

CLASSIFYING SCANNING ELECTRON MICROSCOPE DEFECTS WITH A CNN AND A CUSTOM FOURIER-TRANSFORM LAYER

Magdaléna Skřičilová¹, Oldřich Trenz¹

¹Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1,
613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_033

ABSTRACT

A scanning electron microscope has to be checked periodically, and two of its most common image faults, defocus and astigmatism, are tedious to judge by eye. This article presents a convolutional neural network that sorts pairs of microscope images into four focus and astigmatism classes so the check can run without an operator. Each sample is two shots of the same spot: one in the base position and one with a shifted working distance, because the shifted shot is what makes astigmatism visible. The decisive ingredient is a custom Fourier-transform preprocessing layer that converts each input to its magnitude spectrum before the convolutional part sees it. On a dataset of about three thousand image pairs, the base model variants without the Fourier layer reached 49–65% test accuracy, while the corresponding variants with the Fourier layer reached 89–94%; the best result was 94% with channel-axis concatenation. The same shallow model beat three pretrained ImageNet networks (MobileNetV2 87%, EfficientNetV2L 52%, ResNet50 44%), which we did not expect. The best fusion of the two images was concatenation along the channel axis together with the Fourier layer.

Keywords: convolutional neural network, scanning electron microscope, Fourier transform, image classification, astigmatism

JEL Code: C45

1 INTRODUCTION

In the development and testing of scanning electron microscopes, repeated regression tests are needed to check whether microscope functions continue to work correctly after software or parameter changes. Two image faults addressed in this work are defocus and astigmatism; both can occur for several reasons, including sample or detector contamination, incorrect electron-source settings, magnetic-lens hysteresis or unsuitable microscope parameters. Their visual



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
<https://creativecommons.org/licenses/by-sa/4.0/deed.en>

similarity makes manual assessment difficult, which motivates testing whether a neural network can classify the image states automatically and support an automated regression-testing framework for the microscope.

A single shot is ambiguous: a focused image with astigmatism and a defocused image without it can be visually identical. Software testing treats this kind of input-dependent behaviour as something to exercise systematically rather than check once (Ammann and Offutt, 2008), which is the spirit of running the classifier as a repeatable test. The way out is to take two pictures of the same position and let the network compare them. The aim is to build such a classifier, push its accuracy high enough to be useful, and find which design choices matter.

1.1 Focus and astigmatism correction in the SEM

Automatic focusing and astigmatism correction for electron microscopes is an old problem. Erasmus and Smith (1982) built an early hardware system for it, and Ogasawara *et al.* (1999) showed that the Fourier transform of an SEM image carries a clean signature of both faults, because defocus and astigmatism reshape the frequency spectrum in characteristic ways. Lu, Zhang and Li (2018) later turned this into a simplified single-image correction method based on the spectrum. The shared idea, that the frequency domain rather than the raw pixels is where focus information lives, is the starting point here.

Recent work has moved the same task onto deep learning. Na *et al.* (2021) used a network to discriminate refocused SEM images for materials science, Lee *et al.* (2021) built a dual deep-learning network for robust autofocus, and Schubert, Saxena and Kornfeld (2024) reported fast focus and astigmatism correction. Most of these read a single image. The classifier described below instead reads a pair and asks a four-way question, focused or not and stigmatic or not, which is why two shots are needed.

1.2 Convolutional networks and blur detection

Convolutional neural networks (CNNs) are the standard tool for image classification (LeCun, Bengio and Hinton, 2015; Aggarwal, 2018). The deep ImageNet architectures, AlexNet (Krizhevsky, Sutskever and Hinton, 2012) and ResNet (He *et al.*, 2016) among them, win on natural-image benchmarks with tens to hundreds of layers. Detecting blur is narrower, and there the picture is less one-sided: Szandala (2020) showed that a CNN can replace the classical Laplacian sharpness measure. Whether a deep pretrained network or a small purpose-built one fits the SEM defect task is one of the questions answered here.

