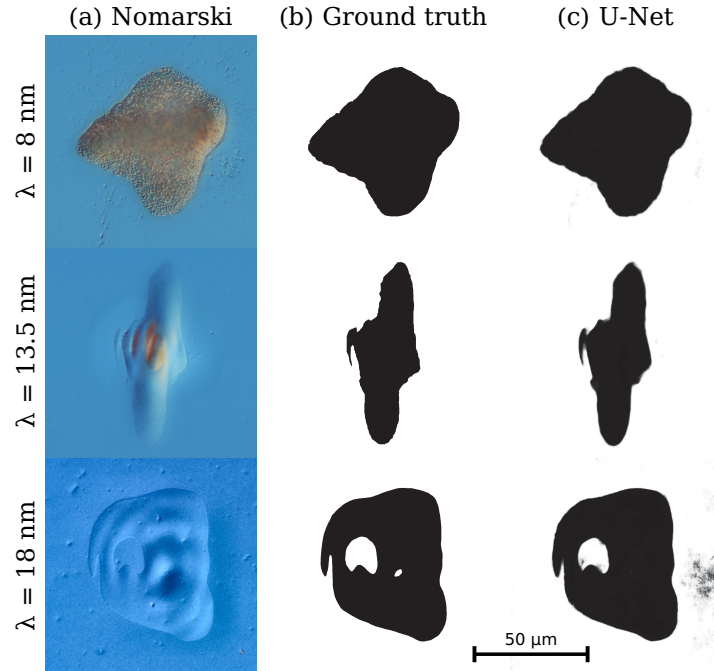


Fig. 3. (a) Nomarski images of ablative imprints in PMMA created by a focused beam of FLASH2 tuned at 8 nm (top row), 13.5 nm (middle row), and 18 nm (bottom row). (b) Manually annotated ground-truth images represented as binary masks (black – imprint interior, white – imprint exterior). (c) Probability maps as recovered by the UNET{7, 8} model. All images are in scale.

all models and analysts compared. To discriminate between models, the values are shown with a higher 4-digit accuracy. Albeit the UNET{6, 15} model wins in the sense of the mean Dice score, the second model UNET{7, 8} is only insignificantly worse within one standard deviation. However, UNET{7, 8} performs much better in terms of MRAD making it the model of best choice if compared to GTA_0 .

Table 1. Results of ground-truth analysts and three best performing U-Net models and comparison to reference GTA_0 . A_{eff} and E_{th} denote the resultant effective area and extrapolated threshold pulse energy and $\langle D \rangle$ and $\langle \delta \rangle$ represent the mean Dice score and relative area deviation related to the reference analyst, respectively.

Analyst	$A_{\text{eff}} [\mu\text{m}^2]$	$E_{\text{th}} [\text{nJ}]$	$\langle D \rangle \pm \sigma_D [1]$	$\langle \delta \rangle \pm \sigma_\delta [\%]$
GTA_0	95.2 ± 6.1	85.5 ± 5.8	1.000 ± 0.000	0.0 ± 0.0
UNET{6, 15}	96.3 ± 6.4	87.3 ± 6.0	0.9790 ± 0.0090	-0.84 ± 1.87
UNET{7, 8}	98.3 ± 6.5	89.2 ± 6.0	0.9789 ± 0.0101	-0.41 ± 2.31
UNET{7, 5}	97.5 ± 6.7	88.1 ± 6.1	0.9786 ± 0.0091	-0.67 ± 1.91
GTA_2	101.3 ± 6.0	90.7 ± 5.5	0.977 ± 0.008	0.9 ± 1.9
GTA_3	105.3 ± 6.2	94.0 ± 5.5	0.975 ± 0.008	1.0 ± 2.0
GTA_1	102.6 ± 5.8	90.8 ± 5.1	0.974 ± 0.011	-0.3 ± 2.3

A graphical comparison of the UNET{7, 8} model prediction maps with reference (GTA_0) masks is presented in Fig. 5(b) for 5 different pulse energies ranging from full power down to

