

Deep learning approaches on dealing with focus blur on optical images

Antti Cederlöf

Abstract

The growth of deep learning methods has proven useful in various image processing tasks such as localized classification and image-to-image translation. Recently, they have been applied in optical histological imaging, where there is an issue of short focal length which causes blurriness in thick samples. To enhance the contents of optical images, or to make them easier to inspect, deep learning methods have been developed that can be categorized in three classes: virtual refocusing, autofocusing, and optical sectioning. This project evaluates the current state of these methods and their generalization capabilities. In addition to a literature review, I performed tests on two algorithms, namely Deep-Z and COMI, with unseen optical image data of artificial beads.

Several fairly recent deep learning algorithms have been developed to perform the previously mentioned tasks. However, there does not seem to be many readily usable published models. The models that do exist can perform well with similar data as was used in training, but they are not intended to generalize on all histological images without re-training. Developing a robust algorithm that could do generalized deblurring on optical images is still a work in process.

1 Introduction and theoretical background

Volumetric information in imaging is essential in today's optical medical imaging methods [1]. In wide-field imaging, the volumetric information is not initially included because of the limited depth-of-field (DOF). Instead, the background and foreground behind and in front of the focal plane appear blurred. However, there is a great potential in deep learning to increase the DOF in optical wide-field imaging.

Deep learning approaches have recently been used in solving various inverse imaging problems including autofocusing, virtual refocusing, labeling, and super-resolution imaging [1]. These methods provide enhanced imaging performance compared to conventional image reconstruction and post-processing algorithms with minimal or no modification to the imaging hardware. These advancements show the possibilities of deep learning in biological image processing.

The first steps towards extending the depth of field with deep learning methods were introduced by Wu et al. in their paper considering holographic imaging [2]. They used a convolutional neural network called U-net, and they proposed that their approach could be used in other imaging modalities

as well. The rise of similar approaches due to that paper can be seen in the high number of other publishings citing it, including references of this project [3, 1, 4].

In this project I am reviewing some of the existing models for dealing with blurry optical images. In addition to a literature review, I will test some of the pre-trained models with new optical image data provided to me by my supervisor Birhanu Belay.

1.1 Neural networks

This project discusses approaches for blur detection and rejection using deep learning, which means that the algorithms consist of neural networks (NN). Neural networks learn a model from data, and they have recently been implemented in various tasks such as classification, regression, and image generation. Their recent increase in popularity is due to increased computational power and large available datasets.

In neural networks, there are multiple neurons connected to each other at different layers inspired by the structure of neurons in human nervous system. The input layer takes in the data for the NN, then the data propagates through hidden layers, and finally gets to the output layer that gives the output that can be compared to a target outcome. The neurons in different layers are connected with varying weights, which are updated by training the network with data samples. Each neuron also has a nonlinear function that maps the outputs of a neuron in a desired way.

Convolutional neural networks (CNN) are a type of neural networks that can learn global and local features by using convolution filters. They are especially useful in computer vision, and in medical field they are used for image processing applications. One popular CNN called u-net [5] was developed for biomedical image segmentation, and it has later been useful for other tasks as well. For example some of the algorithms discussed in this project [3, 6] use u-net as a backbone network structure.

1.1.1 U-net

U-net [5] is a widely used convolutional deep neural network in biomedical image processing. It was created for image segmentation in biomedical imaging to work with very few training images. The invention that differentiated u-net from the previous CNN algorithms was to have a localized output map rather than a single class label. Thus, the network was created to address the problem of poor localization of classification in the previously developed neural networks without needing a large number of samples or high computational power. Having different parts of the image classified separately is crucial in biological imaging processing for example in segmenting cells or determining the in-focus regions in an optical image. There can also be a limited amount of images available for such image processing tasks, so the fact that u-net can perform with low amount of image data is a great advantage.

In U-net, there is a contracting path followed by an expansive one forming a network that is shaped like the letter u, as in figure 1 [5]. This allows to detect properties of the image in multiple

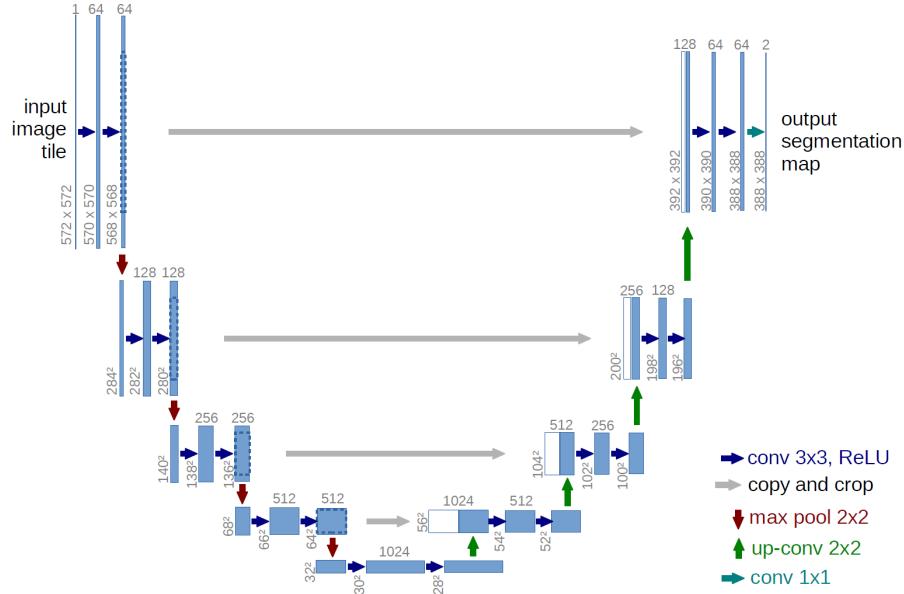


Figure 1: The u-shaped structure of u-net CNN [5]. The input in this case is a 572x572 sized image, and the output is two 388x388 sized maps that contain information on the localized segmentation. The input, output, and hidden layer shapes can differ between different realizations of the u-net.

levels of scaling while maintaining the local information by combining the high resolution images in the contracting path to the upsampled output. The contracting path includes repetition of two successive convolutional layers with rectified linear unit (ReLU) activation function followed by a max-pooling layer for downsampling. In each downsampling step, the number of feature channels is doubled. In the expansive path, there is an upsampling of the feature map and a 2x2 convolution halving the number of feature channels, a concatenation of a correspondingly cropped image from the downsampling path, and two similar convolutional layers with ReLu as in the contracting side. The final step is a 1x1 convolution to map the feature vectors into the desired shaped output.

