

# Cascaded machine learning model for reconstruction of surface topography from light scattering

Mingyu Liu<sup>\*a</sup>, Nicola Senin<sup>a,b</sup>, Rong Su<sup>a</sup>, Richard Leach<sup>a</sup>

<sup>a</sup>Manufacturing Metrology Team, Faculty of Engineering, University of Nottingham, Nottingham, UK

<sup>b</sup>Department of Engineering, University of Perugia, Italy

## ABSTRACT

In this paper, we propose a light scattering method to identify classes of structured surface topographies and estimate their main geometric properties. The method is based on a cascaded machine learning model, designed as a two-layer architecture implemented using neural networks. The first layer consists of a classification model designed to determine which type/class of surface is being observed amongst a set of predefined surfaces. The second layer, cascaded to the first one, is designed to infer geometric properties specific to the individual structured surface being measured within each class, for example, pitch and height for a grating-type surface. The training datasets for the cascaded machine learning model, i.e. scattering signals from different surfaces, are generated through rigorous scattering simulation applied to computer-generated surfaces and based on a boundary element method. Once the model is trained, any scattering signal obtained from a real surface belonging to the considered classes can be fed into the model, and both the surface class and specific values for its geometric properties can be quickly estimated. For validation, we developed a prototype experimental apparatus to generate light scattering data from real surface samples. Different grating patterns (classes) were considered, as well as different values for the main geometric properties specific to each class. Validation consisted both in the assessment of classification performance in recognising instances of each specific class and in quantification of estimation accuracy in determining the geometric properties of each instance, by comparison with measurements performed with atomic force microscopy.

**Keywords:** Surface topography, surface reconstruction, light scattering, cascaded machine learning model

## 1. INTRODUCTION

Measurement of surface topography plays an important role in the manufacturing industry, as surface topography is one of the key factors to ensure the performance of a functional component [1]. Optical surface topography measurement [2], such as imaging confocal microscopy, focus variation microscopy and coherence scanning interferometry, may have difficulty to measure complex surface structures such as sharp edges and vee-grooves with high accuracy [3, 4]. Atomic force microscopy (AFM) [5] and scanning tunnelling microscope (STM) [6] has been demonstrated for on-machine measurement of micro-structured surfaces, but besides the low scanning speed, AFM and STM are limited by a compromise between range and resolution [7]. Light scattering techniques such as scatterometry are used for in-process surface measurement as they can infer surface information from scattering patterns [8], and they have the advantages of being non-contact, high speed and low-cost. Scatterometry has been widely used for the measurement of critical dimension in semiconductor chips [9] and measurement of surface roughness [10]. However, full reconstruction of surface topography from scattering patterns is still challenging, due to the complexity of the inverse scattering problem, in particular, because very similar scattering patterns may be generated by different surfaces.

Our work stems from the consideration that, when information on the measured surface is available in advance (e.g. types of surface structure, range of values for its defining geometric parameters), one can infer a sizeable amount of information from any measured surface by testing against a finite range of alternative estimates. Such an approach works both when assessing what class a surface may belong to (by implementing classifiers), and when assessing the specific values for geometric/dimensional parameters that define the surface within a specific class (by implementing regressors). As the performance of any reconstruction solution based on such an approach is affected by breadth and heterogeneity of classes/parameter values that must be addressed at the same time, in this work we propose a solution specifically

dedicated to addressing surface gratings, a family of structured surfaces characterised by periodic structures. Our solution covers multiple types/classes of gratings, each defined by geometric/dimensional parameters such as pitch and height. The solution consists of a cascaded model, i.e. a sequence of a classifier (to determine what class/type of grating each surface belong to) followed by a regressor (to determine the values for the geometric parameters specific to each class of grating). The model is cascaded because the type of regressor adopted to estimate the geometric parameters (second stage) depends on the result of the classifier (first stage). Both stages are based on machine learning models. The second stage is for each class addressed in the first stage but trained with different data.

The highlight of the proposed cascaded machine learning model is that training can be performed on simulated data, whilst the trained model can then operate on real (experimental) light scattering data. A prototype system was built to demonstrate the surface reconstruction using this method for several types of grating surfaces.