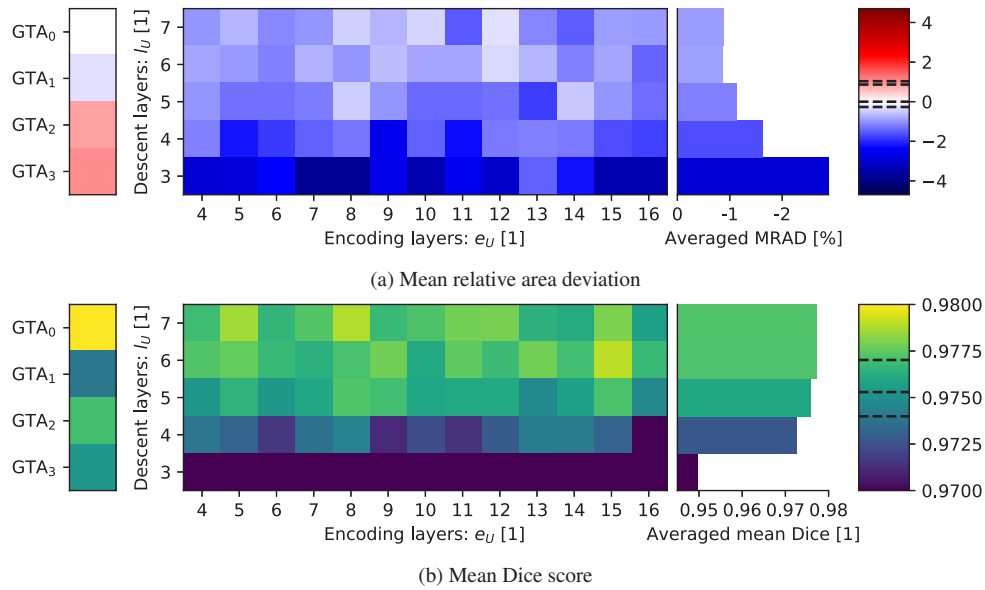


Fig. 4. Mean relative area deviation in percentage (a) and Dice score (b) calculated for 65 U-Net models. Colorbars on the right side assign a color to the corresponding value of mean relative area deviation and Dice score. Bars on the left side compare results provided by individual GTAs whose values are also indicated as dashed lines in colorbars. Horizontal bars next to the table show values averaged over encoding layers for a given descent level. To better adapt the color scale, colors of some data in the Dice score table (all models with $l_U = 3$ and GTA_0) are clipped.

nearly threshold pulse energy. Corresponding Nomarski input images are shown in Fig. 5(a). Comparative maps depict a difference between the reference ground truth and prediction in a shaded color representation indicated by the inset color wheel. While blue and green colors correspond to an interior of the reference mask, red and white colors represent its exterior. The entire color map describes four possible overlap/mismatch combinations of the reference mask and prediction map. Denoting the value in the i -th pixel as $M(i) \in \{0, 1\}$ for the reference binary mask and $Q(i) \in [0, 1]$ for the continuous prediction map, we can distinguish four different cases:

- TRUE POSITIVE: imprint interior correctly predicted inside the mask ($M(i) = 1$ and $Q(i) \geq 0.5$) indicated by the green-to-blue transition
- TRUE NEGATIVE: imprint exterior correctly predicted outside the mask ($M(i) = 0$ and $Q(i) < 0.5$) indicated by the white-to-red transition
- FALSE POSITIVE: imprint interior incorrectly predicted outside the mask ($M(i) = 0$ and $Q(i) \geq 0.5$) indicated by the red-to-white transition
- FALSE NEGATIVE: imprint exterior incorrectly predicted inside the mask ($M(i) = 1$ and $Q(i) < 0.5$) indicated by the blue-to-green transition

Apart from subtle deviations indicated by blueish and reddish colors, the U-Net prediction is very accurate in a broad range of imprint sizes. To qualitatively compare the U-Net prediction with another human analyst, a difference map between GTA_1 and the reference GTA_0 is shown in Fig. 5(c). Regardless of the fact that the difference between the two GTAs acquires only discrete values indicated by white, red, green, and blue colors, segmentations of both the GTA_1 and

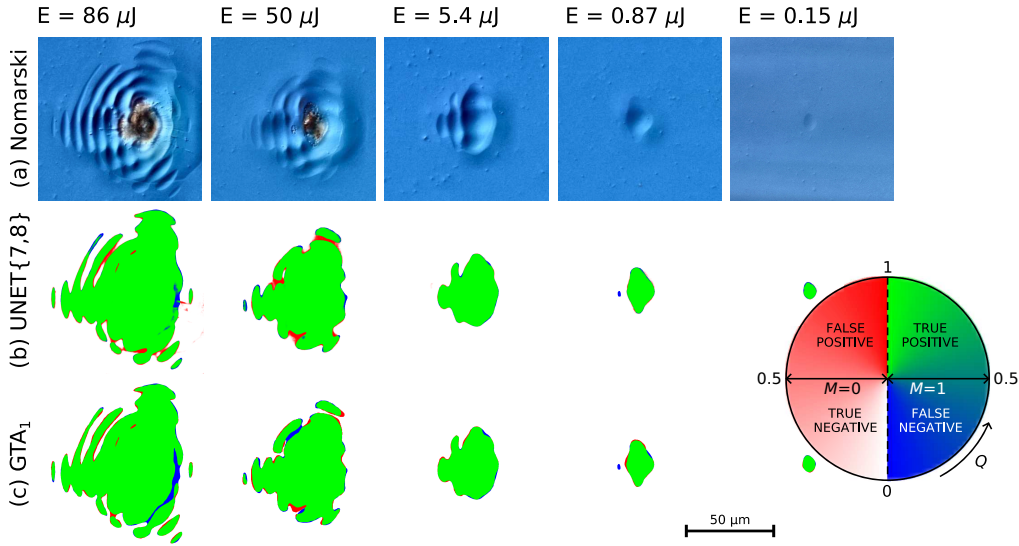


Fig. 5. (a) Nomarski testing images. (b) colored maps depicting a difference between reference GTA_0 's binary masks and probability maps recovered by UNET{7, 8} model. (c) difference between GTA_0 and GTA_1 . Ablative imprints were created at varied pulse energies indicated above images. The last image was acquired with microscope objective 100 $\times$, while the others with 50 $\times$. All images are to scale.

U-Net are of comparable quality. Furthermore, the graphical comparison of two GTAs clearly demonstrates the subjectivity of manual contour annotation.

5.3. Fluence scan evaluation

Retrieved probability maps $Q = U_\theta(X)$ can be processed by various means to obtain threshold contour areas of imprints. The most convenient way is to measure the contour area at a given probability threshold, e.g. 0.5. However, more rigorous (physical) contour area evaluation requires a statistical treatment enabling an uncertainty evaluation as described and mathematically proven in Appendix A. In Fig. 8 the two methods of contour area evaluation are compared with the ground-truth measurement. The mean and variance of the threshold contour area is expressed by Eqs. (10) and (11), respectively. In the following, the physical definition of the contour area/uncertainty will be used.