1.1.2 Generative adversarial networks

Goodfellow et.al. [7] created a CNN framework called generative adversarial network (GAN), that consists of two models: a generator and a discriminator. The models are trained together in an adversarial manner resembling a two player game where they try to avoid mistakes while trying to get the other model to make them. The network has been proven useful in modern day image generation tasks, for example one of the refocusing algorithms described later in this chapter, COMI [8].

The generator is a model that learns a mapping from the input distribution to the output distribution. The discriminator, on the other hand, takes an input that is either the generator's output or a sample from the output distribution dataset, and it tries to determine which one it is. The generator

tries to give outputs resembling the target dataset, so that the discriminator could not tell that it is from the generator. Thus, the generator updates iteratively to output data that resembles the target dataset better, while the discriminator learns to differentiate the generated data from the real data.

One successful approach for GAN is image-to-image translation, where the generator learns the mapping from a source image domain to a target image domain. It has been used for example in creating different styles of paintings from images or generating an object from its outline. [9] The generator in an image-to-image GAN is a model that takes an image, learns the mapping, and outputs another image, so a network like u-net is a suitable option.

1.2 Methods for reducing blur

Several methods have been developed to increase the depth information and reduce the effects of blur with different approaches. In this report, I will categorize the methods in three classes: optical sectioning, refocusing, and autofocusing.

Optical sectioning refers to suppressing the blurred fore- and background information of an image [4]. With this approach, the unwanted blurry information is simply removed from the image. Refocusing methods construct new virtual images at different focal planes using a single [3] or a few [1] partially focused images. Autofocusing is an approach to create a focus map that marks different regions of the image as in-focus or blurred with a distance to the focal plane. This information can then be used to modify the image by either taking a new photo with better focus or using for example a refocusing software. These different approaches on the problem can be used situationally for different needs in defocused images.

1.2.1 Optical sectioning

Optical sectioning is the process of acquiring a clear focal-plane image by suppressing back- and foreground information leaving only the focused parts in the image. Optical sectioning can be traditionally achieved by using specific illumination techniques such as laser scanning confocal microscopy and structured illumination microscopy (SIM) [6]. However, these methods require specific devices with proper illumination that are not included in every conventional optical imaging system. Additionally, there are optical sectioning methods that use deconvolution and do not need any special devices. The problem in convolution based methods is that they are affected by noise, or alternatively they are oversimplified while trying to reduce noise. [4]

Deep learning approaches for optical sectioning can be trained to perform a cross modality image translation from wide-field optical images to corresponding optically sectioned images such as SIM or laser scanning confocal images. They aim to suppress the out-of-focus information of wide-field images without any additional hardware and without compromising on oversimplification or noise.

Optical sectioning methods are useful for example in visualizing a 3D stack of images where the objects are in focus at different levels. Each image thus contains in-focus and out-of-focus information

of the different objects. Optical sectioning improves the depth quality of such image stacks making it easier to separate the objects present in different depths.

1.2.2 Refocusing

Refocusing refers to virtually changing the focal plane of an image so that the defocused parts in the original image are translated into focus. This essentially means virtually changing the focal plane of an image. One approach is to use one or several wide-field images of a thick object with only some parts in focus, and use that to virtually generate a stack of images at different focal planes with different parts in focus [3, 1]. Another approach is to have a fully out-of-focus image of a thin sample or a surface that is then completely refocused into the right focal plane [8].

Wu et al. [3] created a deep neural network, Deep-Z, to create a stack of refocused images at different user-defined focal distances compared to a single wide-field image including objects at different depths. The result is an image stack of generated images at the focal planes defined by the user. Huang et al. [1] continued the research and developed a network, Recurrent-MZ, that takes in several images and improves upon the former's results. Recurrent-MZ also includes optical sectioning to reject defocused data on the image stack while the Deep-Z algorithm aims to recreate the images as they would have been if the photo was taken with different focus.

Zhang et. al. [8] developed a refocusing algorithm, namely COMI, that gives a single output image where the object is in focus and the blurry background is rejected. The algorithms seems to essentially perform both refocusing and optical sectioning. The authors tested the algorithm with public image data and they state that it has excellent generalization capability.

1.2.3 Autofocusing

Autofocus methods estimate the defocus distance of an object. Thus, they only give information on the blur, but they do not correct it by redoing the focusing or discarding the defocused parts. They can instead be used as a guideline for camera optics to make adjustments in focusing.

Deep learning has also been successfully implemented in autofocusing algorithms. Liao et. al. [10] created a deep-learning based autofocusing method that takes a single image as an input and uses that to predict the defocus distance. Conventional autofocusing methods have to use additional hardware such as additional illuminations or multiple shots at different focus. The authors of [10] created a fast autofocus algorithm without any additional hardware requirements so that it could be used for example in augmented reality microscopy, where the system gives information of the object in real time.

2 Materials and methods

I found two relevant publications [3, 8] for this project where the authors published their applicable algorithms. I tested them on optical image data of artificial spherical beads that resemble cells at