2 METHODOLOGY

The task was framed as a four-class classification of image pairs. Each sample is two pictures of the same position: a base (focused) shot and a second shot taken with the working distance shifted, which throws the image out of focus on purpose. The four classes are focused-stigmatic, focused-astigmatic, defocused-stigmatic and defocused-astigmatic. The second shot disambiguates the hard pair from the introduction, since astigmatism shows up as a sideways elongation once the image is defocused. A focused-astigmatic and a defocused-stigmatic sample, identical in the first shot, separate in the second.

Data were collected by a script hooked into the microscope-control software, which cycled through commands to take the four image types at various positions and magnifications. Defocus was induced by shifting the working distance by 0.05 to 0.9 mm depending on magnification. Images were then checked and sorted by hand, which was slow and not error-free, since the labeling task is hard for a person too. The final dataset held about 3 000 image pairs, balanced across classes, shuffled and split into training, validation and test parts, with 80% used for training. The standard tin-balls-on-carbon sample was used throughout, because its regular round shape makes defects easy to see.

Everything ran in Python with TensorFlow and its Keras API on a GPU. Each image was decoded, resized to 128×128 pixels, cast to float and normalized to the range [0, 1]. On the training set only, both images of a pair were augmented with the same random seed (horizontal and vertical flips, an occasional 90° rotation, a brightness change, a random crop) so the pairing stayed consistent.

The base network is deliberately small: three convolutional layers (3×3 filters, 32 then 64 then 96 of them) interleaved with 2×2 max-pooling, then four fully-connected layers ending in the four-class output. It uses ReLU activations, He-normal initialization, L2 weight regularization and a single dropout layer. The whole model is about 1.3 million parameters and 5.04 MB on disk. Training used the Adam optimizer (Kingma and Ba, 2014) with an exponentially decaying learning rate, sparse categorical cross-entropy loss, and roughly 50 epochs on average, with learning-rate reduction and early stopping on the validation loss.

Two design axes were varied. The first is how the two images are fused into one input: side by side (horizontal), stacked as extra channels (channel-axis), or processed in separate branches whose outputs are concatenated. The second is the Fourier-transform layer described below, tested on and off for every fusion type. For comparison, three pretrained ImageNet models from Keras Applications (MobileNetV2, EfficientNetV2L and ResNet50) were also tried as feature extractors. Fine-tuning them gave very poor results, probably because the dataset is small, so only feature extraction is reported.

2.1 The Fourier-transform layer

The key addition is a custom Keras preprocessing layer, with no trainable weights, placed right after the input. For each image it converts to grayscale, runs a two-dimensional FFT, shifts the zero frequency to the centre, takes the log-magnitude spectrum and normalizes it. For the fused inputs the combined image is first split back into its two images, each transformed on its own, then put back together. The point is what the spectrum looks like: a sharp focused image keeps its high frequencies, a defocused one loses them, and astigmatism turns the spectrum into an ellipse stretched to one side. The network therefore sees the defect almost drawn out for it, instead of learning the same cue from raw pixels. Figure 1 shows where the layer sits in the model, and Figure 2 gives a defocused-stigmatic example with its spectrum.