Testing data analysed by all ground-truth analysts and best three U-Net models were processed using the fluence scan evaluation method introduced in Theory 2 and described in detail in Appendix B. Measured areas S_j were paired with the corresponding pulse energies E_j and the resulting sequence $\{S_j, E_j\}_{j=1}^r$, numbering $r = 473$ data points, was sorted in an ascending order with respect to areas. Since the fluence scan was carried out at a fixed in-focus position, the z -dependence is omitted. In order to retrieve the normalized fluence scan curve and evaluate the effective area of the laser beam, the threshold pulse energy must be determined first. For this purpose we construct the so-called Liu's curve $\{\ln(E_j), S_j\}_{j=1}^r$ [45] as depicted by black open circles in Fig. 6. Areas plotted in Fig. 6 were measured by the ground-truth analyst GTA_0 . As illustrated by the red solid line, the threshold pulse energy E_{th} , corresponding to zero contour area, is evaluated using a linear extrapolation of the first $n = 50$ data points in the linear part of the Liu's curve (red open circles).

Extrapolation was performed independently for each analyst using a robust line fitting (Deming regression) procedure outlined in Eqs. (12)-(15) in Appendix B. Contrary to the standard

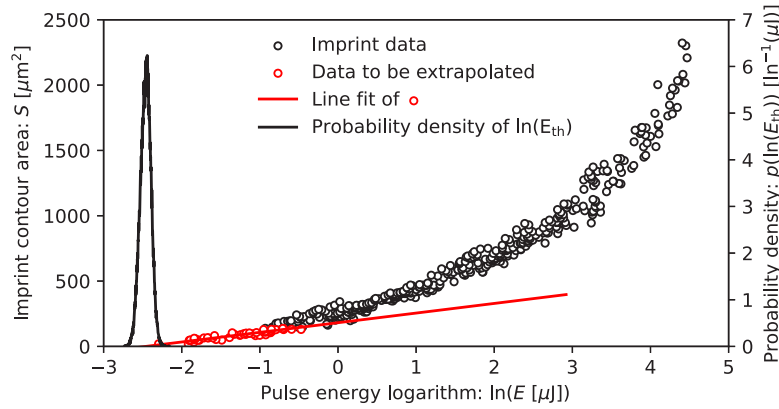


Fig. 6. Liu's plot of imprint data analysed by GTA₀. Red open circles represent a data selection to be processed using a line fit (red solid line). Black solid curve stands for the probability density function of the extrapolated threshold pulse energy logarithm.

method of linear regression, the line fitting procedure performs better on datasets in which both the dependent and independent variable evince an increased measurement error. As the method minimizes orthogonal distance between the line and data points, it is less prone to threshold underestimation. The mean value and uncertainty of the threshold pulse energy was evaluated using an iterative statistical approach of interpenetrating samples [46]. In total, 100000 combinations, containing a half of data points selected for fitting ($\lfloor n/2 \rfloor = 25$), were randomly chosen and independently fitted (extrapolated) to obtain statistics of fitting and derived parameters. In Fig. 6 a histogram of the threshold pulse energy logarithm $\ln(E_{th})$ is shown as a black solid curve the width of which identifies an uncertainty of the threshold value. Due to the nonlinearity of the Liu's curve, the number of extrapolated points n must be identified individually for each analyst. For this purpose we look for a minimum of the mean squared normal difference (fit goodness metric defined in Eq. (16) in Appendix B) plotted as a function of the data subset size $n \in \{4 \dots r\}$. As demonstrated in Fig. 9 in Appendix B, the optimum subset size of GTA₀'s data to be extrapolated is represented by first 50 points of the Liu's curve. All details of the threshold extrapolation procedure are summarized in Appendix B and the resulting values are listed in Table 1.

Following the definition in Eq. (4), the threshold pulse energy value serves as a normalization factor for the fluence scan (sequence) curve $\{S_j, E_{th}/E_j\}_{j=1}^r$ depicted in Fig. 7. Evidently, curves derived by independent analysts are in an excellent agreement. It follows from Eq. (3) that the numerical integral (see Eq. (17) in Appendix C) below the f -scan curve represents the effective area of the beam. The uncertainty of this value was evaluated using the error propagation [47] combining uncertainties of the threshold value E_{th} , pulse energies E_j , and contour areas S_j , as outlined in Eq. (18) in Appendix C. For each particular (j -th) imprint we assume a relative area error determined via standard deviation of the mean contour area measured by all GTAs. While the uncertainty of small close-to-threshold imprints can reach 10%, the relative area error of large imprints drops to 1% or below and the average over all imprints is 1.5%. Uncertainty of area prediction by U-Net was determined via the physical treatment of prediction masks described in Appendix A. In the sense of the relative area error, it performs similarly as for GTAs and the averaged value is 1.9%. To estimate an uncertainty of the pulse energy measurement, we employ the fact that ablative imprints of similar areas should have been created by pulses of similar energy. Hence evaluating the standard pulse energy deviation for imprints of area within the 1.5%-uncertainty bins (similar imprints), we get an average relative pulse energy error of

approx. 13%. Since the contribution of the contour area and pulse energy uncertainty to the relative effective area error decreases approximately with $\sqrt{r}$, the total error is dominated by the uncertainty of the threshold pulse energy E_{th} . Therefore, the reliability of this value is crucial for effective area evaluation.

