Noise robust focal distance detection in laser material processing using CNNs and Gaussian processes

Sepehr Elahi[a], Can Polat[b], Omid Safarzadeh , and Parviz Elahi[b]

[a]Department of Electrical and Electronics Engineering, Bilkent University, 06800, Ankara, Turkey

[b]Department of Physics, Bogazici University, Bebek 34342 Besiktas, İstanbul, Turkey

ABSTRACT

In this work, we investigate the effects of noise on real-time focal distance control for laser material processing by generating the images of a sample at different focal lengths using Fourier optics and then designing, training, and testing a deep learning model in order to detect the focal distances from the simulated images with varying standard deviations of added noise. We simulate both input noise, such as noise due to surface roughness, and output noise, such as detection camera noise, by adding zero-mean Gaussian noise to the source wave and the simulated image, respectively, for different focal distances. We then train a convolutional neural network combined with a Gaussian process classifier to predict focus distances of noisy images together with confidence ratings for the predictions.

Keywords: Focus detection, Fourier optics, machine learning, surface roughness, deep learning, Gaussian process

1. INTRODUCTION

In high precision laser micro-machining, it is required to actively control the machining setup's performance and make the necessary corrections when needed. One of the most critical control mechanisms required for precise machining is controlling the focus position on the work-piece. One focus detection method, which relies on the diffraction effect,[1] is achieved by curve fitting the reflected beam's intensity from the work-piece when landed on a camera. Although they achieve low focal distance prediction error, their method works best on noise-free images. However, in many laser machining setups, noisy images are inevitable, whether due to a low-resolution camera or noise due to the surface roughness of the work-piece, like in micro-machining.

In our work, we investigate the effects of noise on active focal distance detection by simulating the image of a sample at different focal lengths using Fourier optics[2] and then designing, training, and testing a novel deep learning model combined with Guassian processes to classify the focal distances from the simulated images with varying strengths of added noise.

2. FOURIER OPTICS SIMULATION

2.1 Wave Propagation Simulation

We use a similar approach of[1] for the simulation setup. As for the main difference, we consider the initial wave a Gaussian beam and include the noise effects simultaneously at both input and output. Explicit form[3] of the initial beam's electric field can be given as (1).

$$E_0(x, y, z = 0) = E_0 \frac{w_0}{w(0)} e^{-\frac{r^2}{w^2(0)}} , \tag{1}$$

Further author information:
E-mail: parviz.elahi@boun.edu.tr

Optics, Photonics and Digital Technologies for Imaging Applications VII, edited by Peter Schelkens, Tomasz Kozacki, Proc. of SPIE Vol. 12138, 1213802
© 2022 SPIE · 0277-786X · doi: 10.1117/12.2624337

we simulate the free-space propagation for distance L with the Fresnel approximation[4] as shown in (2).

$$U_1(x,y,L) = -i\frac{e^{\frac{i2\pi L}{\lambda}}e^{\frac{i\pi(x^2+y^2)}{L\lambda}}}{\lambda L}\int\int E_0(x',y',z=0)e^{i\pi\frac{x'^2+y'^2}{\lambda L}}e^{-2i\pi\frac{xx'+yy'}{\lambda L}}\,dx'dy', \tag{2}$$

with this propagation, our reflected wave arrives at the surface of the first lens, which collimates the wave when the sample is at focus. A lens's effect can be expressed as a modulation of the incoming wavefront. We can model the lens[5] with a focal length of f_1 as shown below

$$L_1(x,y) = e^{-i\pi\frac{(x^2+y^2)}{\lambda f_1}}, \tag{3}$$

so after the first lens we have the our wave as

$$U_2(x,y) = U_1(x,y)L_1(x,y), \tag{4}$$

again, we propagate the wave to the lens for d distance. This can be calculated by changing L to d and E_0 to U_2 in equation (2). We call this wave U_3. Then we modulate it with a lens having a focal length of f_2. So f_2 instead of f_1 in the equation (3) and call it L_2. We obtain the wave just after the second lens as below

$$U_4(x,y) = U_3(x,y)L_2(x,y), \tag{5}$$

now lastly, we make the wave reach the detection camera by propagating it for f_2 distance using (2) and calling it U_5. In order to obtain the intensity value at the camera, we need to take the absolute square of this wave. Therefore we have the intensity distribution at the camera as