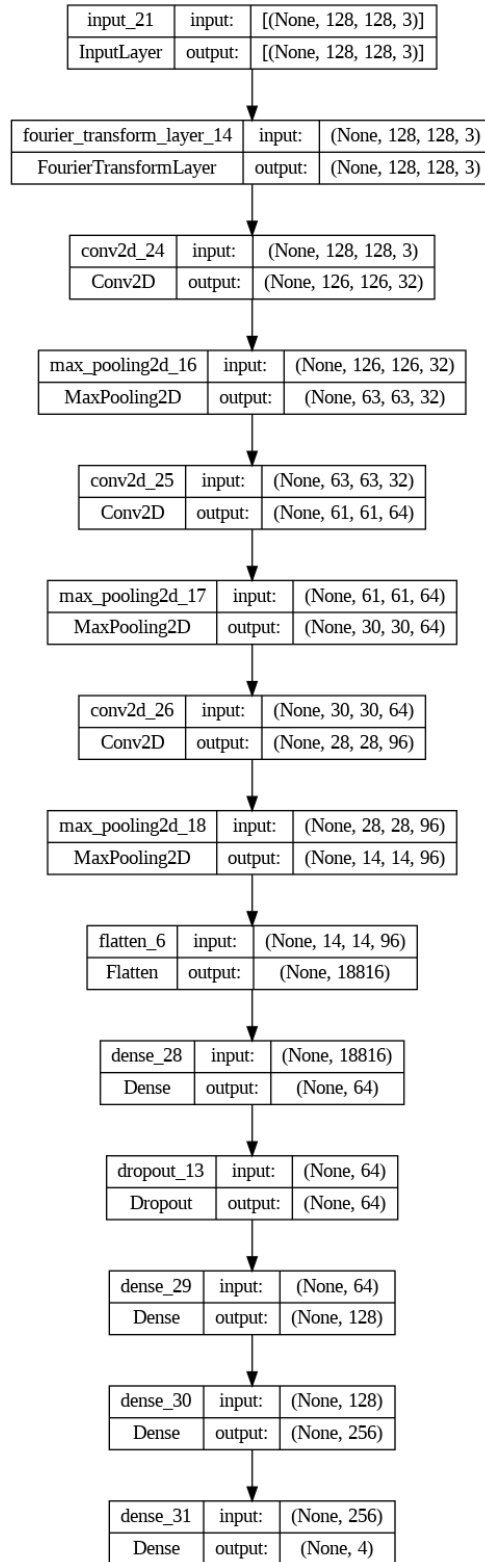


Fig. 1: Architecture of the proposed convolutional neural network with a custom Fourier-transform layer, classifying 128×128 electron microscope images into four focus and astigmatism classes (adopted from Skricilova, 2024)

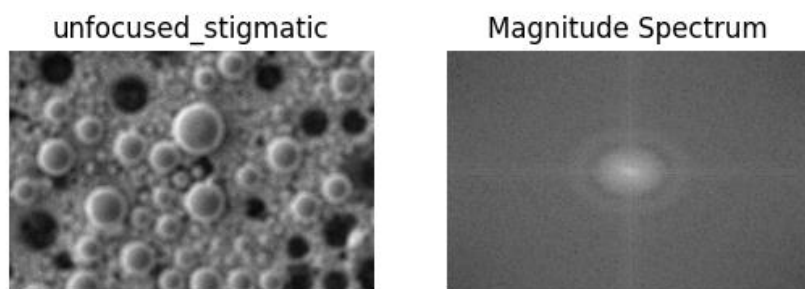


Fig. 2: Example electron microscope image of the defocused-stigmatic class with its corresponding Fourier magnitude spectrum used as model input (adopted from Skricilova, 2024)

3 RESULTS

All accuracies were measured on the held-out test set and are rounded to whole percent. Table 1 reports the base model across the three fusion types, with and without the Fourier layer; Table 2 sets the best base model against the three pretrained networks. The best row in each table is shown in bold.

The Fourier layer changed the outcome on every fusion type, and by a wide margin. Channel-axis fusion went from 55% to 94%, the output-layer variant from 49% to 93%, and horizontal fusion from 65% to 89%. Without the layer the ranking of the fusion types is reversed, with horizontal best at 65% and channel-axis near the bottom at 55%. There is no single best way to combine the two images; it depends on whether the Fourier layer is present (Skricilova, 2024).

The small custom model came out on top at 94%, ahead of all three pretrained networks. Among those, the lightest one, MobileNetV2, did best at 87%, while the two heavier networks landed near chance for a four-class problem (ResNet50 44%, EfficientNetV2L 52%). The Fourier layer did not help the pretrained models, since their best results all came with it switched off, the opposite of what it did for the shallow model. The chosen model for the framework is the custom network with channel-axis fusion and the Fourier layer.

Tab. 1 Base model accuracy by input fusion type, without and with the Fourier-transform layer

Fusion type	Accuracy without FFT	Accuracy with FFT
Horizontal	65%	89%
Channel-axis	55%	94%
Output-layer	49%	93%

Tab. 2 Best base model against pretrained ImageNet models (feature extraction, best parameter combination per model)

Model	Fusion	FFT layer	Accuracy
Custom base model	channel-axis	yes	94%
MobileNetV2	output-layer	no	87%
ResNet50	output-layer	no	44%
EfficientNetV2L	output-layer	no	52%