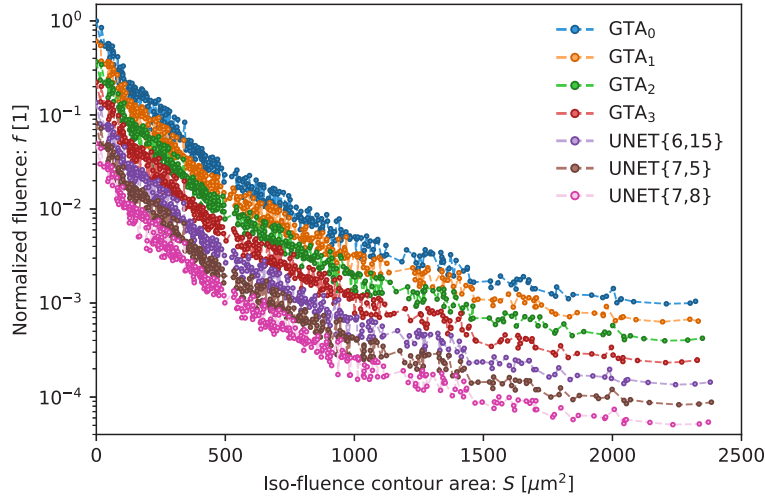


Fig. 7. Fluence scan curves derived from measurements carried out by ground-truth analysts and best performing U-Net models. The curves are offset for clarity.

Values of resulting effective areas are listed in Table 1. All values, except for GTA_3 , are in a very good agreement within the error margin. The uncertainty of manual image segmentation is mostly determined by capabilities and current condition of each individual human analyst. Color perception and contour identification thus may be strongly subjective which in turn leads to an increased variability of results. Especially the threshold pulse energy extrapolation relies on the capability of an analyst to annotate the smallest imprints. Albeit all U-Net models slightly underestimate the prediction in terms of the imprint contour area ($\langle\delta\rangle \leq 0$), MRAD is gradually approaching the reference GTA_0 . We may expect that deeper networks could perform even better. Furthermore, the mean Dice score clearly surpasses other GTAs. U-Net results tend to GTA_0 also in terms of the resultant effective area. This might indicate that the contribution of GTA_0 possibly dominated the training process making the U-Net network biased in favor of this analyst. To prove this, an additional thorough testing of the network performance trained using datasets of varied compositions is necessary. Moreover, in order to independently evaluate the performance of both the U-Net models and human annotators, a precise standardized measurement of imprint contours, e.g. using the atomic force microscopy, is required. This is, however, an extensive topic for another systematic study which would deserve a separate article.

6. Conclusions

We applied the U-Net convolutional neural segmentation network to microscopy images of ablative imprints in PMMA in order to extract ablation threshold contours and their areas. We performed a comprehensive benchmark test by comparing its results with manual annotations carried out by four independent human analysts. We may conclude that the performance of a properly trained network can be distinguished from an experienced human analyst neither qualitatively nor quantitatively within one standard deviation of used similarity metrics. In some aspects, the network may perform even better depending on the size and preprocessing of the training data and available computational capacity. Results of the image segmentation

analysis of all human analysts and three best-performing U-Net models were processed using a well-established fluence scan method to evaluate the effective area and fluence scan curve of a focused beam of the FLASH2 facility tuned at 18 nm. Results are in an excellent agreement within the evaluated error margin. All preprocessing and postprocessing routines reported in this Paper were designed such that the ablation imprint analysis can be carried out automatically. This makes it possible to create a virtual analyst capable to autonomously conduct laborious ablation imprint data evaluation from start to end.

A substitution of a human analyst will enable an efficient processing of much larger ensembles of ablation imprint data making it possible to characterize beam conditions of more complex experiments. Furthermore, autonomously processed imprint data will make it possible to fully employ the $F(z)$ -scan method outlined in Eq. (5). This opens up avenues not only to better characterize the focus and its vicinity, but also to compare other semi-direct and indirect beam characterization methods with a direct in-focus measurement.

Appendix A: imprint area evaluation

Here we describe two approaches how to calculate the imprint area from the predicted probability map $Q \in [0, 1]$. Both methods assume that there is only one imprint (laser shot) in one image, i.e. that all foreground pixels belong to the imprint. The two-dimensional $N \times N$ matrix $Q = U_\theta(X)$, representing the predicted probability map, is flattened (vectorized) into a one-dimensional vector (of length N^2) and sorted in descending order. We interpret the resulting sequence P_n as the probability that the object area contains at least n pixels. Equivalently, P_n also represents the probability that an iso-fluence contour of the area $S = n\Delta S$ is a subset of the threshold contour. Here ΔS is an area occupied by a single image pixel. We can therefore write P as a discrete function of S as $P(S) = P(n\Delta S) = P_n$. Figure 8 shows such probability functions for a ground-truth measurement $P_M(S)$ (black solid line) and U-Net prediction $P_Q(S)$ (blue solid line).

A simple approach to imprint area estimation is to find such S so that $P_Q(S) = 0.5$. A value enumerated for an example imprint is shown by the blue dashed line in Fig. 8. The corresponding area is $1280 \mu\text{m}^2$. Alternatively, in order to take into account the uncertainty of the segmentation, we can consider the numerical derivative of $P_Q(S)$, i.e. the probability density function $p_Q(S) = -dP_Q(S)/dS$, shown as an orange solid curve in Fig. 8. The minus sign corresponds to our choice of sorting the Q values in a descending order. From $p_Q(S)$ we can calculate its mean value $\langle S \rangle$ (first order moment). It turns out that it is simply the sum over all elements $\|Q\| = \sum_i Q(i)$ of the prediction probability map. This can be shown using the integration by parts:

$$\langle S \rangle = \int_0^{S_{\max}} S p_Q(S) dS = \int_0^{S_{\max}} P_Q(S) dS \approx \Delta S \|Q\| = \Delta S \sum_{i \in \Omega} Q(i). \quad (10)$$

