

PAPER • OPEN ACCESS

Image-based monitoring of femtosecond laser machining via a neural network

To cite this article: Ben Mills *et al* 2019 *J. Phys. Photonics* **1** 015008

View the [article online](#) for updates and enhancements.



PAPER

OPEN ACCESS

RECEIVED
30 April 2018

REVISED
9 July 2018

ACCEPTED FOR PUBLICATION
25 July 2018

PUBLISHED
6 December 2018

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Image-based monitoring of femtosecond laser machining via a neural network

Ben Mills , Daniel J Heath , James A Grant-Jacob , Yunhui Xie and Robert W Eason

Optoelectronics Research Centre, University of Southampton, Southampton, SO17 1BJ, United Kingdom

E-mail: bm602@soton.ac.uk

Keywords: laser materials processing, neural networks, ultrafast lasers

Abstract

Femtosecond laser machining offers the potential for high-precision materials processing. However, due to the nonlinear processes inherent when using femtosecond pulses, experimental random noise can result in large variations in the machined quality, and hence methods for closed loop feedback are of interest. Here we demonstrate the application of a neural network (NN), acting as a pattern recognition algorithm, for visual monitoring of the target substrate via a camera that observes the sample during machining. This approach has the advantage that it requires zero knowledge of the underlying physical processes, and hence avoids the need for modelling the complex photon–atom interactions that occur with femtosecond laser machining. The NN was shown to accurately determine the type of material, the laser fluence and the number of pulses, directly from a single image of the sample and within ten milliseconds. This approach provides the potential for real-time feedback for femtosecond laser materials processing.

1. Introduction

1.1. Laser machining

Lasers have transformed manufacturing over the past decades and are now routinely used for applications such as cutting [1, 2], welding [3, 4], drilling [5, 6], ablation [7, 8] and additive manufacturing [9–11]. However, the demand for increased productivity, scalability and improved functionality requires ever-increasing speed and quality of the manufacturing processes. One potential approach to fulfil such a demand is the integration of real-time feedback into the laser machining setup. This could be used for compensating random variations of laser parameters, such as output power or missed laser pulses, or for sample uncertainty, such as the removal of an unknown thickness of an outer material, such as rust in the context of laser cleaning. Feedback in laser-based processing has been studied for processes such as additive manufacturing and laser welding [12–16], where in general, the feedback relies on either optical or acoustic observation of the target material during fabrication processes, where the optical processes generally involve photodiodes, spectrometers, cameras, or infrared cameras using a variety of algorithmic approaches [17].

The development of such feedback algorithms usually require an understanding of the physical processes that underlie the fabrication process, for example, the effect of localised temperatures in the case of laser welding, or the interaction of light and matter in the case of laser ablation. For femtosecond laser ablation, its highly nonlinear nature due to the short time scales means that the particular dominant mechanism for material removal is particularly sensitive to the laser and experimental parameters, and hence developing a theoretical model for use in a feedback loop is considerably challenging. What is desirable, rather, is an approach that requires zero modelling of the complex photon–atom processes involved, and which instead operates purely using pattern recognition. Here, we demonstrate such an approach, through the use of a machine learning algorithm that was trained to determine the particular laser parameters that led to the ablation of the sample. These determined parameters could immediately be compared with the desired parameters, and if different, could be compensated for in real-time, for example by modulating the laser power.