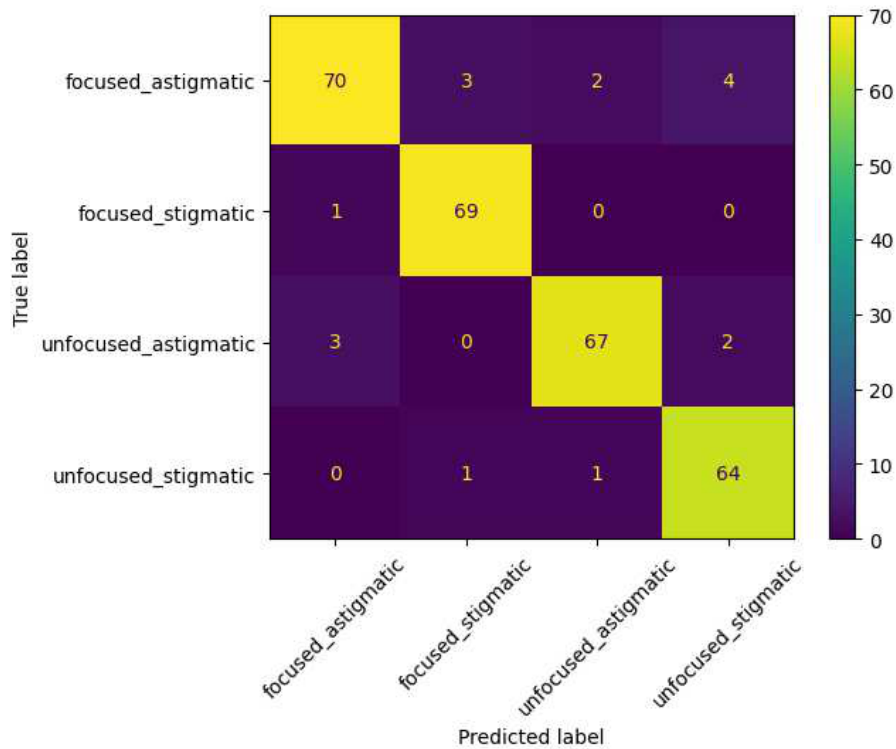


Fig. 3: Confusion matrix of the trained model on the test set across the four classes (focused or defocused × stigmatic or astigmatic) (adopted from Skricilova, 2024)

Where the errors fall (Figure 3) tells the rest of the story. Focused-stigmatic images, the easy case, were classified correctly in almost 100% of cases. Astigmatism was the hard part: images containing it accounted for more than 82% of all misclassifications. The single most common mistake was a focused-astigmatic image read as defocused-stigmatic, the pair that looks identical in the first shot and can only be told apart from the second.

4 DISCUSSION

The FFT-based preprocessing layer appears to help because it changes the representation of the input images before the convolutional layers process them. Defocus is associated with a loss of high-frequency information, while astigmatism can appear as an elongated pattern in the frequency spectrum. This is also why Fourier-based representations have been used in classical focusing and astigmatism-correction methods (Ogasawara *et al.*, 1999; Lu, Zhang and Li, 2018). In this study, adding the FFT layer substantially improved the base model across all tested fusion strategies: the best non-FFT variant reached 65%, whereas the best FFT variant reached 94%. This should be interpreted as an empirical result for the tested configurations, not as proof that the FFT layer alone would always produce the full difference.

A notable result was that the small custom model outperformed the tested pre-trained ImageNet models, and that the Fourier layer improved the custom model but not the best-performing pretrained configurations. A plausible reading is that ImageNet pretraining teaches a network to

chase fine natural-image detail that is not present in a frequency spectrum of a defect. The deep networks keep looking for structure the transformed input has thrown away, whereas a shallow model with few parameters takes the spectrum at face value. This fits the wider observation that a purpose-built CNN can match or beat heavier machinery on a narrow task like blur detection (Szandała, 2020). It is a useful caution against reaching for the largest pretrained model by default.

Two images are the cost of asking a richer question. Single-image spectrum methods (Lu, Zhang and Li, 2018) need only one shot but cannot pull apart a combined astigmatism-and-defocus state; the pair makes that separation possible at the price of a second exposure. Data scarcity was the other constraint: the model overfitted early, and only enlarging the dataset fivefold together with augmentation and L2 regularization brought it under control. L2 alone was the most effective single regularizer but not enough on its own.