## 2. METHOD

### 2.1 Cascaded machine learning model

The diagram of the proposed cascaded machine learning model is shown in Fig. 1. The scattering signal is first fed into a classification machine learning model and the surface class can be identified. For each class of surface, corresponding regression machine learning models are trained to estimate the surface defining parameters for that particular class (e.g. pitch and height) for each specific type of grating. Once class and geometric parameters have been determined, the reference topography that is most representative of the measured one can be obtained by geometric reconstruction.

The classification model can deal with various classes of surfaces (not limited to four as shown in Fig. 1). For each class of surface, multiple defining parameters can be determined (not limited to two as shown in Fig. 1). The regression models are independent of each other and can be reused when adding more surface types into the existing model. As a result, the design of the cascaded machine learning model is highly flexible and extendable.

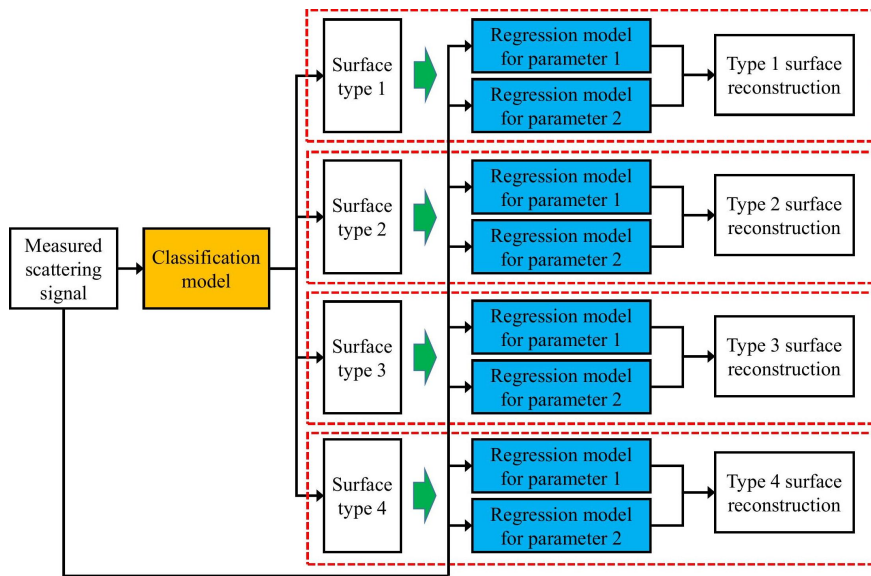


Figure 1. Diagram of the proposed cascaded machine learning model.

### 2.2 Prototype system

A prototype system was developed to evaluate the proposed method, as shown in Fig. 2. A collimated laser beam (wavelength = 633 nm, beam width  $\approx$  1 mm) is incident onto the surface of the measured sample at an angle of  $45^\circ$ . A sensor module (SM) constructed with a pinhole, a focusing lens and a photodiode is mounted on a rotation stage to capture the scattered light in an arc trajectory (similar to a goniometer). The scanning range is from  $0^\circ$  (the initial position is shown in Fig. 2) to  $120^\circ$ . The angular resolution is  $0.1^\circ$  for the scanning. The analogue scattering signal is processed by an amplifier (AMP), then converted to a digital signal by an analogue-to-digital converter (ADC), and finally recorded by a computer. The scattering signal can then be fed into the designed machine learning model.

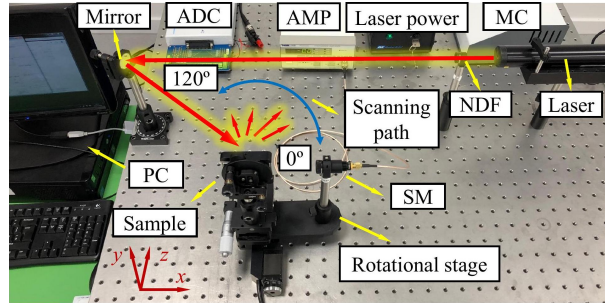


Figure 2. Designed of the prototype system.