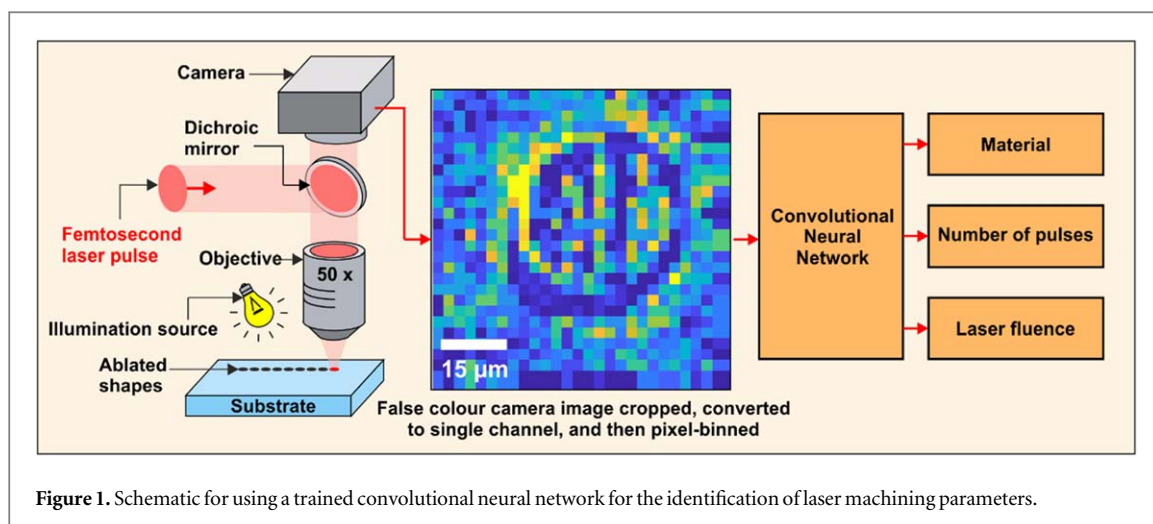


Figure 1. Schematic for using a trained convolutional neural network for the identification of laser machining parameters.

1.2. Neural networks

Discussed here is the sub-domain of machine learning referred to as neural networks (NNs) [18–21], and specifically, convolutional neural networks (CNNs) [22, 23]. In general, a NN is constructed from a large number of interconnected processing elements known as neurons, which form an input layer, one or more hidden layers, and an output layer. Each neuron receives a set of weighted inputs from the previous layer, which it processes and then passes the result to the next layer of neurons. The interconnected network of neurons creates a function that can transform data from one domain into another domain [24], and specifically in this work, can be used to convert an image of a laser-machined sample into a set of experimental parameters. Rather than providing a programmatical description of the physical processes that underlie that transfer function, a NN can learn the transfer function directly from the processing of labelled experimental data, also known as the training dataset. In general, before training commences, all weightings in the network are randomly initialised, and then during training, the weightings are algorithmically optimised via a process known as backpropagation [18]. Once trained on a dataset, the NN will have encoded an internal representation of the dataset and be able to predict the output for input data that it has never seen before.

CNNs are a variant of NN that excel at pattern recognition and identification of objects within images, and hence was chosen for this work. CNNs have seen applications in many diverse areas, such as language analysis [25, 26] and medical imaging and diagnosis [27, 28]. Previous work has shown the application of a (non-convolutional) NN to predict the effectiveness of laser machining given input parameters [29]. Here, we use a CNN for solving the inverse problem, namely, given an image of the surface of a laser-machined sample, determine the experimental parameters that were used, hence demonstrating the potential for real-time corrective algorithms for laser machining. Although femtosecond pulses were used here, as they enable considerably higher precision laser machining due to a reduced heat-affected-zone [30–32], the approach presented here can be similarly applied to other types of laser processing.

A fundamental strategy in using a CNN is the collection of a significant amount of labelled data. In this manuscript, section 2.1 describes the experimental setup and concept, section 2.2 describes the data collection and processing, and section 2.3 discusses the training procedure for the CNN. Section 3 shows results and discussion, and section 4 presents the conclusions. It is important to realise that here the CNN identifies the experimental parameters purely via pattern recognition of images of the laser-machined surfaces, without requiring any understanding of the underlying physics, and hence there is limited discussion in this manuscript on the physical nature of laser ablation.

2. Experimental setup

2.1. Laser machining setup

The schematic for visual-based identification of laser machining parameters is shown in figure 1. Laser pulses were used to machine the sample, whilst the camera imaged the sample during laser machining. The images were processed and then analysed by the trained CNN, and the CNN judged the most likely material, number of pulses, and laser fluence. In this subsection, the laser and imaging components are presented. The data collection and CNN training processes are discussed in sections 2.2 and 2.3, respectively.

The laser used for machining was a Ti:sapphire chirped pulse amplification system, pumped by a 80 MHz oscillator to produce 150 fs pulses with 1 mJ pulse energy at a wavelength of 800 nm. The laser fluence was

Table 1. Experimental parameters for the categories corresponding to laser machining in the silica substrate, showing number of pulses and fluence.