Here the upper integration limit $S_{\max}$ is the area of the processed image and Ω denotes the set of image pixels. It follows from Eq. (10) that the mean $\langle S \rangle$ is equal to the area below the cumulative function $P_Q(S)$ or to the “volume” below the predicted probability map Q . The variance σ_S^2 (central second order moment) is defined by an integral:

$$\sigma_S^2 = \int_0^{S_{\max}} (S - \langle S \rangle)^2 p_Q(S) dS, \quad (11)$$

which can be evaluated numerically, e.g. by the trapezoidal rule. For this purpose, p_Q can be calculated, for example, by taking first order finite differences of P_Q . Alternatively, we can apply per partes integration again. In Fig. 8 the mean value $\langle S \rangle$ and the confidence interval $(\langle S \rangle \pm \sigma_S)$ are depicted as orange dashed line and a semitransparent orange box, respectively. The resulting area is $(1277 \pm 39) \mu\text{m}^2$ which is in a very good agreement with the ground-truth area $1264 \mu\text{m}^2$ below the black curve in Fig. 8.

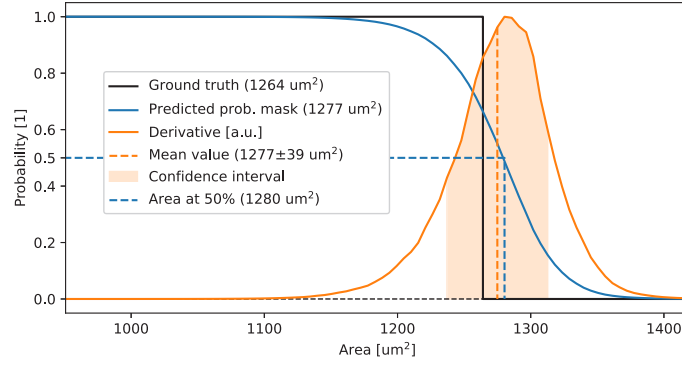


Fig. 8. A comparison of different approaches to imprint area evaluation. The plot depicts the sorted ground-truth binary function $P_M(S)$ (black solid line), predicted probability function $P_Q(S)$ and its normalized numerical derivative $p_Q(S)$ (blue and orange solid lines), the corresponding mean value $\langle S \rangle$ and confidence interval (orange dashed line and semitransparent orange box), and predicted binary mask area at the probability level of 50% (blue dashed line).

Appendix B: threshold extrapolation

To extrapolate the threshold pulse energy, a generalized and more robust approach of the line fit (Deming regression) is used. Contrary to the standard method, the generalized procedure fits the data with a line in an implicit form $ax + by + c = 0$ by minimizing the squared orthogonal distance of fitted data points to that line. Given a set of n coordinate pairs $\{x_j, y_j\}_{j=1}^n$, an averaged sum of squared differences χ^2 can be defined as:

$$\chi^2(a, b, c) = \frac{1}{n} \sum_{j=1}^n (ax_j + by_j + c)^2. \quad (12)$$

Here the fitting parameters a, b are components of a unit vector normal to the fitted line satisfying a condition $a^2 + b^2 = 1$ and c is a negated normal distance between the line and origin. The minimum of the sum of squared normal differences occurs for parameters:

$$a = \frac{\sigma_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \quad (13)$$

$$b = -\frac{\sigma_x \text{sgn}(\sigma_{xy})}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \quad (14)$$

$$c = -a\langle x \rangle - b\langle y \rangle. \quad (15)$$

Here $\langle x \rangle, \langle y \rangle$ are mean values of the variables, $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$ and $\sigma_y^2 = \langle y^2 \rangle - \langle y \rangle^2$ are the corresponding variances and $\sigma_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$ is the covariance. Using the a, b, c parameters, the fitted line can be expressed in an explicit form $y = px + q$ where the slope and y-intercept read $p = -a/b$ and $q = -c/b$, respectively. More importantly, the x-intercept, i.e. the extrapolated threshold, is $x_{\text{th}} = -c/a$. Hence, applying this approach to the logarithmic Liu's sequence, we get $E_{\text{th}} = \exp(-c/a)$.

An uncertainty of the fit is estimated using an iterative statistical approach of random choice. From the set of first $n = 50$ data points (the value to be justified below) a subset of $\lfloor n/2 \rfloor = 25$ points is randomly selected and fitted. In total, 100000 iterations were performed to obtain a good

statistics of derived parameters and to evaluate their means and standard deviations. Such an approach is less sensitive to occasional outlier points occurring in data. The resulting values are: $a = (0.9999074 \pm 0.0000094) \ln^{-1}(\mu\text{J})$, $b = (-0.01359 \pm 0.00069) \mu\text{m}^{-2}$, $c = (2.462 \pm 0.067)$, $p = (73.8 \pm 3.7) \mu\text{m}^2 \ln^{-1}(\mu\text{J})$, $q = (181.4 \pm 4.8) \mu\text{m}^2$, and $E_{\text{th}} = (0.0855 \pm 0.0058) \mu\text{J}$.