### 2.3 Machine learning models

Two types of machine learning models are implemented in the proposed method, one for the classification stage, the other for the regression stage (estimation of the geometric parameters of the surface). The surface classes selected for this work are one-directional gratings so that two-dimensional (2D) scattering signals measured orthogonally to the grating direction can be considered as an adequate approximation of a fully three-dimensional (3D) case, and numerical simulation of the far-field is less complex and time consuming. The simulation can, therefore, be implemented to replicate the experimental set-up, with the same angle for the incident light ( $45^\circ$ ) and the same angular resolution in the resulting, simulated spectrum ( $0.1^\circ$  over a  $120^\circ$  arc).

A conceptual diagram for the classification stage (first stage of the cascaded model) is shown in Fig. 3, the core element being a neural network. In training, computer-generated surface instances (2D cross-sectional grating profiles) are inserted into the light scattering simulation model based on a boundary element method (BEM) [11, 12]. The far-field spectrum resulting from the simulation, binned into  $0.1^\circ$  arc intervals, is normalised (intensities converted to the 0 to 1 interval) and fed into the input layer of the neural network (number of neurons equal to the number of bins). The neural network is fully connected, with a hidden layer and an output layer. The number of nodes in the output layer is equal to the number of classes handled by the classifier. Each node containing a 0 to 1 value representing the likelihood for each class. The classification result is taken as the node with the highest likelihood value. The neural network is trained by minimising the loss function designed as sparse categorical cross-entropy [13] and the optimisation algorithm is designed as the adaptive moment estimation [14]. The activation functions are rectified linear units (ReLU) [15]. For training, multiple scattering spectra are generated for each class by simulation. Each is characterised by small variations (incident illumination angle, the relative position of the illuminated region of the grating, values of the parameters defining the individual grating within each class) so that intrinsic variability in the input can be incorporated and classifier robustness can be increased.

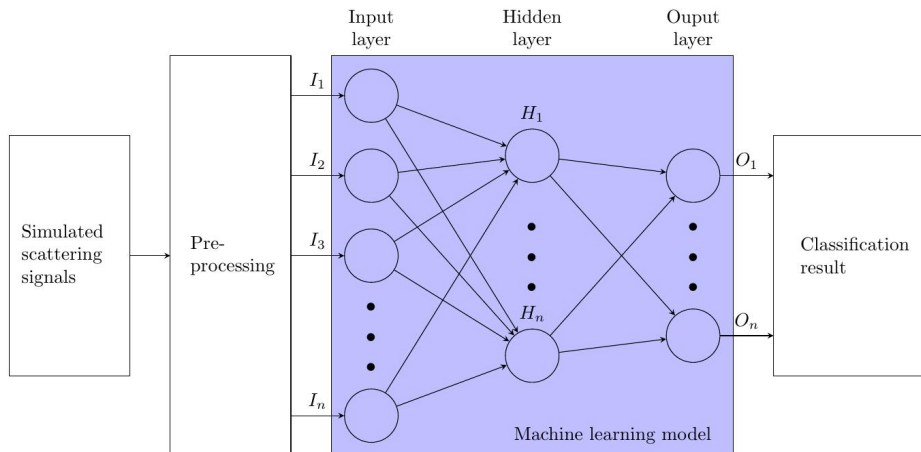


Figure 3. Diagram of the machine learning model for surface classification.

In Fig. 4, the conceptual diagram for the regression stage of the cascaded model is shown. Whilst each regression unit will have the same architecture, training will be different as each is trained on data relative to a specific class of grating,

and only one of its defining parameters. The pitch of a sinusoidal grating-class is considered as an example in Fig. 4. The regression stage is also powered by a neural network made of three layers: an input layer, a hidden layer and an output layer. As for the classifier, the input layer of the regressor consists of binned intensities of the scattering spectrum, again normalised to the 0 to 1 interval. However, compared to the classifier, there is only one neuron in the output layer, containing the value for the estimated geometric parameter (e.g. pitch). Again simulation is used to train the model, and again natural variability is included in the training set (incident illumination angle, the relative position of the surface cross-section and geometric parameters themselves) to increase the robustness of the regressor to variations across instances of the specific class of surfaces covered. The loss function for the neural network is modelled as the mean square error (MSE), the optimisation algorithm is RMSprop [16]. The activation functions are ReLUs [15]. As stated previously, a dedicated and separately trained regression model is needed of every geometric parameter to be estimated for each one of the classes being covered by the cascaded model.