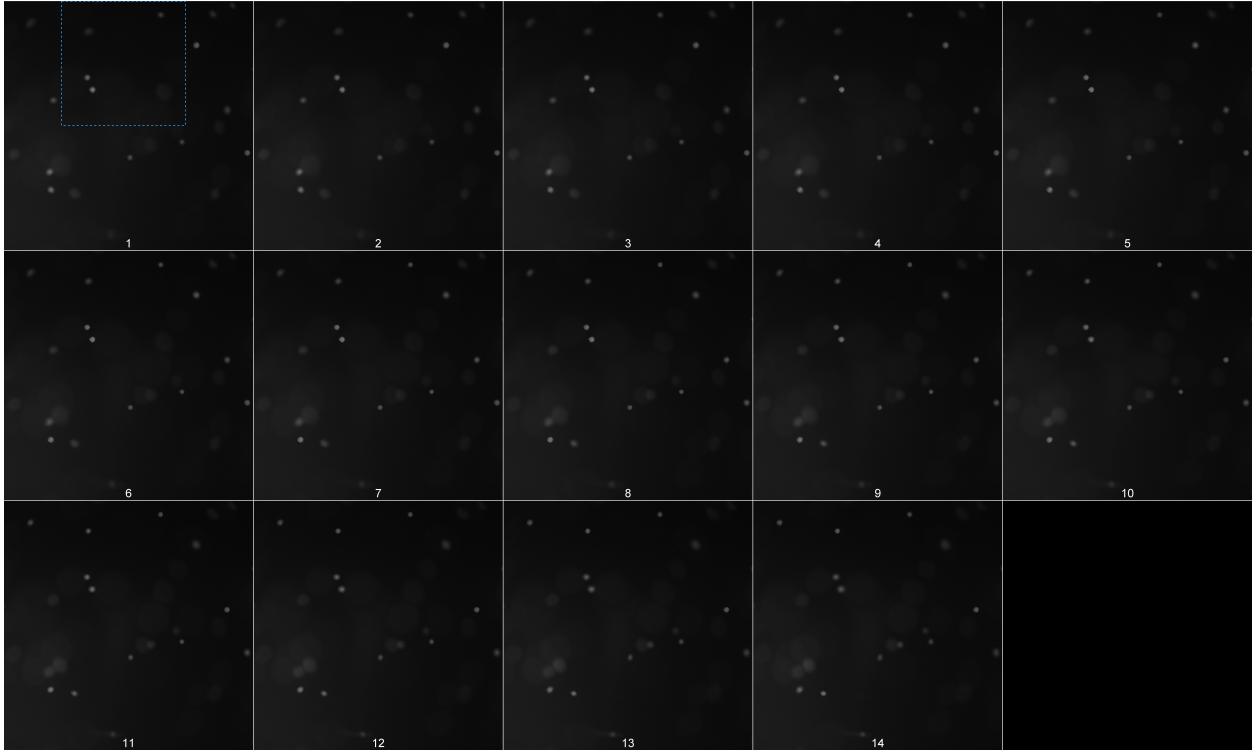


Figure 2: Images of fluorescent beads at 14 different focal distances. These images are used in the tests in this project.

14 different focal planes as shown in figure 2. The tests are done on pre-trained algorithms provided by the authors, so the results should not be as good as they would if the algorithms were trained by similar images that I use for testing. The goal of these tests is to see how well the algorithms perform and generalize on optical image data that is different from the data that was used in training.

The algorithms that were used on testing are Deep-Z and COMI. They were introduced in section 1.2.2, and I will go briefly through their technicalities in the following parts. Their detailed descriptions can be found in the original papers [3, 8].

2.1 Deep-Z

The Deep-Z algorithm is a least squares GAN framework [3]. The generator part is a U-net inspired CNN with ReLU nonlinearity functions. The discriminator is a CNN of six consecutive convolutional layers with leaky ReLU nonlinearity, a flattening layer, and a fully connected output layer that produces values within $(0, 1)$.

The authors [3] trained the Deep-Z model with fluorescent images axially focused at different depths and the corresponding ground truth fluorescence images at the target focus plane. They tested the algorithm on fluorescent images of a *Caenorhabditis elegans* worm that were also used in my example test shown in figure 3. The images were taken with different objective lenses, and there

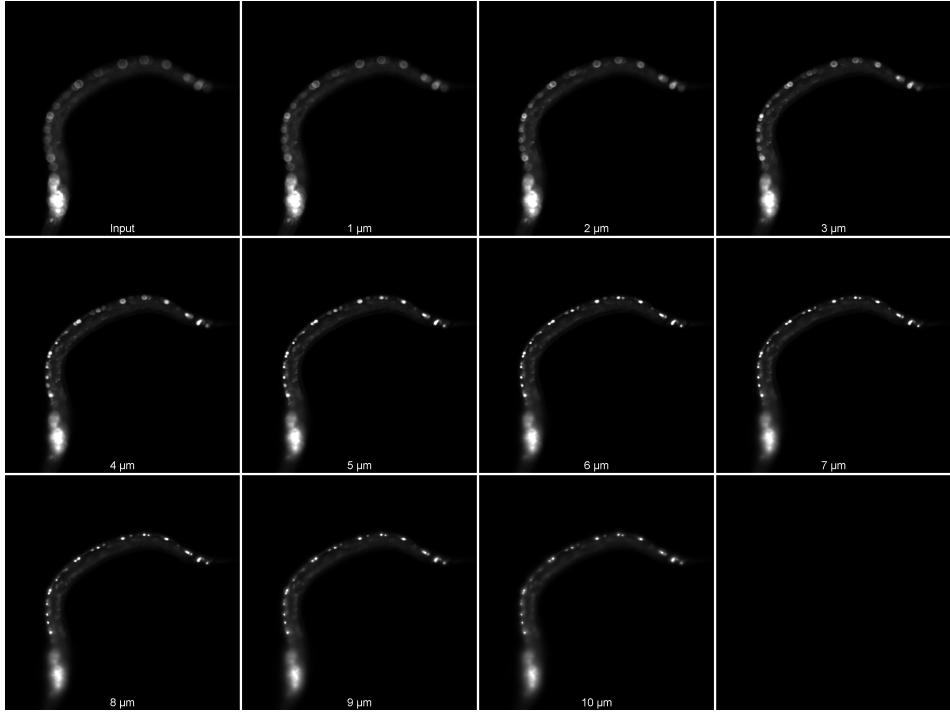


Figure 3: Example result from the Deep-Z algorithm using fluorescent images of a *Caenorhabditis elegans* worm [3]. Input image in the top left corner followed by digitally refocused images at different focal distances from the input image ($1 \mu\text{m}$ – $10 \mu\text{m}$).

is a trained model for each lens configuration used. I tried each lens configuration model to see how they work on the test data in figure 2. The Deep-Z algorithm is provided by the authors as a ready to use ImageJ plugin only requiring the user to input a 1024×1024 image and follow the instructions.