$$I(x,y) = |U_5|^2. \tag{6}$$

2.2 Input and Output Noise Simulation

We consider the input noise, which corresponds to surface roughness, as a Gaussian,[6] by adding a random phase of the initial electric field, given by

$$E_0'(x,y,z=0) = E_0\frac{w_0}{w(0)}e^{-\frac{r^2}{w^2(0)}-i\eta_{\text{in}}}, \tag{7}$$

where η_{in} is the zero mean Gaussian input noise with standard deviation σ_{in}. Thus, we write $\eta_{\text{in}} \sim \mathcal{N}(0,\sigma_{\text{in}}^2)$. There is an important parameter when considering surface roughness which is correlation length. Sample simulation for different correlation length can be seen in Figure 2. As for the output noise, we also consider a Gaussian, but this time we add it to equation (6) in both dimensions. We base our output noise level compatible with available cheap camera's readout noise. Thus, given that $I(x,y)$ is the image intensity at point (x,y), the noisy image intensity is given by

$$I'(x,y) = I(x,y) + \eta_{\text{out}}, \tag{8}$$

where $\eta_{\text{out}} \sim \mathcal{N}(0,\sigma_{\text{out}}^2)$.

2.3 Simulation Parameters and Generated Images

Since we are interested in the laser machining application, our range for defocus positions will be based on the Rayleigh length.[7] Starting from the calculation of beam waist $(2\omega_0)$ as shown below

$$2\omega_0 = \left(\frac{4\lambda}{\pi}\right)\left(\frac{F}{D}\right), \tag{9}$$

where λ is the beam's wavelength, F is the focal length of the lens that focuses the light onto the sample, and D is the diameter of the source beam. One can obtain the value of Rayleigh length (L_R) from the equation below

$$L_R = \frac{\pi\omega_0^2}{\lambda}, \tag{10}$$

Table 1: Simulation parameters for the data generation.

Rayleigh length	Laser wavelength	Collimating beam diameter	Beam waist	Defocus distances	Input noise stds.	Output noise stds.
20 µm	976 µm	2 mm	5 µm	$\{-250, -200, \ldots, 0, \ldots, 200, 250\}$ µm	$\{0, 0.33, 0.67, 1.0\}$ µm	$\{0, 0.01, 0.02, 0.03\}$ µm

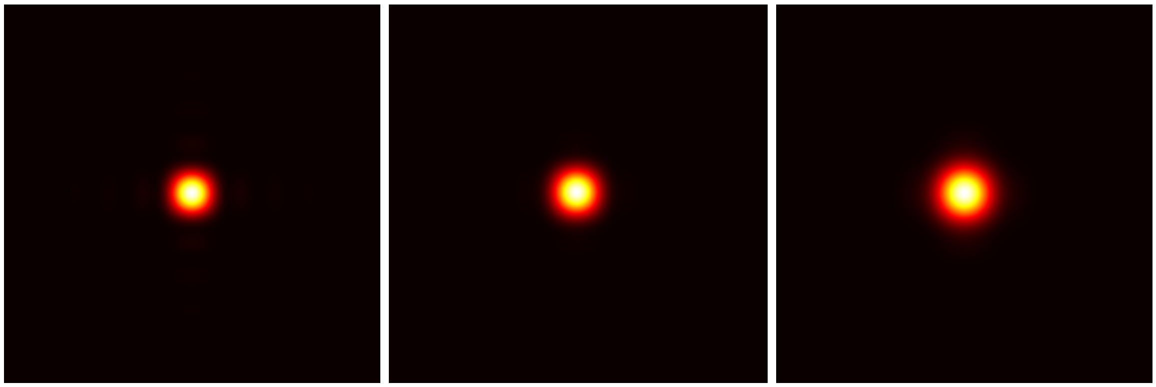

Figure 1: Simulated images for non-noised waves. Diffraction rings appear when the sample is 100 µm closer than the focus distance (left), a small, focused beam when the sample is at the focus (middle), and a widespread beam when the sample is 100 µm away than the focus distance (right).