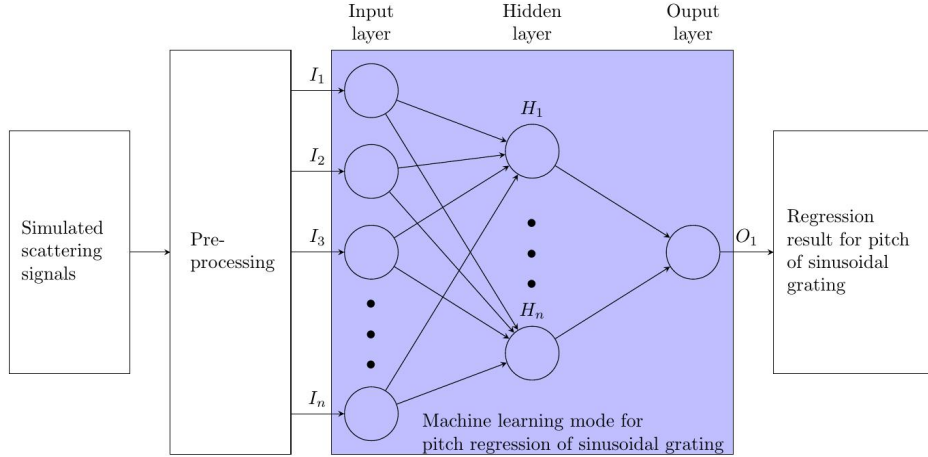


Figure 4. Diagram of the machine learning model for estimation of one of the surface defining parameters (e.g. pitch of sinusoidal grating).

### 3. RESULTS AND DISCUSSION

The range of surfaces selected for this work covered two distinct classes of gratings (blazed grating and sinusoidal grating) and two defining geometric parameters for each (pitch of the grating and peak-to-valley amplitude, referred to as height). Four physical samples were used for validation: two for the blazed grating class and two for the sinusoidal grating class). The blazed grating physical samples had spatial frequencies of 300 lines/mm and 600 lines/mm (corresponding pitch values  $3.33\ \mu\text{m}$  and  $1.67\ \mu\text{m}$  respectively), where their nominal height values were both  $0.20\ \mu\text{m}$ . The two sinusoidal gratings had spatial frequencies of 125 lines/mm and 400 lines/mm (corresponding pitch values  $8.00\ \mu\text{m}$  and  $2.50\ \mu\text{m}$  respectively), where their nominal height values were  $0.20\ \mu\text{m}$  and  $0.12\ \mu\text{m}$  respectively. The reported spatial frequencies are nominal values thus, the surfaces were first measured by an AFM (Bruker FastScan) for validation and the results are shown in Fig. 5.

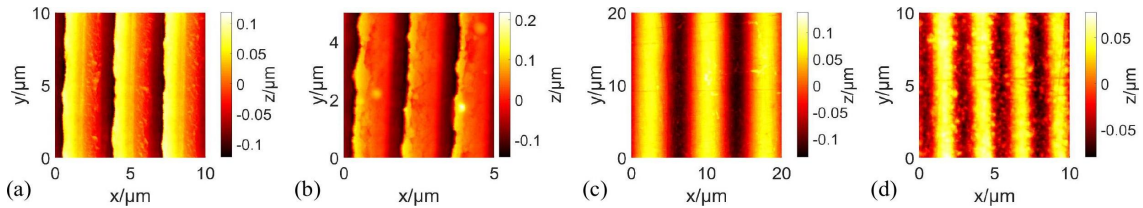


Figure 5. AFM results for the surfaces, (a) blazed grating 300 lines/mm, (b) blazed grating 600 lines/mm, (c) sinusoidal grating 125 lines/mm, and (d) sinusoidal grating 400 lines/mm.