Deep-Z performs digital refocusing, resulting in images with different focal planes. The result can be a single image at a specific focal distance from the input image, or a stack of images at different focal distances. An example result of the Deep-Z algorithm with the original data from the authors is shown in figure 3.

2.2 COMI

COMI is inspired by a cycle GAN network [9], that has two mirror-symmetric GAN networks both with their own discriminator and generator [8]. One side creates an in-focus image from an out-of-focus source image and tries to discriminate the result from in-focus target domain images. The other side does the opposite by generating an out-of-focus image from the in-focus target image, and discriminating the result from the out-of-focus source images. Essentially, the different sides learn to do the inverse operations of each other.

The COMI model was trained and tested with two datasets: Leishmania parasite and bovine pulmonary artery endothelial cell images. The images in the datasets have an uniform blur, meaning

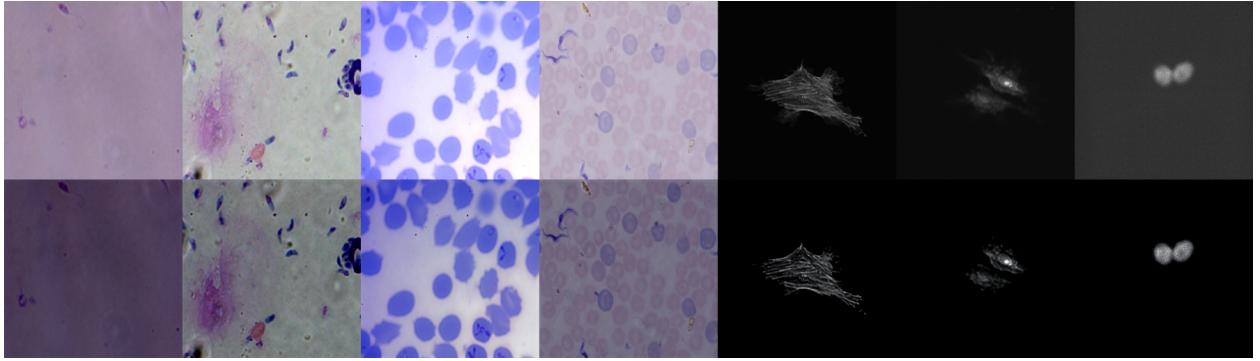


Figure 4: Example performance of the COMI algorithm using a collection of images from the dataset provided by Zhang et. al. [8]. Top row: input images, bottom row: corresponding COMI outputs.

that the whole image is either in-focus or it is evenly out-of-focus. However, my test images have beads at different heights, so the blur is not uniform, as can be seen in figure 2.

The COMI algorithm is published as a python code in the authors' github page [11]. It includes separately training a model and testing one. The authors also provide a pre-trained model, which I will be using in this project. I got the code in the file *test.py* working for me with a few modifications:

- set the correct path of the images in variables *set* and *z_depth*,
- set all directory names for saving as desired,
- remove and skip all lines with *multi_gpu_model* (I have a single gpu system),
- since I do not have ground truth images, skip the parts in the method *test* where the ground truth image is downloaded and numerical tests are run.

COMI is trained to refocus an evenly blurry image into a single image where the whole content is in focus. Example results using data from the creators of COMI [8] is shown in figure 4.

2.3 Evaluation metrics

The purpose of reducing blurriness in histological images is to get a sharper yet realistic image which allows to better see the details of the target tissue. The blur reduction performance can be evaluated visually, but it can be subjective and a bit tricky. While the blurriness and sharpness is quite easy to see, the correctness of details is not so clear to evaluate, especially if there is no ground truth data. The evaluation would become easier with ground truth data, but it is still not straightforward. Simple error measures such as mean squared error do not describe the level of blur reduction that well, since we are not interested in the exact pixel values, but rather restoring the structure in detail.

Deep-Z results are quite clear to assess with just visual evaluation. COMI results will additionally be evaluated using two commonly used image quality metrics that were also used in the original COMI paper: structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) [8].

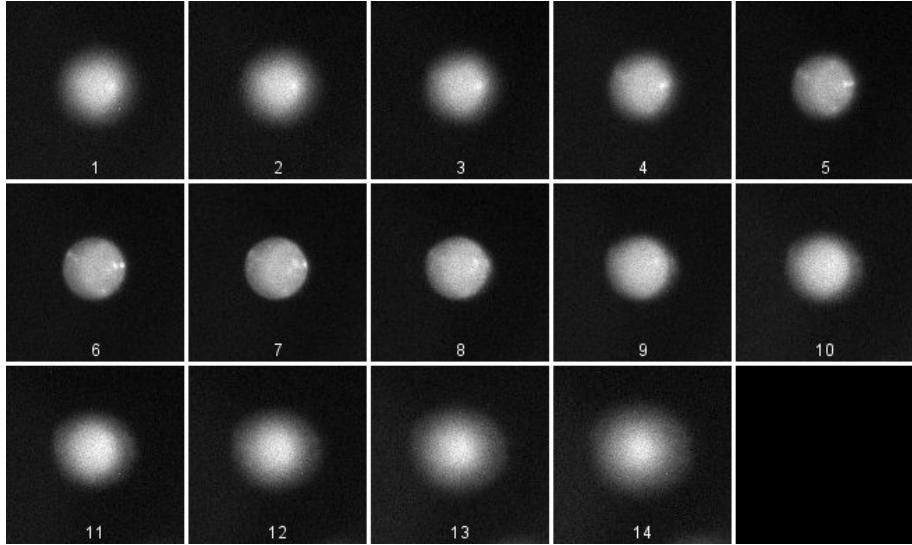


Figure 5: Test data in figure 2 cropped to only one bead. Used in tests for COMI.

SSIM is a metric used to measure perceived similarity. It was developed to correlate to the human visual system assuming that human perception is adapted for extracting structural information. [12] PSNR is a mathematically simple metric that expresses the ratio between the maximum value of the object and noise. It does not assess the structure as well as SSIM [12], but it works well assessing the quality of noisy data [13]. These metrics are used to evaluate the COMI algorithm performance using cropped images of the test data with a single bead shown in figure 5, so that the effects of changing background and beads at different depths is minimized.