	A	B	C	D	E	F	G	H	I	J
Pulses	1	1	1	1	1	1	1	1	1	1
Fluence (J cm^{-2})	1.63	1.71	1.80	1.89	1.98	2.08	2.18	2.29	2.41	2.53

Table 2. Experimental parameters for the categories corresponding to laser machining in the nickel substrate, showing number of pulses and fluence.

	K	L	M	N	O	P	Q	R	S
Pulses	1	2	3	1	2	3	1	2	3
Fluence (J cm^{-2})	0.41	0.41	0.41	1.43	1.43	1.43	2.45	2.45	2.45

controlled via a graded neutral density filter and the number of pulses were controlled via access to the laser cavity Pockels cells, where the repetition rate of pulses used during the experiment was 20 Hz. The laser pulse intensity was spatially homogenised (via a Pi-Shaper 6_6) and then shaped into a circular profile using a digital micromirror device as a spatial light modulator [33, 34]. However, it should be noted that this approach will work for almost any beam shape, given that the beam shape is constant for all categories, and that different categories result in a different sample appearance. The shaped pulses were focussed onto the sample using a 50x microscope objective (Nikon ELWD) to a spot size of $\sim 30 \mu\text{m}$ (diameter), with a working distance of 10.1 mm. This extended working distance objective was chosen in order to minimise the recast of removed material on to the surface of the objective. A co-linear imaging line was used to record images of the sample during machining, using a CMOS camera (Thorlabs DCC1645C). The samples were mounted on an XYZ translation stage (Thorlabs LNR50S), with software corrections to ensure the samples were always at the same focal position. The samples used here were silica (a silica glass slide) and nickel (an electroless-nickel coated mirror, with a coating thickness of $5 \mu\text{m}$).

2.2. Data collection and analysis

A fundamental requirement for using a NN is the collection of suitable training data. Here, we chose to limit our training data to 19 categories, labelled A to S, where each category corresponded to a unique set of experimental parameters, namely material, number of laser pulses, and laser fluence. The choice of categories, in this proof-of-principle demonstration, was chosen to produce an asymmetric parameter space, i.e. one where the values within each parameter were not evenly distributed. This shows a key advantage of using a CNN in this approach, as the CNN acts purely as a pattern recognition device and works independently of the underlying physical experimental parameters. In other words, a CNN does not require a systematic approach to data collection. The parameters are shown in tables 1 and 2, for the silica and nickel substrate, respectively.

For silica (see table 1), laser fluences from 1.63 to 2.53 J cm^{-2} in incremental steps of 5%, all for a single laser pulse, were used, giving a total of 10 categories. For nickel (see table 2), three different laser fluences, 0.41 , 1.43 , and 2.45 J cm^{-2} , were used in combination with 1, 2, and 3 pulses, giving a total of 9 categories. To ensure unbiased training data, the images were collected by randomly choosing the category and position of each machined region for each of the two target substrates. This ensured that any systematic inhomogeneity across the sample or changing trend in any laser parameter during the experiment would be randomly distributed within the dataset and hence not learnt as an identification factor by the CNN. For each of the images, a non-ablated region of the material was used. Thus, each image contained a single machined region and corresponded to a single category. In total, 1800 pairs of images were collected, where each pair consisted of an image taken before, and an image taken after, the laser machining of each region. The 1800 images were split evenly between the two materials, thus 900 images per material. The 1800 images were randomly allocated into the training dataset (1620 images) and the validation dataset (180 images). For the training dataset, the number of images in categories A to S were, 86, 64, 81, 77, 86, 83, 68, 85, 99, 85, 73, 87, 96, 89, 93, 99, 89, 85 and 95, respectively. For the training dataset, the number of images in categories A to S were 9, 5, 12, 6, 9, 11, 8, 8, 10, 8, 8, 10, 11, 11, 12, 11, 8, 12 and 11, respectively. The images were cropped to the area of interest, background subtracted, pixel-binned (with 16 pixels being reduced to 1) to increase the signal-to-noise levels, and converted to a single channel by taking the average of the RGB channels. The final processed images contained 28 by 28 pixels. Although the final images were single channel 28 by 28 pixel images, for visual clarity, the figures in this manuscript show the single channel images in a 'blue to yellow', rather than a typical 'black to white', colour map.

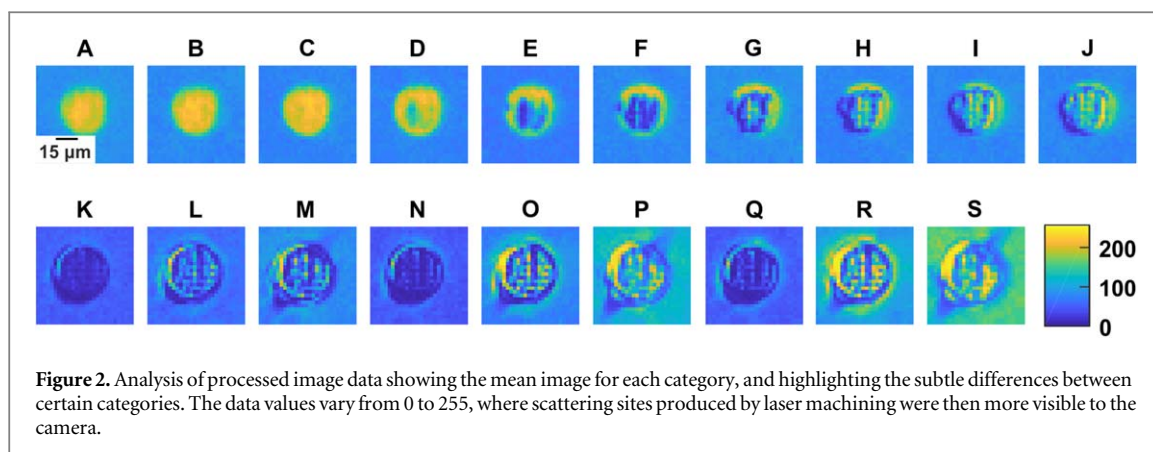


Figure 2, which shows the mean of all images within each category, highlights the subtle differences between each category, and hence demonstrates the considerable challenge of the identification of the experimental parameters from any single image. Whilst the underlying processes that govern the particular appearance of each category are complex, here the CNN acts purely as a pattern recognition algorithm, and therefore zero knowledge of the underlying physical processes is required for identification of the experimental parameters corresponding to any particular image. The differences between the images are a consequence of the scattering caused by laser processing of the surface, and hence different experimental conditions resulted in the variations in the images for each category. Silica is highly transparent to 800 nm wavelength light and hence absorption for this material during ablation would have been predominantly via multiphoton absorption [35]. Due to the resultant nonlinear processes, a thresholding effect appears to be observed, where above a specific fluence, the surface modification mechanism changed from melting (categories {A–C}) to ablation (categories {D–J}). For nickel, increasing the fluence and/or increasing the number of pulses intensified the contrast of the edges of the machined region, which can be attributed to laser machining deeper into the material and producing kerf at the edges. In this case, the illumination was from one direction, as this resulted in shadows and hence information on the depth of the ablation. The illumination was constant for all experiments, and hence independent of the category.