The classification part of the cascaded machine learning model was designed to recognise twelve types of surfaces. They are blazed gratings with spatial frequencies of 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm, and sinusoidal gratings with spatial frequencies of 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm, and square gratings with

spatial frequencies of 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm. The twelve classes are summarised in Table 1. Essentially, the adopted approach was to subdivide the values of the defining geometric parameters (pitch) into intervals, so that a more significant part of the reconstruction problem could be handled by the classifier (as opposed to simply discriminating between blazed and sinusoidal gratings), therefore, simplifying the work of the regressor model (as each specific regressor would have to deal with a narrower range of values). For each class, six observations were obtained by simulation, incorporating the intrinsic variability as illustrated earlier, leading to a total of  $12 \times 6^5 = 93312$  datasets.

Table 1. Surface classes defined for the classification model.

Surface type	Pitch/ $\mu\text{m}$	Height/ $\mu\text{m}$	Incident angle/ $^\circ$	Sample offset in z direction/mm	Alignment error in x direction/mm
Blazed grating 125 lines/mm	7.20 to 8.80	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Blazed grating 300 lines/mm	2.99 to 3.66	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Blazed grating 400 lines/mm	2.25 to 2.75	0.10 to 0.13	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Blazed grating 600 lines/mm	1.50 to 1.83	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Sinusoidal grating 125 lines/mm	7.20 to 8.80	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Sinusoidal grating 300 lines/mm	2.99 to 3.66	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Sinusoidal grating 400 lines/mm	2.25 to 2.75	0.10 to 0.13	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Sinusoidal grating 600 lines/mm	1.50 to 1.83	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Square grating 125 lines/mm	7.20 to 8.80	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Square grating 300 lines/mm	2.99 to 3.66	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Square grating 400 lines/mm	2.25 to 2.75	0.10 to 0.13	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00
Square grating 600 lines/mm	1.50 to 1.83	0.18 to 0.22	40.50 to 49.50	-1.00 to 1.00	-1.00 to 1.00

The four physical samples were illuminated using the prototype system and the measured scattering signals (normalised) are shown in Fig. 6. Spectral binning led to 1201 data points for each spectrum. The scattering signals were then fed into the trained classifier. The classification results are shown in Fig. 7. The results show that all surfaces are successfully classified with very high probabilities.