3 Results and discussion

3.1 Deep-Z

As stated before, the Deep-Z plugin comes with several pre-trained models for different lens setups. The performance of the four available models is shown in figure 6, where the input images are cropped to 1024x1024 as shown in the first image of figure 2. The input image in those tests is image 7 in figure 2. As seen in the model comparison, the models learn to refocus differently on different lenses, so it would be best to use the correct model for testing images taken with the specific lens. However, the authors of [3] did not provide the code to train a model, so in this project the following tests are done with the model *Leica_20X0.8_worm_TxRd* which based on several tests using the test images visually seems to refocus the images most realistically.

Since the lens configuration of the model and the test images do not match, the focal distances are not consistent in the Deep-Z model and in the test data. It follows that a $1\mu m$ difference according to the model would not necessarily mean a similar difference in the focal distance of the images.

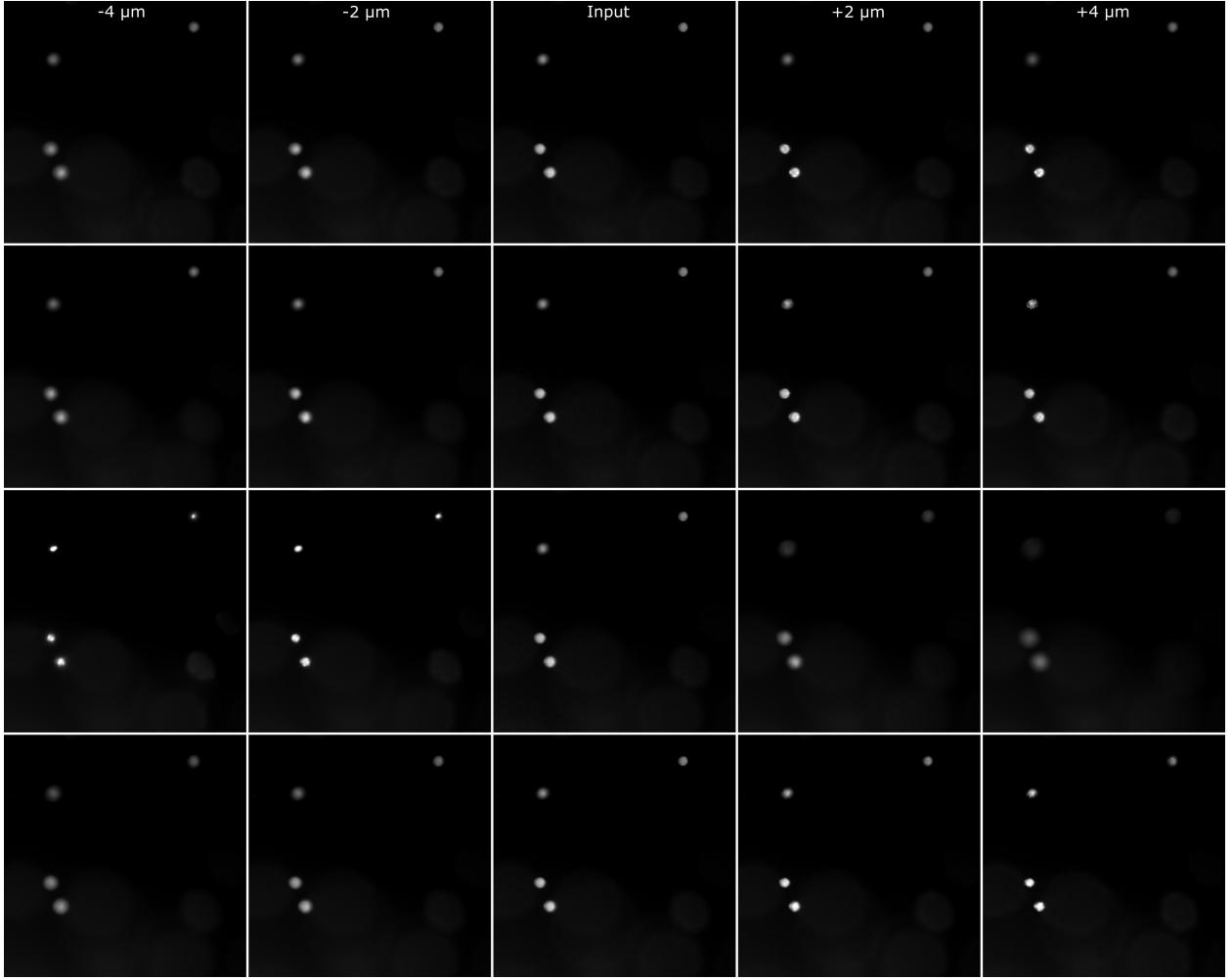


Figure 6: Comparison between the four pre-trained models of Deep-Z algorithm with different lens setups [3]. From top to bottom: Leica_20X0.8_worm_FITC, Leica_20X0.8_worm_TxRd, Leica_40X1.6_worm_TxRd, and Olympus_20X0.75_worm_TxRd.

Therefore, it would not be easy or even reasonable to do numerical measures between the virtually refocused images and the test images. Another reason why numerical measurements are not applicable is the fact that the algorithm doesn't seem to know the direction of the focal plane, as can be seen in figure 7.

The refocusing results of the Deep-Z algorithm are shown in figure 7. There the inputs for the algorithm are images 3, 7, and 11 in figure 2, cropped as shown in the first image in that figure. While the algorithm seems to successfully refocus the beads and reduce blurriness, there are some problems with the refocusing.