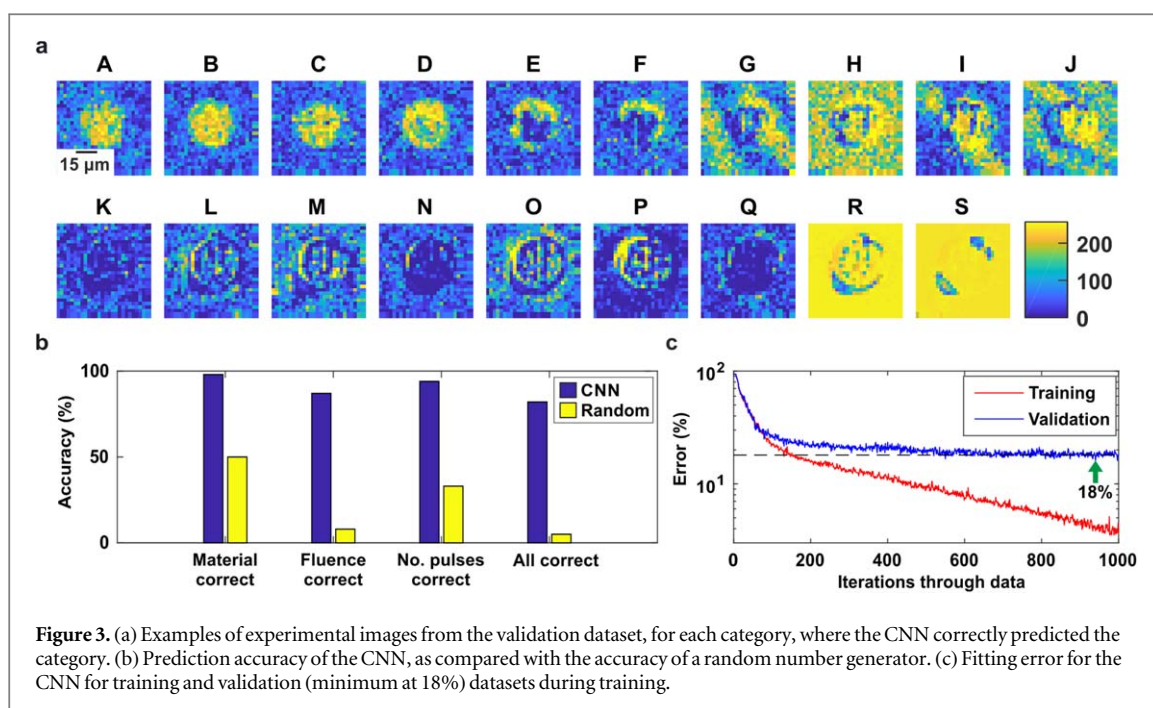
2.3. CNN architecture and training

When training and evaluating a NN, a key strategy is the separation of processed data into a training dataset and a validation dataset. The training dataset is used for training the NN. The validation dataset is used for evaluating the trained NN. At no point is the validation dataset used in training, and hence the accuracy on the validation dataset is indicative of the application of the NN to unseen data. This procedure is critical, as a NN will often over-fit to the training dataset, and hence the prediction accuracy on the training dataset can be misleading. Here, 90% of the 1800 processed images were randomly selected, and used as the training dataset. The remaining 10% were used as the validation dataset.

The CNN consisted of three convolutional layers (each with 32 filters of size 3×3 and with same padding) that were each followed by a max pooling of 2×2 and stride of 1, and a fully connected layer of 1024 neurons and classification layer (19 categories), using a learning rate of 0.0003, batch size of 10, and drop-out of 0.5. During training, each of the convolutional filters in the CNN were algorithmically optimised via gradient descent [22, 23]. The images in the training dataset were randomly ordered and used as the training input for the CNN, where the CNN was trained for 1000 epochs (where one epoch is defined as the processing of the entire training dataset exactly once). Once trained, for each image in the validation dataset, the CNN prediction was compared to the known category, hence providing a measure for which of the experimental parameters were correctly predicted in each case. The total training time was approximately one hour, and the time for identifying the category for a single image was approximately 30 ms.

3. Results and discussion

Figure 3(a) shows randomly selected images from the validation dataset where the trained CNN correctly predicted the category. Whilst there are distinct differences between the materials (categories {A–J} and {K–S}), the difference between validation images from adjacent categories, such as {B–D}, is extremely subtle. The diagonal features in the categories {G–J} are attributed to the random formation of debris causing a shadow within the silica from the directional illumination of the sample, and hence was averaged out in the mean images



in figure 2. The dramatic change in appearance in figure 3 for the images in categories {R, S} was likely caused by additional kerf at the edges of the laser-machined structure, which was found to more likely to occur at higher fluences. Despite the subtle differences between categories and the existence of low-probability cases (such as random debris and/or kerf), the CNN was still able to categorise these particular images correctly.