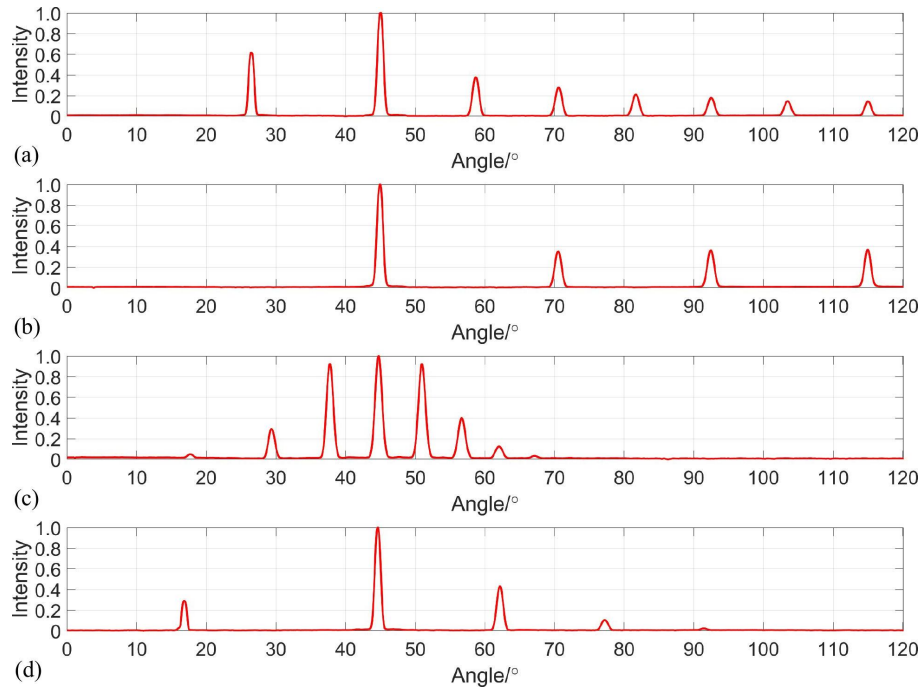


Figure 6. Measured scattering signals for the surfaces, (a) blazed grating 300 lines/mm, (b) blazed grating 600 lines/mm, (c) sinusoidal grating 125 lines/mm, and (d) sinusoidal grating 400 lines/mm.

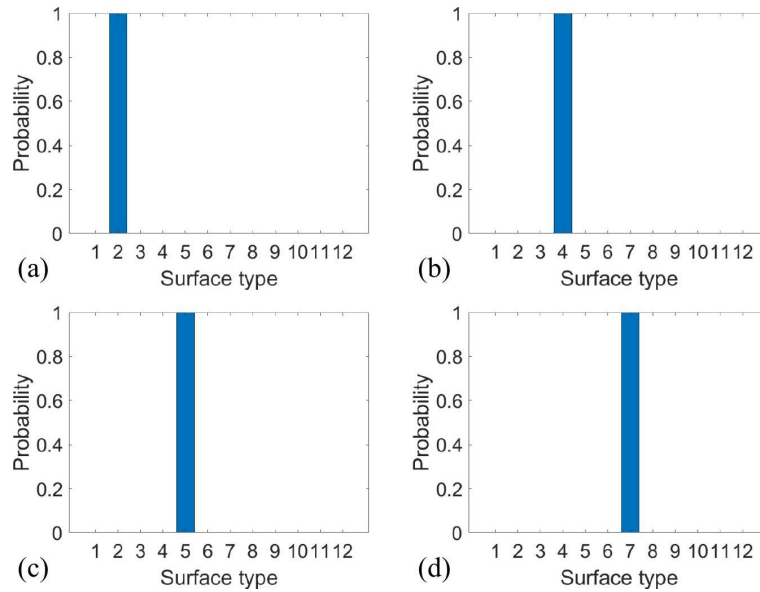


Figure 7. Classification results for, (a) blazed grating 300 lines/mm, (b) blazed grating 600 lines/mm, (c) sinusoidal grating 125 lines/mm, and (d) sinusoidal grating 400 lines/mm. Horizontal axes: surface type 1 to 12 represent blazed gratings with 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm, sinusoidal gratings with 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm and square gratings with 125 lines/mm, 300 lines/mm, 400 lines/mm, 600 lines/mm, respectively. Vertical axes: the probabilities (likelihood) that the instance belongs to a specific class.

Training datasets were also generated by simulation for the regression models. There were a total of  $7776 \times 10 = 77760$  datasets for each one of the  $12 \times 2$  regression models (twelve classes, two parameters per class – pitch and height). The results obtained by feeding the trained regressors with real spectra obtained from the four physical samples are shown in Table 2.

To evaluate the results for the regression models, least-squares best-fit values for the surface defining parameters were determined using the AFM data. The best-fit results are shown in Table 3. The results show that all predicted results using the proposed method are close to the best-fit values using the AFM data, which demonstrate the accuracy of the regression model.

Table 2. Regression results for surface defining parameters.

Surface type	Estimated pitch value/ $\mu\text{m}$	Estimated height value/ $\mu\text{m}$
Blazed grating 300 lines/mm	3.26	0.19
Blazed grating 600 lines/mm	1.60	0.20
Sinusoidal grating 125 lines/mm	7.53	0.18
Sinusoidal grating 400 lines/mm	2.64	0.10

Table 3. Best-fit results for surface defining parameters from AFM data.