Figure 7 shows that the algorithm does not know which direction, + or - in the figure, is the right one for refocusing. Instead, the algorithm seems to determine a direction in depth where all the blurry

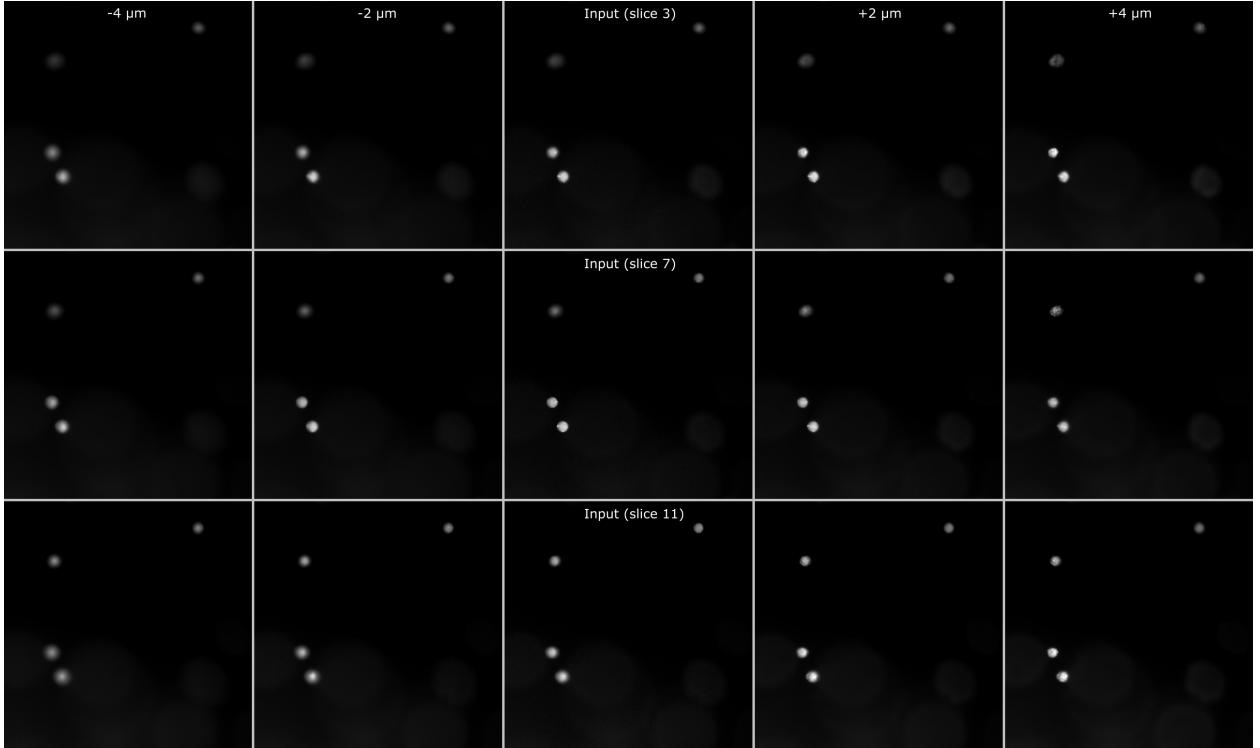


Figure 7: Refocusing results with the Deep-Z algorithm [3]. Three cropped input slices: 3, 7, and 11 from figure 2, and their refocusing results at distances $-4\mu m$, $-2\mu m$, $+2\mu m$, and $+4\mu m$ defined by the pre-trained model. Each row shows the refocusing using the corresponding input image in the middle column.

objects get more in focus. That can be seen in the first and third rows of the figure, where the two beads on the bottom left parts of the image are both blurry in the input image, but the planes 3 and 11 are at the opposite directions of the in-focus plane of the beads, which is approximately at slice 7 as seen in figure 2. Still in both cases, the algorithm tries to focus the beads on the same direction, as the beads appear more in-focus in the $+2\mu m$ and $+4\mu m$ images. The model does, however, detect that those beads are in-focus in input slice 7 and it does not try to deblur them any further.

The algorithm seems to also have some issues with the refocusing of a bead. As can be seen in the $+4\mu m$ images in figure 7, the beads are reconstructed with small dots inside them. While the beads do have an uneven surface, the algorithm seems to exaggerate it. It looks like the model is trying to do similar reconstruction as with the authors' data in figure 3, where the images do in fact have a structure with smaller dots inside the object.

The problem with creating dots inside the beads is caused by the model trying to find smaller structures than there actually is in the image. Downscaling the input image instead of cropping, and using the model with the highest magnification as shown in figure 8 gets rid of this problem as the model is apparently used to objects of that scale. This shows that the algorithm is not scale-invariant

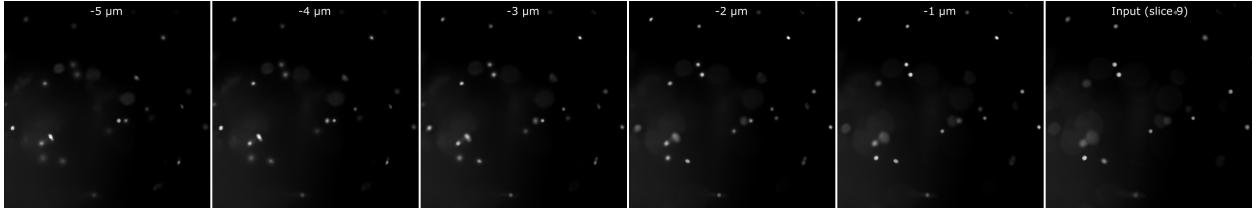


Figure 8: Deep-Z results using model Leica_40X1.6_worm TxRd and a rescaled 1024x1024 test image from figure 2 as an input.

and further indicates that training a new model for each lens system would be required.

3.2 COMI

Results using the COMI algorithm are shown in figure 9 using the test images, and in figure 10 using cropped test images of only one bead. By comparing the figures 2 and 9, one can see that the algorithm reduces blurred background information essentially performing optical sectioning. While the optical sectioning increases the observed quality of the image, there seems to be no significant refocusing of the beads.

The results using SSIM and PSNR are shown in the graphs in figure 11. The graphs show the SSIM and PSNR values of the test images in figure 5 as *inputs* and in figure 10 as *comi outputs* compared to slice 7 in figure 5, that is visually evaluated to be the most in-focus. These numerical tests show that the COMI algorithm indeed does not do any notable virtual refocusing, since both the SSIM and the PSNR values are significantly higher on the unprocessed input data than they are using the outputs of COMI. This only shows that the COMI outputs are less similar than the input images at different focuses due to the reduction of background.