The 94% should be read with one caveat. The study used a single sample type, tin balls on carbon, whose round regular shape makes defects stand out. On an irregular natural sample the Fourier layer might read the sample's own shape as astigmatism, so the figure holds for this well-behaved specimen rather than for arbitrary samples. The model is presented as suitable for the regression-testing framework; its behaviour once deployed there was not measured here.

5 CONCLUSION

A three-convolution CNN using image pairs, channel-axis fusion and a custom Fourier-transform layer achieved 94% test accuracy when classifying SEM images into four focus and astigmatism classes. This result indicates that the proposed approach is promising for integration into an automated regression-testing framework, although the study was limited to one sample type and the behaviour of the classifier inside a deployed testing workflow was not measured. The tested FFT variants substantially outperformed the corresponding variants without FFT, and the small custom model also outperformed the three tested pretrained ImageNet models. Useful next steps are training on more varied and irregular samples, adding further defect classes and evaluating the classifier in the target testing framework.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Skricilova, 2024) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

- AGGARWAL, C. C. 2018. *Neural Networks and Deep Learning*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-94463-0>
- AMMANN, P., OFFUTT, J. 2008. *Introduction to Software Testing*. Cambridge: Cambridge University Press. ISBN 978-0-521-88038-1.

- ERASMUS, S. J., SMITH, K. C. A. 1982. An automatic focusing and astigmatism correction system for the SEM and CTEM. *Journal of Microscopy*. 127(2), 185-199. ISSN 0022-2720. <https://doi.org/10.1111/j.1365-2818.1982.tb00412.x>
- HE, K., ZHANG, X., REN, S., SUN, J. 2016. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770-778. ISBN 978-1-4673-8851-1. <https://doi.org/10.1109/CVPR.2016.90>
- KINGMA, D. P., BA, J. 2014. Adam: A Method for Stochastic Optimization. arXiv preprint. *arXiv*:1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>
- KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*. 60(6), 84–90. <https://doi.org/10.1145/3065386>
- LECUN, Y., BENGIO, Y., HINTON, G. 2015. Deep learning. *Nature*. 521(7553), 436-444. ISSN 0028-0836. <https://doi.org/10.1038/nature14539>
- LEE, W., KIM, H., HAN, S., YOON, J. 2021. Robust autofocusing for scanning electron microscopy based on a dual deep learning network. *Scientific Reports*. 11(1), 20933. ISSN 2045-2322. <https://doi.org/10.1038/s41598-021-00412-5>
- LU, Y., ZHANG, X., LI, H. 2018. A simplified focusing and astigmatism correction method for a scanning electron microscope. *AIP Advances*. 8, 015124. ISSN 2158-3226. <https://doi.org/10.1063/1.5009683>
- NA, J., KIM, G., KANG, S.-H., KIM, S.-J., LEE, S. 2021. Deep learning-based discriminative refocusing of scanning electron microscopy images for materials science. *Acta Materialia*. 214, 116987. <https://doi.org/10.1016/j.actamat.2021.116987>
- OGASAWARA, M., SATO, M., KOMURO, M., KAGA, R. 1999. Automatic Focusing and Astigmatism Correction Method Based on Fourier Transform of Scanning Electron Microscope Images. *Japanese Journal of Applied Physics*. 38(2R), 957. ISSN 0021-4922. <https://doi.org/10.1143/JJAP.38.957>
- SCHUBERT, P. J., SAXENA, R., KORNFELD, J. 2024. DeepFocus: fast focus and astigmatism correction for electron microscopy. *Nature Communications*. 15, 948. <https://doi.org/10.1038/s41467-024-45042-3>
- SKRICILOVA, M. 2024. *Classification of electron microscope images using an artificial neural network* [in Czech]. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno. .
- SZANDAŁA, T. 2020. Convolutional Neural Network for Blur Images Detection as an Alternative for Laplacian Method. In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. pp. 2901-2904. ISBN 978-1-7281-2547-3. <https://doi.org/10.1109/SSCI47803.2020.9308594>

Contact information

Magdalena Skricilova: e-mail: Hugouse@seznam.cz

Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz