Semi-Synthetic Test Data Generation and Content-aware Image Restoration of Fluorescent Microscopy Images

Anonymous ECCV submission

Paper ID 100

Abstract. Although fluorescence microscopes theoretically allow imaging with resolutions at the diffraction limit, in practice the resolution is often drastically limited by several factors: Image quality of fluorescence microscopes is always limited by the interaction of the light with the tissue, often resulting in a low image intensity and weak contrast of the biological structure. Excitation of fluorophores and molecules outside the imaged area leads to photobleaching or phototoxic effects in the sample. In addition, diffraction effects and the low intensity of fluorescent light reduce the quality of the images. In this paper, we present an application of using deep learning to improve the image quality of a fluorescence microscope in biological imaging. We trained a predefined architecture of a deep convolutional neural network provided by the CSBDeep project with semi-synthetic training data of mitochondria and cardiomyocytes. The results show good improvement of the image quality.

Keywords: fluorescence microscopy, biological imaging, deep convolutional neural network

1 Introduction

Fluorescence microscopy has become established, especially in biology and the life sciences, as a technique for observing cellular and subcellular processes. In general, limitations in microscopy can emerge from the feasible acquisition speed, exposure time and achievable spatial resolution. For example, a very good image contrast can be achieved with an intensive exposure, but only at a high acquisition speed, because the sample is bleached faster. Additionally, the resolution of fluorescence microscopes is limited by interactions of imaging light source with the sample, which are leading to a decrease in image quality. The effect of photobleaching, caused by the finite number of fluorophores in the sample, which can lose their function under exposure to light, limits the amount of light energy that can be transferred to the sample. This again limits parameters such as exposure time or exposure intensity. Furthermore, the light is not only absorbed by fluorophores, but also by the molecules in sample. This can lead to phototoxic effects, which might destroy the sample. Hence, fluorescence microscopes make special demands on the exposure and detection path. The use of special

light sources is necessary and sensitive cameras are used for the detection of the low-intensity fluorescent light.

More generally, the resolution of a light microscope is limited by diffraction effects of light. Thus, a point light source, by diffraction at the imaging lenses, is represented in the image as a Airy disk. The three-dimensional intensity distribution of a point light source when imaging with an ideal lens is called point spread function (PSF). Images recorded with a light microscope are therefore always distorted. Mathematically, this image distortion corresponds to a convolution of the PSF with the real intensity distribution. The PSF of a microscope is completely determined by the numerical aperture of the objectives, the excitation wavelength and the refractive index of the immersion medium.

Further progress in the field of fluorescence microscopy was achieved with the development of the *confocal* microscope and the modern "STED" microscope. The latter even overcomes physical limitations in resolution [1]. A confocal microscope solves the above mentioned problems by using point illumination and inserting a pinhole aperture into the detection path. Fluorescent light that does not come from the focus or focal plane is blocked by the pinhole. The entire sample can be imaged by moving it through the focal plane. The individual images are then digitally concatenated.

In order to be able to evaluate even low quality images, it is possible to digitally restore images afterwards. Low quality images often occur in laboratory practice, when using standard fluorescent microscopes with comparably low resolution or if the imaging parameters cannot be optimized for the best contrast, to avoid photobleaching and phototoxicity. For the purpose of image restoration, a neural network can be trained on suitable training images to recognize and restore the underlying structure in image areas of poor quality.

In this article, we analyze the application of a neural network to subcelluar structures. More precisely, we make the following contributions:

- Because the generation of a large dataset in a biological setting is time consuming, we decided to set up two semi-synthetic datasets. We describe a method for the generation of semi-synthetic training data, which generates several different training data from each high-resolution *confocal* image. Therefore we added typical distortions to the high-resolution confocal images mimicking optical and sample-related resolution limitations.
- To improve image quality in post-processing, we trained two predefined content-aware image restoration (CARE) networks by the CSBDeep project [2] based on high resolution fluorescent images of a confocal microscope.
- The trained CARE networks are applied to previously unknown microscope images to verify training success.

Both trained networks lead to a significant enhancement of the images. By improving contrast, reducing noise and deconvolution of the image, the biological structures are displayed in greater detail. With the restoration of the images by self-learning algorithms, deficits in fluorescence microscopy are compensated afterwards.

2   Related Work

Microscopy is fundamental for observing cellular and subcellular processes. The challenge is to create a good contrast between the structure and the environment. The most commonly used contrast generation method in biology and life sciences is based on fluorescence. In this physical effect, a material absorbs light at a fixed wavelength and emits lower-energy light after a short time. This difference in energy results in a difference in wavelengths between the excited and emitted light. By using colour filters, a good contrast can be obtained by filtering out the excitation wavelength. Different proteins, the so-called fluorophores, are used as fluorescent materials. One example is the frequently used GFP (green fluorescent protein). Specific components of the cell can be tagged by the fluorescent protein, by encoding the gene for the protein into the cells.