Due to the nonlinearity of the Liu's curve, the result of the fit and extrapolation may depend on the size n of the fitted dataset, i.e. the number of first points in the Liu's curve subject to extrapolation. In order to find the optimum n , we repeat the above mentioned approach for varied subset sizes gradually increased in range $n \in \{4 \dots r\}$. In each step the minimum mean squared normal difference, an indicator of the fit goodness, is evaluated as:

$$\chi_{\min}^2 = a^2 \langle x^2 \rangle + b^2 \langle y^2 \rangle + c^2 + 2ab \langle xy \rangle + 2ac \langle x \rangle + 2bc \langle y \rangle, \quad (16)$$

where the angle brackets denote averaging over all n subset points and the parameters a, b, c are the fitting parameters evaluated using Eqs. (13)–(15). In total 5000 iterations were carried out for each subset size while performing the fits on $\lfloor n/2 \rfloor$ points. The minimum subset size is $n = 4$ since the linear fit requires at least 2 points. Figure 9 shows the calculated minimum χ -squared averaged over all (5000) iterations as a function of n . The curve exhibits a clear minimum at $n = 50$ identifying the first linear part of the Liu's curve which can be best fitted/extrapolated by a line.

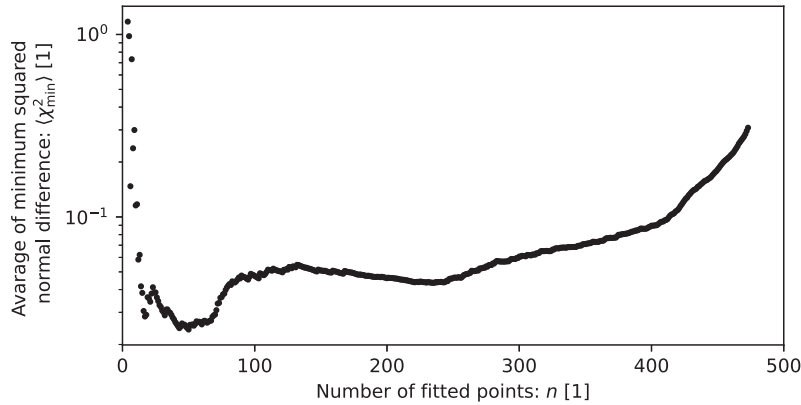


Fig. 9. Averaged minimum squared normal difference as a function of the fitted data subset size.

Appendix C: effective area

The effective area is numerically calculated from the normalized fluence scan (curve) sequence $\{S_j, E_{\text{th}}/E_j\}_{j=1}^r$ sorted in an ascending order with respect to contour areas S_j . It follows from the normalization that the point $\{0, 1\}$ can be artificially added to the sequence. The effective area of the beam is calculated by the numerical integration of the f -scan curve using, for example, the trapezoidal rule:

$$A_{\text{eff}} = \frac{1}{2} E_{\text{th}} \sum_{j=0}^{r-1} \left(\frac{1}{E_j} + \frac{1}{E_{j+1}} \right) (S_{j+1} - S_j), \quad (17)$$

where the zeroth point was added to the sequence using values $E_0 = E_{\text{th}}$ and $S_0 = 0 \mu\text{m}^2$. The uncertainty of the effective area ΔA_{eff} can be evaluated using the error propagation method [47]

combining errors of the threshold pulse energy ΔE_{th} , areas ΔS_j , and pulse energies ΔE_j as:

$$\Delta A_{\text{eff}} = \left\{ \left(\frac{\partial A_{\text{eff}}}{\partial E_{\text{th}}} \right)^2 (\Delta E_{\text{th}})^2 + \sum_{j=1}^r \left[\left(\frac{\partial A_{\text{eff}}}{\partial E_j} \right)^2 (\Delta E_j)^2 + \left(\frac{\partial A_{\text{eff}}}{\partial S_j} \right)^2 (\Delta S_j)^2 \right] \right\}^{\frac{1}{2}}. \quad (18)$$

Here we use the fact that the zeroth point of the f -scan sequence is fixed and therefore its uncertainty is zero.

Funding. Grantová Agentura České Republiky (20-08452S); Horizon 2020 Framework Programme (Grant Agreement No 730 872, VOXEL H2020-FETOPEN-2014- 2015-RIA 665 207); Fundação para a Ciência e a Tecnologia (IC&DT—AAC n. o 02/SAICT/2017—X-ELS 31 868, PD/BD/105879/2014 (PD-F APPLAuSE)).