We consider focus distances from −250 µm to 250 µm with steps of 50 µm. In Figure 1 we show the simulation output of three different focus distances with no noise. Moreover, we consider low and high correlation lengths for the surface roughness with four equally spaced input noise standard deviation from 0 nm to 1000 nm. Simulation for two different correlation lengths are given in the Figure 2. Finally, since the detection system for the reflected light is not isolated and bound to environmental noise, we also consider the noise at the output with standard deviations of $\{0, 0.01, 0.02, 0.03\}$ µm, an example of which can be found in the Figure 3. All simulation parameters are given in Table 1.

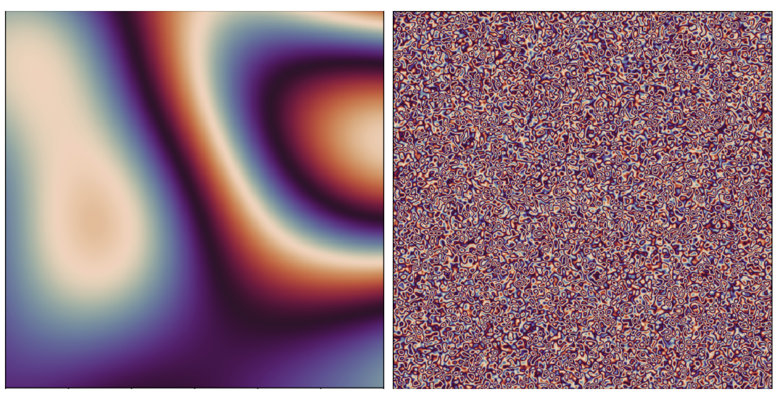

Figure 2: Simulation of surface roughness. Correlation length is 300 µm (left) and when 5 µm (right).

3. FOCUS DISTANCE CLASSIFICATION

Machine Learning methods have been applied to many areas of physics, including statistical physics, quantum physics, quantum computing, cosmology, and chemical physics.[8] Machine Learning methods are also used for laser machining to improve its accuracy, speed, and online modeling of laser machining at scale.[9] Some scholars also used deep learning for system monitoring via visual observation of the work-piece during laser processing.[10] They have used Convolutional Neural Networks (CNN) to detect single-axis beam translation. CNNs are able

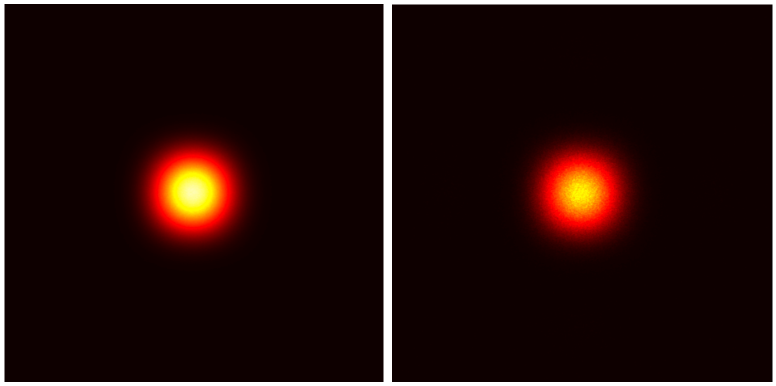

Figure 3: Simulated images with added output noise. Standard deviation of the Gaussian noise is 100 nm on the left and 500 nm on the right.

to automatically learn important image features without any human input. Moreover, they can be trained to be robust to noise.

Although CNNs perform extremely well on image-related classification tasks, they cannot provide an uncertainty measurement for their predictions. As it is important to have such a measure when all possible classes are not available, we combine CNNs with Gaussian processes (GP), which do provide us with uncertainty measures in the form of the posterior variance[11,12]

3.1 CNN and GPs for Image Classification

Our CNN model takes as input an image and classifies its focal distance class. It consists of two convolution layers with RELU[13] activation functions followed by max-pooling layers. Pooling layers reduce dimensions of data and provide outputs for successive layers through one neuron. More specifically, we used max-pooling to use the maximum value of each neuron to create a feature map. The architecture of our model is given in Table 2.

We then take the output of the penultimate 256-dimensional output of the CNN and feed it to a GP classifier that classifies the focus distance while providing a confidence rating of how sure it is about its prediction. Notice that the CNN model bust be trained first, after which its weights are frozen, its last layer removed, and it is connected to a GP classifier which is then retrained on the same data as the CNN.