The research question we address in this paper is how to improve the image quality by post-processing; in particular, how deficits in the observation of biological structures by fluorescence microscopy can be reduced with the help of artificial intelligence. A detailed summary of the development of neural networks (NNs) can be found in [3]. The algorithms of the 21st century have become much more powerful and efficient, although the basic principle of NNs has not changed much. The learning ability of modern NNs is limited by the performance of current hardware, especially GPUs. Open source projects such as Tensorflow and Keras make this new technology more accessible to people from other fields. Although deterministic algorithms for denoising and deconvolution of images already exist, the advantage of self-learning algorithms is a reduction in processing time and better restoration quality. The one-time effort to obtain a well-trained neural network is compensated by restoration quality of special structures.

Convolution Neural Networks (CNNs) are a class of neural networks, which receive a matrix of data as input. These can be, for example, images, which are constructed from $H \times W \times C$ (height $\times$ width $\times$ color channel) pixels. CNNs use special operations to reduce the information contained in an image. The goal of the processing is the recognition of repetitive structures and characteristics in the image. The recognized characteristics can then be classified so that a CNN can infer the image content from a combination of certain structures. CNNs are already widely used in image recognition and image restoration [4, 5]. In particular, the application of CARE networks for image restoration of fluorescence microscopy images is now widespread. The approach is suggested in [6] and [7]. An widely used approach for generating datasets is taking high-quality images with low exposure and subsequently with a long exposure to get a corresponding low-quality image as done in [8]. A new way by Krull et al. [9] waives on supervised learning for noise reduction and shows great results with a self-supervision network. Within the framework of the CSBDeep project, Weigert et al. developed an environment for digital image restoration of fluorescence images [2]. The aim of the project is to compensate for limitations in images taken with a fluorescence microscope by means of Deep Learning. Specifically, the authors provide, among other things, a Python package in whose environment a CNN can be trained. The CNN is based on the U-Net structure, which consists of a

special sequence of convolution levels, pooling levels and upsampling levels [10]. As output the network makes a probability prediction for the value of each pixel in the image. The architecture of the CNN is designed for biomedical applications and stands out from other CNNs in that it achieves very good results with only a few training images. An analysis of the training is done using Tensorflow. To restore microscopy images with a trained CNN, the CNN can be applied to the images via a Fiji plugin.

In this paper we want to test the approach on semi-synthetic training data for noise reduction and restoration of fluorescent images of two specific biological structures.

3   Methods

In this section we describe our approach for the restoration of fluorescence microscopy images using a CARE network. In the first part we report how the dataset was created and the second part the training process is depicted.

3.1   Semi-Synthetic Datasets

One of the main challenges on the way to a well trained CNN for image restoration is the generation of training data. The quality of the image restoration of a trained network depends largely on the quality and quantity of training data. The possibility to improve training data by increasing the number of training images can be very time-consuming, especially in biological applications.

In general, there are three approaches to obtain training images. The first and simplest approach is to take pictures with a microscope. In order to obtain a wide range of training data, acquisition parameters such as acquisition time and exposure intensity are varied. The problem with this method is that it can be very time consuming to acquire a sufficient amount of suitable images. Furthermore, noise and reduced resolution due to the used optical devices limit the quality of the training images. The second possibility is the simulation of the observed biological structures. For this purpose, ground truth images are calculated with a computer based on structural properties of the sample. These artificial training data have the advantage of being independent of the above mentioned limitations. However, if the structures are too complex, it turns out to be difficult to simulate them accurately. The bridge between these two possibilities is the third option: semi-synthetic training data. These training data are images from a high-resolution fluorescence microscope, which are digitally distorted afterwards. For this purpose, a series of transformations are applied to the ground truth images. First, the image is folded using the microscope's point spread function. This makes the image more blurred. Then, Poisson noise and Gaussian noise are added to the image. Finally, the intensity of the image is reduced. The advantage of semi-synthetic training data is that by changing the transformation parameters, such as the amount of noise and the intensity reduction factor, many different training images can be generated.

The generation of semi-synthetic training data by reproducing the image distortion of the microscope is more efficient than the other two options. Another aspect is the diversity of the training images: The training data should represent as wide a range of structures as possible; this can be easily taken in account while producing the training images by varying the distortion parameters.