Surface type	Fitted pitch value/ $\mu\text{m}$	Fitted height value/ $\mu\text{m}$
Blazed grating 300 lines/mm	3.32	0.18
Blazed grating 600 lines/mm	1.67	0.20
Sinusoidal grating 125 lines/mm	7.70	0.18
Sinusoidal grating 400 lines/mm	2.53	0.10

After the surface defining parameters were determined, surface topographies were reconstructed from the estimation of their geometric parameters. The reconstructed surfaces were compared with the AFM data and the results are shown in Fig. 8 to 11. The AFM results were registered with the reconstructed surfaces using a method described elsewhere [17]. The difference between reconstruction and AFM measurement were quantitatively determined as shown in Fig. 8 to 11(b). The RMS of the error (the local difference between the AFM and reconstructed topography) were  $0.018 \mu\text{m}$ ,  $0.025 \mu\text{m}$ ,  $0.010 \mu\text{m}$  and  $0.015 \mu\text{m}$  respectively. The results show that the reconstructed surfaces have sub-micrometre level differences compared to the AFM results, which demonstrates the effectiveness of the proposed method.

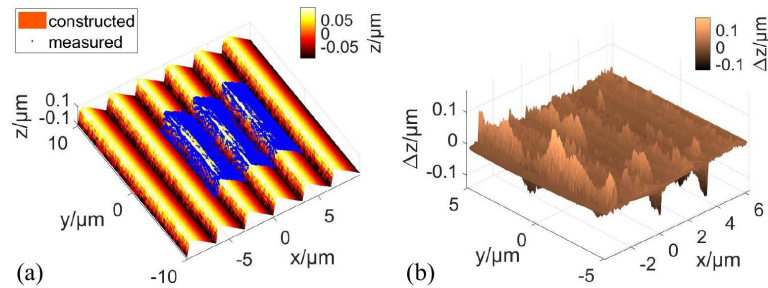


Figure 8. Reconstructed blazed grating surface (300 lines/mm), (a) registered with AFM data, (b) reconstruction error.

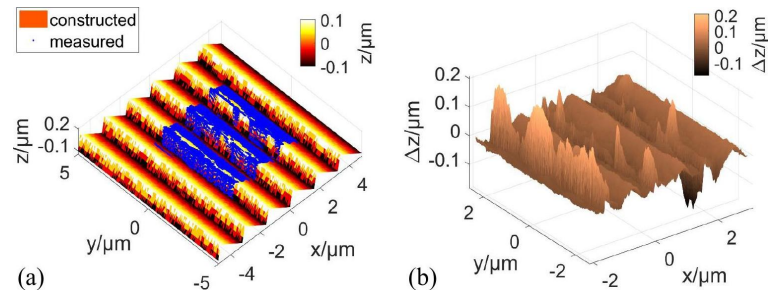


Figure 9. Reconstructed blazed grating surface (600 lines/mm), (a) registered with AFM data, (b) reconstruction error.



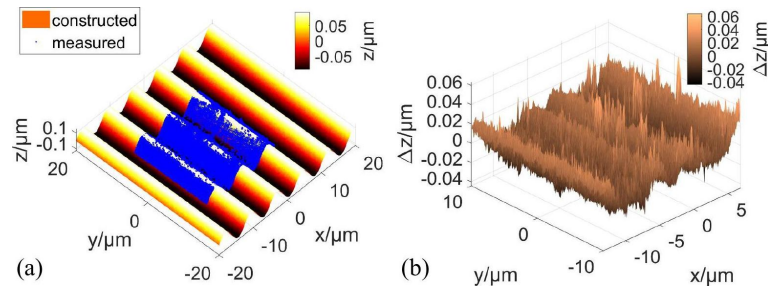


Figure 10. Reconstructed sinusoidal grating surface (125 lines/mm), (a) registered with AFM data, (b) reconstruction error.