Acknowledgement. J. Ch. developed the method of ablation imprints and statistical methods of data analysis. V. V. performed numerical experiments, network training and benchmarking. J. H., J. K., and B. P. designed and created the U-Net code for ablation imprints segmentation. Š. J., Z. K., V. H., K. F., V. V., and L. V. manually annotated ablation imprints as ground-truth analysts GTA_{0,1,2,3,4,5}, respectively. J. Ch., V. V., T. B., S. D., V. H., L. J., B. K., M. K., M. R. -L., L. V., T. W., and E. P. carried out the experiment at FL24/FLASH2. J. Ch., V. V., and J. K. wrote the manuscript. The authors acknowledge the FLASH2 facility for beamtime.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. H. N. Chapman, S. P. Hau-Riege, and M. J. Bogan, *et al.*, “Femtosecond time-delay x-ray holography,” *Nature* **448**(7154), 676–679 (2007).
2. K. Kharitonov, M. Mehrjoo, M. Ruiz-Lopez, B. Keitel, S. Kreis, M. Seyrich, M. Pop, and E. Plonjes, “Flexible ptychography platform to expand the potential of imaging at free electron lasers,” *Opt. Express* **29**(14), 22345–22365 (2021).
3. I. Inoue, Y. Deguchi, B. Ziaja, T. Osaka, M. M. Abdullah, Z. Jurek, N. Medvedev, V. Tkachenko, Y. Inubushi, H. Kasai, K. Tamasaku, T. Hara, E. Nishibori, and M. Yabashi, “Atomic-scale visualization of ultrafast bond breaking in x-ray-excited diamond,” *Phys. Rev. Lett.* **126**(11), 117403 (2021).
4. A. Sorokin, S. Bobashev, T. Feigl, K. Tiedtke, H. Wabnitz, and M. Richter, “Photoelectric effect at ultrahigh intensities,” *Phys. Rev. Lett.* **99**(21), 213002 (2007).
5. B. Nagler, U. Zastra, and R. R. Fäustlin, *et al.*, “Turning solid aluminium transparent by intense soft x-ray photoionization,” *Nat. Phys.* **5**(9), 693–696 (2009).
6. S. Toleikis, R. R. Fäustlin, and L. Cao, *et al.*, “Soft X-ray scattering using FEL radiation for probing near-solid density plasmas at few electron volt temperatures,” *High Energy Density Phys.* **6**(1), 15–20 (2010).
7. S. M. Vinko, O. Ciricosta, and B. I. Cho, *et al.*, “Creation and diagnosis of a solid-density plasma with an x-ray free-electron laser,” *Nature* **482**(7383), 59–62 (2012).
8. O. Ciricosta, S. M. Vinko, and B. Barbrel, *et al.*, “Measurements of continuum lowering in solid-density plasmas created from elements and compounds,” *Nat. Commun.* **7**(1), 11713 (2016).
9. S. M. Vinko, V. Vozda, and J. Andreasson, *et al.*, “Time-resolved xuv opacity measurements of warm dense aluminum,” *Phys. Rev. Lett.* **124**(22), 225002 (2020).
10. J. Rossbach, J. R. Schneider, and W. Wurth, “10 years of pioneering x-ray science at the free-electron laser FLASH at DESY,” *Phys. Rep.* **808**, 1–74 (2019).
11. C. Bostedt, S. Boutet, D. M. Fritz, Z. Huang, H. J. Lee, H. T. Lemke, A. Robert, W. F. Schlotter, J. J. Turner, and G. J. Williams, “Linac coherent light source: The first five years,” *Rev. Mod. Phys.* **88**(1), 015007 (2016).
12. T. Tschentscher, C. Bressler, J. Grünert, A. Madsen, A. Mancuso, M. Meyer, A. Scherz, H. Sinn, and U. Zastra, “Photon beam transport and scientific instruments at the european XFEL,” *Appl. Sci.* **7**(6), 592 (2017).
13. R. W. Schoenlein, S. Boutet, M. P. Minitti, and A. M. Dunne, “The linac coherent light source: Recent developments and future plans,” *Appl. Sci.* **7**(8), 850 (2017).
14. K. Tiedtke, A. Azima, and N. von Barga, *et al.*, “The soft x-ray free-electron laser FLASH at DESY: beamlines, diagnostics and end-stations,” *New J. Phys.* **11**(2), 023029 (2009).
15. A. A. Sorokin, Y. Bican, S. Bonfigt, M. Brachmanski, M. Braune, U. F. Jastrow, A. Gottwald, H. Kaser, M. Richter, and K. Tiedtke, “An x-ray gas monitor for free-electron lasers,” *J. Synchrotron Radiat.* **26**(4), 1092–1100 (2019).
16. S. L. Pape, P. Zeitoun, M. Idir, P. Dhez, J. J. Rocca, and M. François, “Electromagnetic-field distribution measurements in the soft x-ray range: Full characterization of a soft x-ray laser beam,” *Phys. Rev. Lett.* **88**(18), 183901 (2002).