In our application, the observed biological structures are the two following types:

– HeLa cells with fluorescently labeled mitochondria [11] (Fig. 2a).
– Cardiomyocytes (heart muscle cells) that have a sarcomere with a repeating, rectangular structure of intermediate z-discs (Fig. 3a). These z-discs are fluorescently labeled.

These two biological samples are chosen because they are used in current experiments on laser nanosurgery in one of our institutes that aim to observe the effect of the laser on the biological structures. Furthermore, both samples have a characteristic structure, which we want to resolve in detail.

We proceeded as follows to obtain semi-synthetic data. First, high quality images are taken with an confocal microscope from *Leica*. Parameters like exposure time, excitation intensity and recording time are changed to get a diverse training set. Excitation is realised with a Helium-Neon (HeNe) laser at a wavelength of $534\,nm$. The images are post-processed with Fiji [12]: The contrast is adjusted and a minimum function is applied to the image to remove acquisition noise. This minimum function determines the minimum value in a small area around each pixel, sets the pixel to this value and the rest of the pixels to the black level. Special care is taken to not to make structural changes to the data during post-processing. To this end, we ensured that the change in pixel values in an area is much smaller than the structural size of the mitochondria or z-discs. Depending on the magnification, an adjustment to 0.5 pixels to 4 pixels was made, which corresponds to a length of approximately $1\mu m$ to $5\mu m$.

Next, to generate the training data from the post-processed images, typical distortions of fluorescence microscopes are simulated by a Python script. To simulate realistic images from a fluorescence microscope, the following sequence of transformations is applied:

1. Each image is convoluted with the point-spread function of the microscope, which we calculated from the imaging objectives.
2. A Poisson noise is added to the image.
3. A Gaussian noise is added to the image. The standard deviation and the intensity of the noise are chosen randomly from an interval.
4. Afterwards the image intensity is randomly modulated by a factor which is selected from a predefined interval. Simulating over- and underexposure corresponds to a factor greater than one or between 0 and 1, respectively.

From the original data, three sets with the distorted training data are created. For each set the interval for the random parameters, which are the Gaussian noise sigma and intensity as well the intensity modulation factor, are shifted. The

suitability of the training images is then estimated, based on prior experience on microscope images.

The microscopic ground truth images of the mitochondria are of size 1024×1024 px. From 78 of confocal images we create 234 corresponding semi-synthetic training images. In a similar manner, the cardiomyocyte dataset consists of 31 microscopic ground truth images of size 512×512 px from which 93 training images are generated.

The generated images are now made compatible with the CSBDeep environment. This requires the use of two functions provided by the CSBDeep package. The function *raw data* re-reads the ground truth images and associated training images and converts them into an internal file format. By the function *create patches*, the training images are divided into small areas (patches). As parameters of the function, it is specified how many areas are obtained from an image and their corresponding size. These areas are fed into the neural network as input during training. The choice of the patch size affects the training quality and should be chosen according to the size of the observed structure. For our datasets we set the patch size to 32×32 px for the mitochondria and to 16×16 px for the cardiomyocytes.

4 Training and testing

CNNs are composed of layers. The input and output of each layer are feature maps, each of which represents a specific property in the image. Each layer is divided into three operations, which are performed on the maps. Specifically, each layer consists of a filter bank, a nonlinear activation function and a pooling operation. A CNN is typically composed of several combinations of these three operations. The use of several levels enables the network to learn the hierarchies found in images. For example, an object in an image consists of different image parts, image parts in turn consist of different motifs and the motifs consist of combinations of edges. When a CNN is applied to an image, the image is searched for combinations of structures by scanning the image with characteristic matrices. Certain values in the output matrices therefore represent different structures. These values are classified during training and can be used to draw conclusions about the objects in the image if training is successful.

The training data generated above are read in again. During a training epoch, CNN propagates all training data through the different levels and performs backpropagation on all training data. Each epoch is divided into smaller steps (iterations). In each training step, a subset of the data of a specified batch size is propagated back and forth through the network. The choice of the number epochs, the steps per epoch and the batch size is therefore decisive for the training. The overall goal of the training is to minimize a specified loss function by varying the weights. The learning rate indicates the adjustment factor of the weight functions of the neural network. The learning rate is reduced by the network in the course of the training. The aim is to select the parameters in such a way, that a converging network is obtained. A converging network is charac-

terised by the fact that a generalisation is made to the general structure (of mitochondria and Z-discs) on the basis of the given training data.