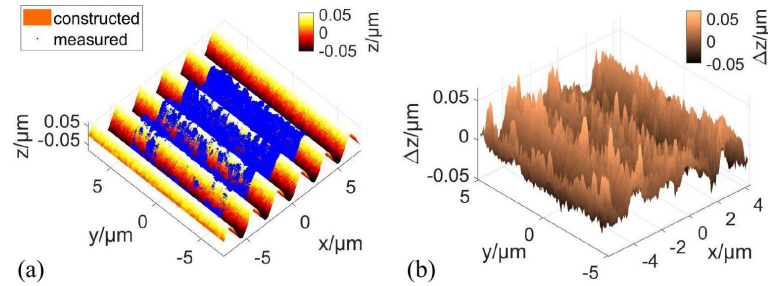


Figure 11. Reconstructed sinusoidal grating surface (400 lines/mm), (a) registered with AFM data, (b) reconstruction error.

## 4. CONCLUSIONS

This paper presents a cascaded machine learning model for the reconstruction of surface topography from light scattering, given the a priori knowledge of classes and ranges of geometric parameters that must be addressed in the inspection. The cascaded machine learning model is designed as a top-down structure that begins with a classification model followed by multiple regression models. Experiments show that the proposed method is capable of reconstructing surface topographies from light scattering signals. The proposed method is intrinsically extendable to cover more surface classes in multiple scenarios of industrial inspection.

## ACKNOWLEDGEMENT

We acknowledge the support from the Engineering and Physical Sciences Research Council (EP/R028826/1), European Union's Horizon 2020 Research and Innovation Staff Exchange Programme (MNR4SCell, 734174), and the access to the University of Nottingham's Augusta HPC service.

## REFERENCES

- [1] Whitehouse, D. J., [Handbook of surface metrology], CRC Press, (1994).
- [2] Leach, R. K., [Optical Measurement of Surface Topography], Springer, (2011).
- [3] Fang, F., Zeng, Z., Zhang, X., and Jiang, L., "Measurement of micro-V-groove dihedral using white light interferometry," *Optics Communications*, 359, 297-303 (2016).
- [4] Gao, F., Leach, R. K., Petzing, J., and Coupland, J. M., "Surface measurement errors using commercial scanning white light interferometers," *Measurement Science and Technology*, 19(1), 015303 (2007).
- [5] Gao, W., Aoki, J., Ju, B.-F., and Kiyono, S., "Surface profile measurement of a sinusoidal grid using an atomic force microscope on a diamond turning machine," *Precision Engineering*, 31(3), 304-309 (2007).
- [6] Goto, S., Hosobuchi, K., and Gao, W., "An ultra-precision scanning tunneling microscope Z-scanner for surface profile measurement of large amplitude micro-structures," *Measurement Science and Technology*, 22(8), 085101 (2011).
- [7] Leach, R. K., Jones, C. W., Sherlock, B., and Kryszynski, A., "The high dynamic range surface metrology challenge." 149-152 (2013).



- [8] Vorburger, T. V., Silver, R., Brodmann, R., Brodmann, B., and Seewig, J., [Light scattering methods] Springer, (2011).
- [9] Raymond, C. J., [Scatterometry for semiconductor metrology] CRC Press, (2001).
- [10] Bennett, J., and Mattsson, L., [Introduction to surface roughness and scattering, 2nd edition], Optical Society of America, (1999).
- [11] Simonsen, I., "Optics of surface disordered systems," *The European Physical Journal Special Topics*, 181(1), 1-103 (2010).
- [12] Thomas, M., Su, R., Nikolaev, N., Coupland, J., and Leach, R. K., "Modelling of interference microscopy beyond the linear regime," *Optical Engineering*, 59, 034110 (2020).
- [13] De Boer, P.-T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y., "A tutorial on the cross-entropy method," *Annals of Operations Research*, 134(1), 19-67 (2005).
- [14] Kingma, D. P., and Ba, J., "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, (2014).
- [15] Agarap, A. F., "Deep learning using rectified linear units (ReLU)," arXiv preprint arXiv:1803.08375, (2018).
- [16] Tieleman, T., and Hinton, G., "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," COURSE: Neural Networks for Machine Learning, 4(2), 26-31 (2012).
- [17] Liu, M. Y., Cheung, C. F., Feng, X., Wang, C. J., and Cao, Z. C., "Any-degrees-of-freedom (anyDOF) registration for the characterization of freeform surfaces," *Precision Engineering*, 62, 170-180 (2020).