17. B. Keitel, E. Plönjes, S. Kreis, M. Kuhlmann, K. Tiedtke, T. Mey, B. Schäfer, and K. Mann, "Hartmann wavefront sensors and their application at FLASH," *J. Synchrotron Radiat.* **23**(1), 43–49 (2016).
18. H. M. Quiney, A. G. Peele, Z. Cai, D. Paterson, and K. A. Nugent, "Diffractive imaging of highly focused x-ray fields," *Nat. Phys.* **2**(2), 101–104 (2006).
19. N. D. Loh, D. Starodub, and L. Lomb, *et al.*, "Sensing the wavefront of x-ray free-electron lasers using aerosol spheres," *Opt. Express* **21**(10), 12385 (2013).
20. A. Schropp, R. Hoppe, V. Meier, J. Patommel, F. Seiboth, H. J. Lee, B. Nagler, E. C. Galtier, B. Arnold, U. Zastra, J. B. Hastings, D. Nilsson, F. Uhlén, U. Vogt, H. M. Hertz, and C. G. Schroer, "Full spatial characterization of a nanofocused x-ray free-electron laser beam by ptychographic imaging," *Sci. Rep.* **3**(1), 1633 (2013).
21. B. Nagler, A. Aquila, S. Boutet, E. C. Galtier, A. Hashim, M. S. Hunter, M. Liang, A. E. Sakdinawat, C. G. Schroer, A. Schropp, M. H. Seaberg, F. Seiboth, T. van Driel, Z. Xing, Y. Liu, and H. J. Lee, "Focal spot and wavefront sensing of an x-ray free electron laser using ronchi shearing interferometry," *Sci. Rep.* **7**(1), 13698 (2017).
22. J. Chalupský, P. Boháček, T. Burian, V. Hájková, S. Hau-Riege, P. Heimann, L. Juha, M. Messerschmidt, S. Moeller, B. Nagler, M. Rowen, W. Schlotter, M. Swiggers, J. Turner, and J. Krzywinski, "Imprinting a focused x-ray laser beam to measure its full spatial characteristics," *Phys. Rev. Appl.* **4**(1), 014004 (2015).
23. H. Yumoto, H. Mimura, T. Koyama, S. Matsuyama, K. Tono, T. Togashi, Y. Inubushi, T. Sato, T. Tanaka, T. Kimura, H. Yokoyama, J. Kim, Y. Sano, Y. Hachisu, M. Yabashi, H. Ohashi, H. Ohmori, T. Ishikawa, and K. Yamauchi, "Focusing of x-ray free-electron laser pulses with reflective optics," *Nat. Photonics* **7**(1), 43–47 (2012).
24. M. Schneider, C. M. Günther, B. Pfau, F. Capotondi, M. Manfredda, M. Zangrando, N. Mahne, L. Raimondi, E. Pedersoli, D. Naumenko, and S. Eisebitt, "In situ single-shot diffractive fluence mapping for x-ray free-electron laser pulses," *Nat. Commun.* **9**(1), 214 (2018).
25. T. A. Pikuz, A. Y. Faenov, Y. Fukuda, M. Kando, P. Bolton, A. Mitrofanov, A. V. Vinogradov, M. Nagasono, H. Ohashi, M. Yabashi, K. Tono, Y. Senba, T. Togashi, and T. Ishikawa, "Soft x-ray free-electron laser imaging by LiF crystal and film detectors over a wide range of fluences," *Appl. Opt.* **52**(3), 509 (2013).
26. J. Chalupský, J. Krzywinski, L. Juha, V. Hájková, J. Cihelka, T. Burian, L. Vyšín, J. Gaudin, A. Gleeson, M. Jurek, A. R. Khorsand, D. Klinger, H. Wabnitz, R. Sobierajski, M. Störmer, K. Tiedtke, and S. Toleikis, "Spot size characterization of focused non-gaussian x-ray laser beams," *Opt. Express* **18**(26), 27836–27845 (2010).
27. J. Chalupský, P. Boháček, V. Hájková, S. Hau-Riege, P. Heimann, L. Juha, J. Krzywinski, M. Messerschmidt, S. Moeller, B. Nagler, M. Rowen, W. Schlotter, M. Swiggers, and J. Turner, "Comparing different approaches to characterization of focused x-ray laser beams," *Nucl. Instrum. Methods Phys. Res., Sect. A* **631**(1), 130–133 (2011).
28. J. Chalupský, T. Burian, V. Hájková, L. Juha, T. Polcar, J. Gaudin, M. Nagasono, R. Sobierajski, M. Yabashi, and J. Krzywinski, "Fluence scan: an unexplored property of a laser beam," *Opt. Express* **21**(22), 26363 (2013).
29. B. Rösner, F. Döring, P. R. Ribič, D. Gauthier, E. Principi, C. Masciovecchio, M. Zangrando, J. Vila-Comamala, G. D. Ninno, and C. David, "High resolution beam profiling of X-ray free electron laser radiation by polymer imprint development," *Opt. Express* **25**(24), 30686 (2017).
30. N. Gerasimova, S. Dziarzhyski, H. Weigelt, J. Chalupský, V. Hájková, L. Vyšín, and L. Juha, "In situ focus characterization by ablation technique to enable optics alignment at an XUV FEL source," *Rev. Sci. Instrum.* **84**(6), 065104 (2013).
31. Y. Rivenson, Z. Gorocs, H. Gunaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
32. D. Piccinotti, K. F. MacDonald, S. Gregory, I. Youngs, N. Zheludev, and I. , "Artificial intelligence for photonics and photonic materials," *Rep. Prog. Phys.* **84**(1), 012401 (2021).
33. Y. Zhang, M. A. Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, and P. Villanueva-Perez, "Phasegan: a deep-learning phase-retrieval approach for unpaired datasets," *Opt. Express* **29**(13), 19593–19604 (2021).
34. D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, vol. 25 F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., (Curran Associates, Inc., 2012).
35. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, (Springer International Publishing, 2015), pp. 234–241.
36. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).
37. J. Chalupský, L. Juha, and V. Hájková, *et al.*, "Non-thermal desorption/ablation of molecular solids induced by ultra-short soft x-ray pulses," *Opt. Express* **17**(1), 208 (2009).
38. N. Medvedev, J. Chalupsky, and L. Juha, "Microscopic kinetics in poly(methyl methacrylate) exposed to a single ultra-short xuv/x-ray laser pulse," *Molecules* **26**(21), 6701 (2021).
39. E. Plönjes, B. Faatz, M. Kuhlmann, and R. Treusch, "FLASH2: Operation, beamlines, and photon diagnostics," *AIP Conf. Proc.* **1741**, 020008 (2016).
40. M. Manfredda, C. Fava, S. Gerusina, R. Gobessi, N. Mahne, L. Raimondi, A. Simoncig, and M. Zangrando, "The evolution of kaos, a multipurpose active optics system for euv/soft x-rays," *Synchrotron Radiation News* **35**, 29–36 (2022).
41. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT University, 2016).

42. I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 28 of Proceedings of Machine Learning Research S. Dasgupta and D. McAllester, eds. (PMLR, Atlanta, Georgia, USA, 2013), pp. 1139–1147.
43. A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information* **11**(2), 125 (2020).
44. A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **38**(1), 35–44 (2004).
45. J. Liu, "Simple technique for measurements of pulsed gaussian-beam spot sizes," *Opt. Lett.* **7**(5), 196–198 (1982).
46. P. C. Mahalanobis, "Recent experiments in statistical sampling in the indian-statistical-institute," *J. R. Stat. Soc. Ser. A-Stat. Soc.* **109**, 325–378 (1946).
47. S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers* (Springer-Verlag, 1999).