

Super-resolution for 2.5D height data of microstructured surfaces using the vdsr network

Stefan Siemens^{1,*}, Markus Kästner¹, and Eduard Reithmeier¹

¹Leibniz University Hannover, Institute of Measurement and Automatic Control, Nienburger Straße 17, 30167 Hannover, Germany

Abstract. In this work super-resolution imaging is used to enhance 2.5D height data of thermal sprayed Al₂O₃ ceramics with stochastically microstructured surfaces. The data is obtained by means of a confocal laser scanning microscope. By implementing and training a Very Deep Super-Resolution neural network to generate residual images an improvement of the peak signal-to-noise ratio and structural similarity index can be observed when compared to classic interpolation methods.

INTRODUCTION

Advances in manufacturing lead to increasing demands on the quality and durability of technical components. The topography of a components surface such as the microstructure have major impact on its performance. Surface metrology therefore plays an important role in the research and development of technical components. In recent years, optical measurement methods have increasingly been used for surface metrology as they allow for a contactless and fast measurement of the topography. With regards to measuring microstructures confocal microscopes are among the most widely used measuring instruments [1]. Generally, it is desirable to measure with the highest possible resolution to capture small structures. At the same time a large area should be measured in order to obtain reliable results. These two objectives are in conflict with each other, since lenses with a higher magnification and resolution usually have a smaller field of view. One approach to circumvent this is super-resolution imaging, where low-resolution (LR) images are upscaled to a higher resolution (HR) to obtain fine details whilst keeping a large field of view.

BACKGROUND

Mathematically the degradation of the HR image (denoted by x) to the LR image (denoted by y) can be modelled by the degradation function Φ and the degradation parameters θ_η [3]:

$$y = \Phi(x, \theta_\eta). \quad (1)$$

The goal of the super-resolution algorithm is to find a function Φ^{-1} and parameters θ_ζ that invert the degradation function to get an estimate of the HR image \hat{x} :

$$\hat{x} = \Phi^{-1}(y, \theta_\zeta). \quad (2)$$

As Φ and θ_η are unknown and different HR images can result in the same LR image this task is an ill posed problem. There are many approaches to generate a good HR approximation [2, 3]. The most widely used are classic upsampling methods like the nearest-neighbor, bilinear or bicubic interpolation. In recent years different machine learning approaches have been developed, for example linear networks or GAN models [3].

In this work we chose to train a Very Deep Super-Resolution (VDSR) neural network, as it has the advantage of being relatively easy to train whilst yielding good performance metrics. It was published by Kim et al. in 2016 and is inspired by the VGG-Net [4]. The VDSR network uses a pre-upsampling approach to super-resolution. This means that the input is not the LR image but an interpolated low-resolution image (ILR) which is the LR image interpolated to the dimensions of the HR image using classic upsampling methods [4]. The network then learns to generate the residual mapping between the HR and ILR image. This approach simplifies the training objective as the network only has to fill in high frequency information. This in turn allows for the training of deeper networks compared to earlier works like by Dong et al. [5]. With exception to the first layer (which is the input layer) and the last two layers (the reconstruction and output layer) the network consists of 20 pairs of convolution layers each followed by a ReLU activation layer. The convolution layers each consist of 64 filters of the size 3×3 .

DATA & TRAINING SETUP

The examined surfaces are thermal sprayed Al₂O₃ ceramics with a stochastic pore distribution, an average pore area of 15 μm^2 and a porosity of 8 % to 9 %. The confocal laser scanning microscope (CLSM) Keyence VK-210X was used to acquire so called 2.5 D height images where each pixel of the image has a corresponding height value of the measured surface.

* Corresponding author: stefan.siemens@imr.uni-hannover.de

In total 680 height images were taken each with a size of 1024x768 pixels. 600 images were used for training and 80 for testing. The lens used has a magnification of 50x and a numerical aperture of 0.95. An example surface is shown in Figure 1.

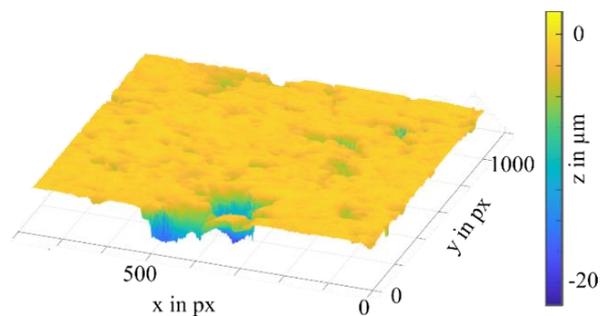


Fig. 1. 2.5D height image of a thermal sprayed Al_2O_3 ceramic.