The effect of the COMI algorithm can be further inspected on the profile graphs of the beads at their center point shown in figure 12. They show that the COMI output has similar width and other properties, shown as shapes and notches on the graphs, as the input images but with higher contrast and more even background. This also shows that the algorithm does optical sectioning rather than virtual refocusing.

4 conclusion

How blur appears in optical images seems to vary in different imaging setups and it is difficult to model generally in a deep learning deblurring model. Therefore, the current algorithms work better on similar data that was used in training it. Before achieving good results with these algorithms on a specific data, a model should be trained with similar data. This is usually time consuming and it requires very high computational power to train the deep neural networks.

The problem of deblurring optical images with deep learning approaches is an ongoing research

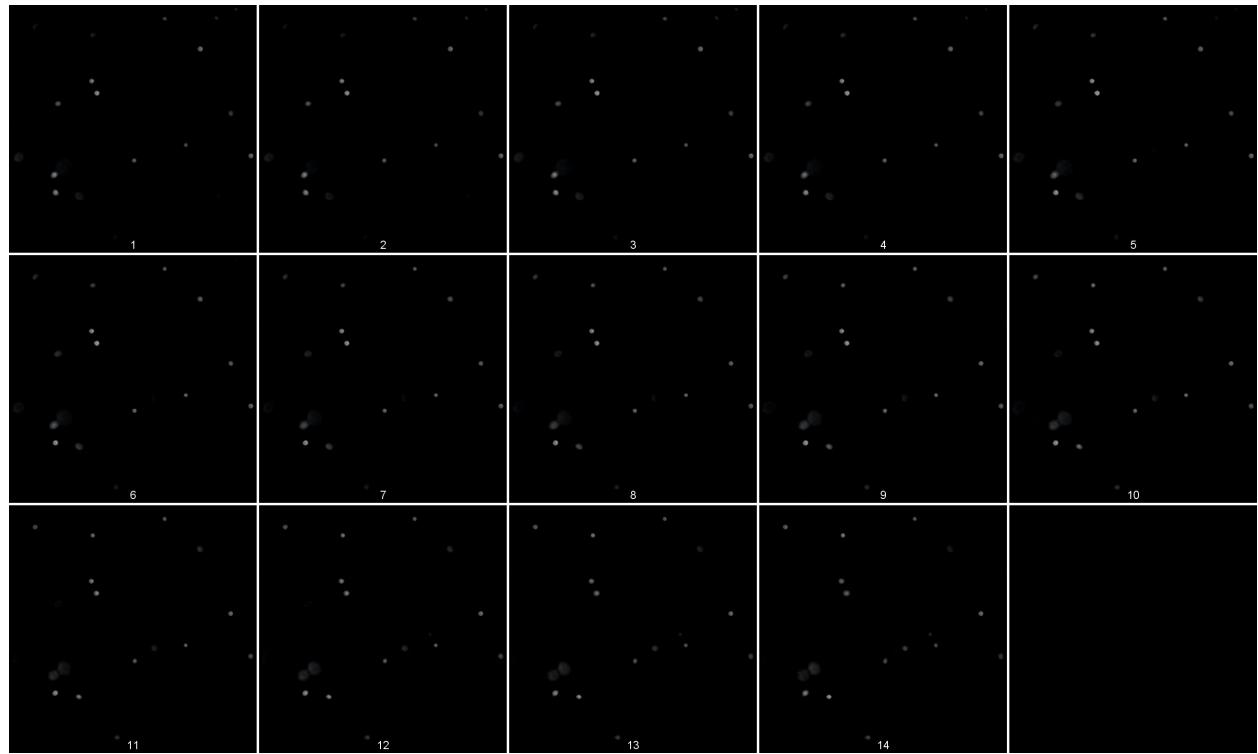


Figure 9: Results of COMI algorithm using the test data in figure 2.

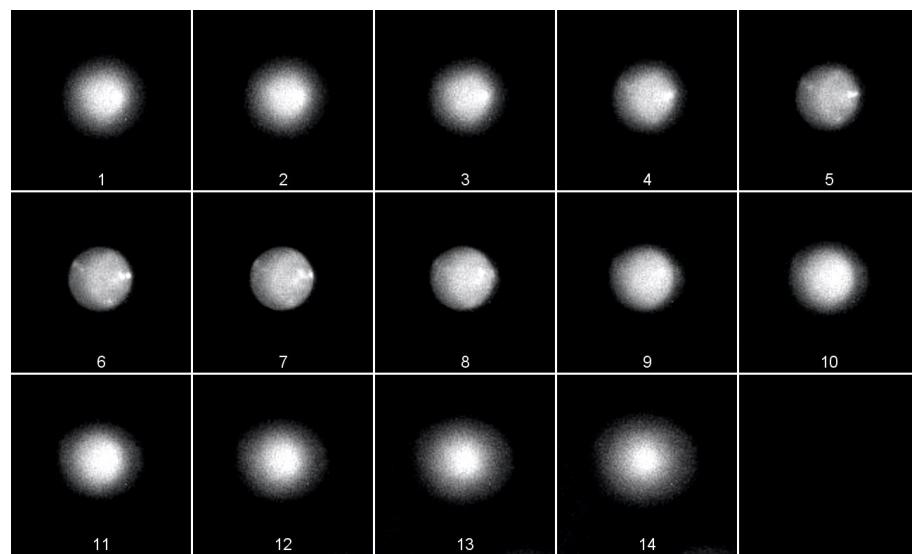


Figure 10: Results of COMI algorithm using the cropped test data in figure 5.