3.2 Data Generation

We generate 100 images per focal distance and input and output noise standard deviations. We generate two sets of images per the mentioned specs, one for training and validation and the other for testing. Furthermore, we use a 85%/15% split for training/validation.

3.3 CNN Training

Using a batch size of 16 and SGD with a momentum of 0.93 and a learning rate of 0.001, we train the CNN for 200 epochs, picking the weights from the epoch corresponding to the highest validation accuracy. The model is trained on a PC running Ubuntu 20.04 LTS with an i7 6800k equipped with 32 GB of RAM and a GTX 1080Ti.

3.4 GP Training

We then take the trained CNN and strip off the output layer so that the CNN outputs a 256-dimensional vector. This vector is essentially a 256-dimensional encoding of the image. We fit a GP with a zero mean and SquaredExponential kernel onto the feature vectors of every image in the training set. Technically, the GP is a multioutput GP with 11 outputs, one for each class. The i^{th} GP's posterior mean is the probability that the class of the input image is i, and the posterior variance of the i^{th} GP indicates how unsure the model is about its prediction. The higher the variance, the less sure the model is.

Table 2: The architecture of the CNN network used for focus distance classification.

Layer	**Parameters**
Input	Size: 1x40x32
Conv	Out channels: 32 Kernel size: (3,3)
Max-pool	Kernel size: (2,2) Stride: (2,2)
Conv	Out channels: 64 Kernel size: (3,3)
Max-pool	Kernel size: (2,2) Stride: (2,2)
Flatten	Output size: 1x3072
Dropout	Probability: 0.5
Dense	Output size: 512
Dense	Output size: 256
Dense	Output size: 11

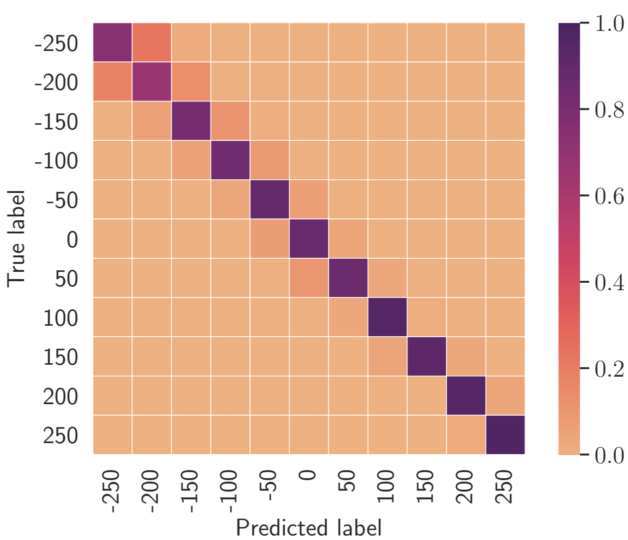

Figure 4: Confusion matrix of our CNN+GP model when tested on the testing data.

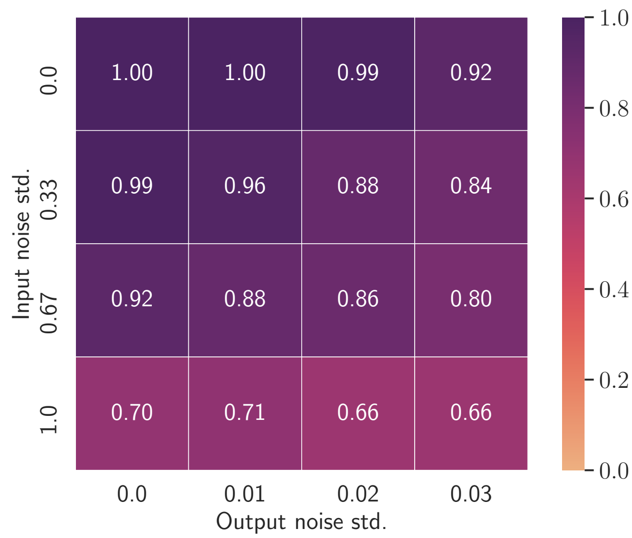

Figure 5: Classification testing accuracy of the CNN+GP model corresponding to images of each input-output standard deviation noise pair.