To train the network 20 patches with a size of 128x128 pixels were extracted from each height image for a total of 12.000 training patches. For this work synthetically degraded data was used. Before being fed into the VDSR the patches were first scaled down by a factor of 2, 3, 4 or 8 and then scaled up by the same factor using bicubic interpolation. Additionally, the training data was augmented with a random rotation by 0° , 90° , 180° or 270° and a random reflection along the x- and y-axis. The network was trained for 150 epochs using a stochastic gradient descent. The hyperparameters were adopted from Kim et al. [4]. The output of the network is the predicted residual image which gets added to the ILR image. The loss function for training is given by half the MSE between ground truth and the predicted residual.

RESULTS

For testing purposes, the image data was also synthetically degraded using bicubic interpolation and scaling factors of 2, 3 and 4. The peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) between the prediction and ground truth were calculated and averaged for all testing images. The results were compared to an upscaling of the LR image using only the nearest-neighbor, bilinear or bicubic algorithm. The results are shown in Table 1. The performance of the VDSR network exceeds the other methods in most scenarios, especially for larger scaling factors. An example is shown in Figure 2. The visual impression is consistent with the previous results.

Table 1. Testing results for different methods and scale factors.

Magnification / Method	2x: PSNR / SSIM	3x: PSNR / SSIM	4x: PSNR / SSIM
Nearest-neighbor	27.13 dB / 0.95	22.07 dB / 0.83	22.21 dB / 0.82
Bilinear	28.50 dB / 0.95	23.44 dB / 0.86	23.47 dB / 0.84
Bicubic	30.18 dB / 0.97	23.71 dB / 0.87	24.28 dB / 0.86
VDSR	30.72 dB / 0.96	23.92 dB / 0.87	25.14 dB / 0.88

While the VDSR results clearly outperform the nearest-neighbor and bilinear approach they are hard to distinguish from the bicubic interpolation.

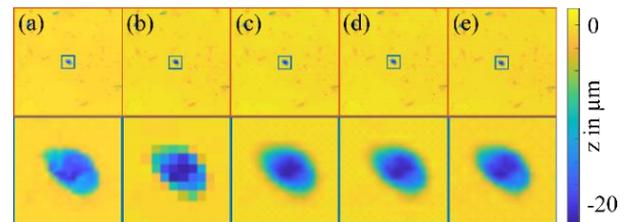


Fig. 2. Overview (top) and detailed view (bottom) of results for a 4x upscaling using different methods. Ground truth (a), nearest-neighbor (b), bilinear (c), bicubic (d) and VDSR (e).

Looking at the residual images (see Figure 3) it can be seen that the VDSR network successfully generates high frequency information which leads to the increase in PSNR and SSIM.

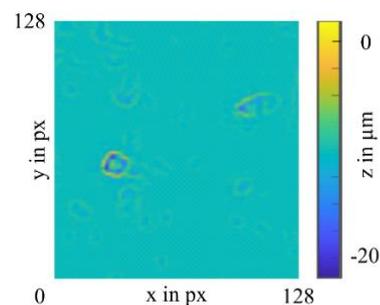


Fig. 3. Example for a residual image generated by the VDSR Net for a 4x upscaling.

CONCLUSION

This work shows that the VDSR network can be used to super-resolve 2.5D height image data acquired by means of a confocal laser-scanning microscope. It is relatively easy to train and in many cases outperforms classic interpolation methods. Limitations of this work are that real degradation processes are unknown and likely not the same as a synthetic down-sampling.

References

1. A.M. Martinez, J. de Vicente Y Oliva, *Materials*, **12(24)**, 4137 (2019)
2. W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, Q. Liao, *IEEE Transactions on Multimedia*, **21(12)**, 3106-3121 (2019)
3. S. Anwar, S. Khan, N. Barnes, *ACM Computing Surveys*, **53(3)**, 1-34 (2020)
4. J. Kim, J.K. Lee, K.M. Lee, *2016 CVPR*, 1646-1654 (2016)
5. C. Dong, C.C. Loy, K. He, X. Tang, *Computer Vision – ECCV 2014*, **8692**, 184-199 (2014)