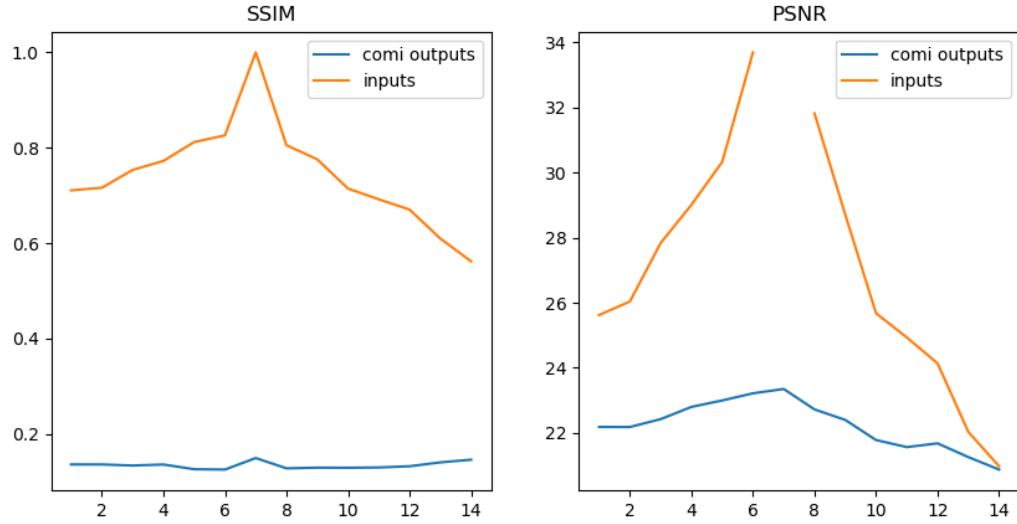


Figure 11: Graphs showing the SSIM and PSNR values of the test (input) images and their corresponding output images from the COMI algorithm at different focal distances compared to the in-focus slice 7 in figure 5.

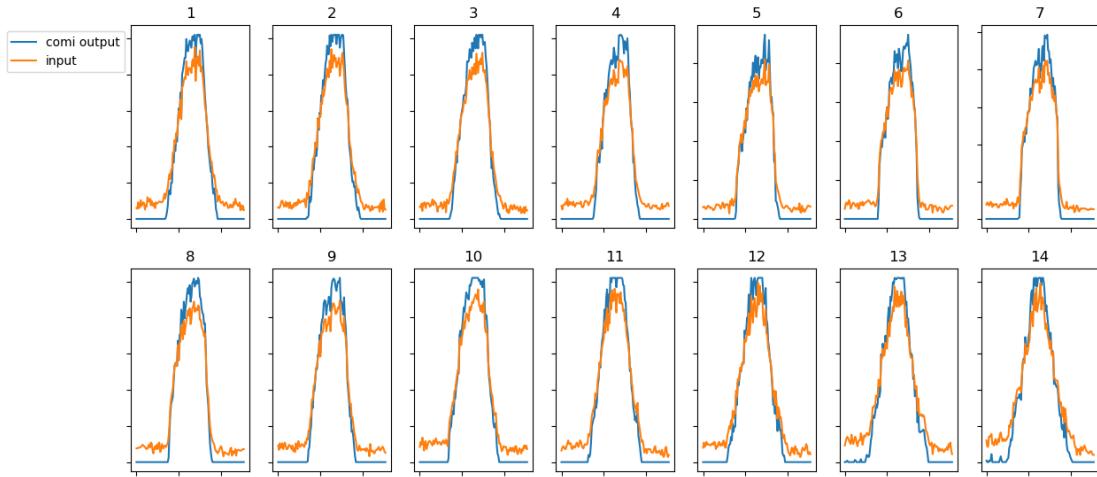


Figure 12: Grayscale value profiles of the test (input) images and COMI output images of a single bead at different focal distances.

topic. Many recent models have been developed to solve that, but there seems to be no functional pre-trained deblurring algorithm available that could generalize well on any unseen data. The algorithms tested in this project did some deblurring on the unseen test data, but the results were lacking compared to tests with the authors' data. The work has been going on beside the Deep-Z and COMI algorithms, for example in [1, 10], but the algorithms were not yet available for testing. Once there are more advanced algorithms available for testing, an extensive comparison on their generalization possibilities could be advantageous.

For me as a student, this work has been an important learning process on deep neural networks. It has taught about the use of advanced deep learning models in large and with physiological samples. After doing this project I am more prepared to create similar networks of my own to solve complex issues and inspect the usability of data and results.

References

- [1] Luzhe Huang, Hanlong Chen, Yilin Luo, Yair Rivenson, and Aydogan Ozcan. Recurrent neural network-based volumetric fluorescence microscopy. *Light, science applications*, 10(1):62–62, 2021.
- [2] Yichen Wu, Yair Rivenson, Yibo Zhang, Zhensong Wei, Harun Günaydin, Xing Lin, and Aydogan Ozcan. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica*, 5(6):704–710, 2018.
- [3] Yichen Wu, Yair Rivenson, Hongda Wang, Yilin Luo, Eyal Ben-David, Laurent A Bentolila, Christian Pritz, and Aydogan Ozcan. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nature methods*, 16(12):1323–1331, 2019.
- [4] Xiaoyu Zhang, Yifan Chen, Kefu Ning, Can Zhou, Yutong Han, Hui Gong, and Jing Yuan. Deep learning optical-sectioning method. *Optics express*, 26(23):30762–30772, 2018.
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241, Cham, 2015. Springer International Publishing.
- [6] Huimin Zhuge, Brian Summa, Jihun Hamm, and J. Quincy Brown. Deep learning 2d and 3d optical sectioning microscopy using cross-modality pix2pix cgan image translation. *Biomedical optics express*, 12(12):7526–7543, 2021.
- [7] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv.org*, 2014.
- [8] Chi Zhang, Hao Jiang, Weihuang Liu, Junyi Li, Shiming Tang, Mario Juhas, and Yang Zhang. Correction of out-of-focus microscopic images by deep learning. *Computational and structural biotechnology journal*, 20:1957–1966, 2022.
- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV)*, volume 2017- of *IEEE International Conference on Computer Vision*, pages 2242–2251, NEW YORK, 2017. IEEE, IEEE.
- [10] Jun Liao, Xu Chen, Ge Ding, Pei Dong, Hu Ye, Han Wang, Yongbing Zhang, and Jianhua Yao. Deep learning-based single-shot autofocus method for digital microscopy. *Biomed. Opt. Express*, 13(1):314–327, Jan 2022.

- [11] Chi Zhang, Hao Jiang, Weihuang Liu, Junyi Li, Shiming Tang, Mario Juhas, and Yang Zhang. Comi : Correction of out-of-focus microscopic images by deep learning. <https://github.com/jiangdat/COMI>, 2022.
- [12] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [13] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.