3.5 Results

We test our trained CNN+GP model and present its confusion matrix in Figure 4. Notice that the model confuses some labels, but only with neighboring labels. For instance, it mistakes −50 µm for −100 µm. Overall, the model achieves a 86% testing accuracy, which is considerably better than the baseline of $1/11 = 9.1\%$. We also present individual testing accuracies corresponding to each input-output noise standard deviation pair, given in Figure 5. As expected, the higher the noise, the lower the accuracy. It also appears that the output noise has a larger effect when adjusted for standard deviation, as the going from no output noise to 0.03 std. results in an 8% accuracy drop while going from no input noise to 0.33 std. results in only a 1% accuracy drop.

Lastly, we measure the inference speed of our model to find that the CNN model on its own achieves 1200 Hz on the CPU, while the combined CNN+GP model achieves 430 Hz. Thus, for setups where extreme speed is needed while uncertainty measures are not, a CNN-only setup would be better suited.

4. DISCUSSION

We explored the effect of noise in a focus control mechanism for a laser machining system by simulating images with input and output noise. We then designed and trained a combined CNN and GP model that is able classify the focus distance of a simulated noisy image and output an uncertainty measure. Therefore, our model is not only robust to both input and output noise, but it can also determine when the image is too noisy to make meaningful predictions.

We analyzed the classification accuracy of the trained model when tested on unseen noisy images of increasing standard deviation and observed that although its performance decreases with increasing noise, it manages to outperform the baseline by a considerable amount. Lastly, we tested the inference speed of our model and showed that even on a GPU-less machine it can achieve extremely fast inference speeds.

High output noise caused the reflection of the light from a focused sample to look like it was coming from a defocused position. This result can easily interfere with the auto-focusing system and cause damage to the work-piece. There can be possible solutions to this problem: designing a model with higher noise levels or using a high resolution and low readout noise camera. Nonetheless, the latter would be a defective approach since instead of increasing the cost of the experimental setup, one can improve the model quickly with little more data. However, our method offers high accuracy for the given noise levels with the defocus range. Therefore, stating that there is no need for expensive equipment for these parameters.

ACKNOWLEDGMENTS

This work is partially financed through the BAP, Start-Up project 21B03SUP3 at Boğaziçi University, awarded to Parviz Elahi.

REFERENCES

[1] Xu, S.-J., Duan, Y.-Z., Yu, Y.-H., Tian, Z.-N., and Chen, Q.-D., "Machine vision-based high-precision and robust focus detection for femtosecond laser machining," *Opt. Express* **29**, 30952–30960 (Sep 2021).

[2] Goodman, J. W., "Introduction to fourier optics, roberts & co," *Publishers, Englewood, Colorado* (2005).

[3] Yariv, A. and Yeh, P., [*Optical waves in crystals*], vol. 5, Wiley New York (1984).

[4] Ware, M. and Peatross, J., [*Physics of Light and Optics (Black & White)*], Brigham Young University, Department of Physics (2020).

[5] Saleh, B. E. and Teich, M. C., [*Fundamentals of photonics*], john Wiley & sons (2019).

[6] Ogilvy, J. A., "Wave scattering from rough surfaces," *Reports on Progress in Physics* **50**, 1553–1608 (dec 1987).

[7] Damask, J. N., [*Polarization optics in telecommunications*], vol. 101, Springer Science & Business Media (2004).

[8] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová, L., "Machine learning and the physical sciences," *Reviews of Modern Physics* **91**(4), 045002 (2019).

[9] Mills, B. and Grant-Jacob, J. A., "Lasers that learn: The interface of laser machining and machine learning," *IET Optoelectronics* **15**(5), 207–224 (2021).

[10] Xie, Y., Heath, D. J., Grant-Jacob, J. A., Mackay, B. S., McDonnell, M. D., Praeger, M., Eason, R. W., and Mills, B., "Deep learning for the monitoring and process control of femtosecond laser machining," *Journal of Physics: Photonics* **1**(3), 035002 (2019).

[11] Milios, D., Camoriano, R., Michiardi, P., Rosasco, L., and Filippone, M., "Dirichlet-based gaussian processes for large-scale calibrated classification," *Advances in Neural Information Processing Systems* **31** (2018).

[12] Williams, C. K. and Rasmussen, C. E., [*Gaussian processes for machine learning*], vol. 2, MIT press Cambridge, MA (2006).

[13] Fukushima, K. and Miyake, S., "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in [*Competition and cooperation in neural nets*], 267–285, Springer (1982).