Figure 3(b) shows the percentage of (1) correct material, (2) correct fluence, (3) correct number of pulses and (4) all experimental parameters correct, as determined by the CNN [98%, 87%, 94%, 82%] and via a random number generator [50%, 8%, 33%, 5%], respectively. The random number generator picked a category at random for each image, hence guessing the material correctly in approximately 50% of cases, and is presented here to highlight the number of categories within each experimental parameter dimension. The CNN correctly determined the fluence for silica in 76% of the trials, and the fluence and number of pulses for nickel in 98% and 92%, respectively. A larger dataset would almost certainly improve this accuracy, due to additional information for differentiation of categories and the inclusion of low-probability edge cases, such as debris. Figure 3(c) shows the fitting error of the CNN for training and validation datasets during training. The validation error reached a minimum at 18% (corresponding to an 82% prediction accuracy for all experimental parameters), whilst the training error continued to decrease exponentially due to overfitting, which is generally an indicator that a CNN had the capacity to encode additional information. Prediction of the category for each validation image by the trained CNN took less than ten milliseconds, and hence using CNNs in this manner has huge potential for any laser-based fabrication approach where observation of the work piece is possible and could be used to enable real-time feedback.

Whilst three-dimensions of experimental parameters (material, fluence and number of pulses) were explored here, this approach could be extended indefinitely to any number of dimensions, given suitable training data and computing capability. In addition, parameters that are continuous rather than discrete, such as laser fluence, could have been output as a continuous variable rather than as a discrete category, and it would be expected that integration of this approach within an industrial setting would take full advantage of such a continuous output. However, due to the low number of combinations of experimental parameters in this proof-of-principle demonstration, a purely categorisation method was chosen. As the CNN works purely as a pattern recognition device, non-unique solutions i.e. different experimental conditions that result in the same appearance of the sample, could potentially limit this approach. This could occur, for example, if the laser power increased but machining time decreased unexpectedly. However, in general, non-unique solutions are very rare, as nonlinear processes, such as multiphoton absorption, and cumulative approaches, such as the effect of debris and melting on subsequent machining, means that the particular combinations of experimental parameters that lead to the same result are very specific and rarely encountered. If, however, a non-unique solution was encountered, meaning that expected changes to the experimental parameters still resulted in the originally desired sample appearance, the fact that the CNN would not detect the change would not be an issue as the sample would still be machined as per the initial requirements.

4. Conclusion

By training a CNN on a dataset of images of laser-machined surfaces, the CNN was able to determine the experimental parameters used during laser machining. The CNN achieved this despite having zero knowledge of the underlying physical processes that govern laser machining, and instead acted purely as a pattern recognition algorithm. Once trained, the CNN was able to identify the experimental parameters directly from images that were not part of the training dataset, and hence this approach could be a central component of a visual-based real-time closed loop feedback system for laser machining. In practice, this could be achieved via a CNN that continuously monitors the surface of the target substrate, where the output of the CNN (i.e. the predicted experimental parameters) at each moment in time could be compared to the desired experimental parameters. In the event of a difference between the predicted and desired parameters, a corrective action, such as decreasing or increasing the laser power, or even stopping the laser machining process entirely, could be automatically and immediately applied. This approach could be applied to both subtractive and additive laser processing, and for any size scale, given that the sample can be observed during fabrication, and where different experimental parameters result in a different appearance of the sample. With suitable training data, we anticipate that this approach could also be adapted to ensure the correct fabrication of a specific pattern or structure, where the desired pattern or structure is used as an input to the process.

Acknowledgments

This work was supported under EPSRC grant numbers EP/N03368X/1 and EP/N509747/1. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research. Supporting data for this submission can be found at <http://doi.org/10.5258/SOTON/D0429>.

ORCID iDs

Ben Mills  <https://orcid.org/0000-0002-1784-1012>

Daniel J Heath  <https://orcid.org/0000-0003-3566-1813>

James A Grant-Jacob  <https://orcid.org/0000-0002-4270-4247>

Robert W Eason  <https://orcid.org/0000-0001-9704-2204>

References

- [1] Bärsch N, Körber K, Ostendorf A and Tönshoff K H 2003 *Appl. Phys. A* **77** 237–42
- [2] Niziev V G and Nesterov A V 1999 *J. Phys. D: Appl. Phys.* **32** 1455
- [3] Tamaki T, Watanabe W, Nishii J and Itoh K 2005 *Japan. J. Appl. Phys.* **44** L687
- [4] Kaplan A 1994 *J. Phys. D: Appl. Phys.* **27** 1805
- [5] Luft A, Franz U, Emsermann L and Kaspar J 1996 *Appl. Phys. A* **63** 93–101
- [6] Weck A, Crawford T H R, Wilkinson D S, Haugen H K and Preston J S 2008 *Appl. Phys. A* **90** 537–43
- [7] Chichkov B N, Momma C, Nolte S, Von Alvensleben F and Tünnermann A 1996 *Appl. Phys. A* **63** 109–15
- [8] Liu X, Du D and Mourou G 1997 *IEEE J. Quantum Electron.* **33** 1706–16
- [9] Gu D D, Meiners W, Wissenbach K and Poprawe R 2012 *Int. Mater. Rev.* **57** 133–64
- [10] Frazier W E 2014 *J. Mater. Eng. Perform.* **23** 1917–28
- [11] Santos E C, Shiomi M, Osakada K and Laoui T 2006 *Int. J. Mach. Tools Manuf.* **46** 1459–68
- [12] Craeghs T, Clijsters S, Kruth J P, Bechmann F and Ebert M C 2012 *Phys. Proc.* **39** 753–9
- [13] Lott P, Schleifenbaum H, Meiners W, Wissenbach K, Hinke C and Bültmann J 2011 *Phys. Proc.* **12** 683–90
- [14] Mazumder J, Dutta D, Kikuchi N and Ghosh A 2000 *Opt. Lasers Eng.* **34** 397–414
- [15] Fathi A, Khajepour A, Toyserkani E and Durali M 2007 *Int. J. Adv. Manuf. Technol.* **35** 280–92
- [16] Hu D and Kovacevic R 2003 *Int. J. Mach. Tools Manuf.* **43** 51–60
- [17] Purtonen T, Kalliosaari A and Salminen A 2014 *Phys. Proc.* **56** 1218–31
- [18] Hinton G E and Salakhutdinov R R 2006 *Science* **313** 504–7
- [19] Rowley H A, Baluja S and Kanade T 1998 *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 23–38
- [20] Specht D F 1991 *IEEE Trans. Neural Netw.* **2** 568–76
- [21] Hecht-Nielsen R 1992 *Neural Networks for Perception* (Amsterdam: Elsevier) pp 65–93
- [22] Lawrence S, Giles C L, Tsoi A C and Back A D 1997 *IEEE Trans. Neural Netw.* **8** 98–113
- [23] Krizhevsky A, Sutskever I and Hinton G E 2012 *Advances in Neural Information Processing Systems* pp 1097–105
- [24] Hartman E J, Keeler J D and Kowalski J M 1990 *Neural Comput.* **2** 210–5
- [25] LeCun Y and Bengio Y 1995 *The Handbook of Brain Theory and Neural Networks* (Cambridge, MA: MIT Press) p 3361
- [26] Kalchbrenner N, Grefenstette E and Blunsom P 2014 arXiv:1404.2188
- [27] Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H and Ozcan A 2017 *Optica* **4** 1437–43
- [28] Cireşan D C, Giusti A, Gambardella L M and Schmidhuber J 2013 *Int. Conf. on Medical Image Computing and Computer-assisted Intervention* pp 411–8 (Berlin: Springer)
- [29] Madić M, Marković D and Radovanović M 2012 *Int. J. Adv. Intell. Paradigms* **4** 299–312
- [30] Pronko P P, Dutta S K, Squier J, Rudd J V, Du D and Mourou G 1995 *Opt. Commun.* **114** 106–10

- [31] Gattass R R and Mazur E 2008 *Nat. Photon.* **2** 219
- [32] Mills J D, Kazansky P G, Bricchi E and Baumberg J J 2002 *Appl. Phys. Lett.* **81** 196–8
- [33] Mills B, Feinaeugle M, Sones C L, Rizvi N and Eason R W 2013 *J. Micromech. Microeng.* **23** 035005
- [34] Mills B, Grant-Jacob J A, Feinaeugle M and Eason R W 2013 *Opt. Express* **21** 14853–8
- [35] Deng Z and Eberly J H 1985 *J. Opt. Soc. Am. B* **2** 486–93