We obtained optimal settings for training the CARE network on our datasets by a grid search trying different parameters. We reached an optimal setting at 1000 epochs with 500 steps on both datasets with a batch size of 16. As mentioned above, the input are patches of size $32 \times 32px$ for the mitochondria and patches of size $16 \times 16px$ for the cardiomyocytes. The training is performed using the Adam optimizer with an initial learning rate of 0.0002 (mitochondria dataset) and 0.0004 (cardiomyocytes dataset). The datasets are split into 80% training and 20% validation data. Evaluation of the networks performances is achieved by a combination of evaluation of the training history with Tensorflow, intensity profile comparison and visual review of neural network enhanced images. Training is performed on a *NVIDIA GTX 1080* GPU.

5   Results

In this section, the training performance of both networks is evaluated. The control of the training for the two different cases is done by an analysis of the training history. The training history is a graph showing the performance of the network over the training duration. The analysis is done by comparing the performance of the network on the training data and the on validation data in terms of the mean absolute error (MAE) on intensity values. Furthermore, the development of the loss function can be examined. The loss function differs from the error function, because we use a probabilistic CARE model. Since the global minimum of the loss function is searched in the training, a convergence should be attained here as well.

Looking at the training history of the mitochondria in Figure 1a–1c, the network converges well with increasing training time. The mean absolute errors converges to 0.046 on the training and validation set, while omitting over- and underfitting of the training data. The learning rate is reduced in steps from 0.002 to zero, resembling that the adjustment of the weights is decreased and that the number of epochs is sufficient.

The training history for the cardiomyoctes dataset shows similar characteristics (Fig. 1d–1f). Here, the mean absolute error reaches an approximate value of 0.07. The error function and the loss function are converging on the training and validation set.

Further evaluation is done by comparing the restored images with training images. The image sequence in Figure 2 shows the result of the training for mitochondria and in Figure 3 for the cardiomyocytes. By visual comparison, the restored images approximate the original image well. A significant reduction in noise can be observed. Furthermore, the contrast, especially in parts of the image with lower intensity, has been improved.

This first impression is quantified by comparing the intensity profile (see Figure 4). The intensity data can be determined with the image processing program Fiji by extracting the intensity of the individual pixels along the marked

horizontal lines in Figures 2 and 3. The differences between the intensity pro-
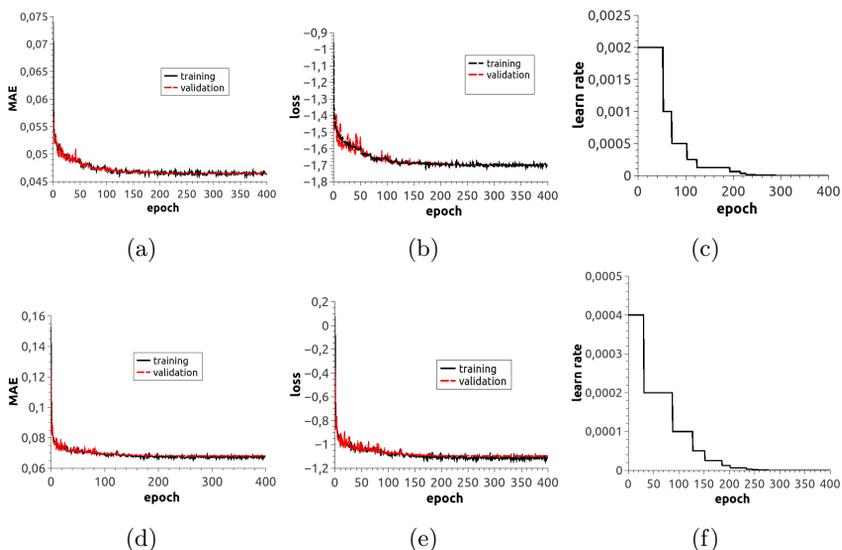

Fig. 1: The averaged absolute error, the value of the loss function and the learning rate are plotted over the training epoch for the mitochondria dataset (Figure 1a to 1c) as well for the cardiomyocyte dataset (Figure 1d to 1f). The networks clearly converge with increasing training time.

files of the ground truth picture and the training picture are clearly visible. Due to adding of digital distortion, characteristic features of the profile vanish. Especially structures with low intensities are no longer clearly identifiable. This corresponds to the real conditions of a fluorescence microscope. The restored images obtained by the CARE network again show a sharp profile. Hence, we confirm quantitatively the significant reduction of noise in the restored images.

Especially impressive is the agreement with the profile of the original image. The position of almost every maximum could be restored. The intensity values of some maxima is increased, which is represented in the restored picture by a higher contrast of structures with overall lower intensities. Noticeable is, however, that the peaks are slightly widened. Visually, this creates the impression of a slightly blurred picture. The evaluation of further sets of corresponding ground truth, training and restored pictures led to similar results.

We evaluated the performance of the network on images, which were not included in the dataset. Images for testing the trained neural networks are acquired with the *Zeiss Axio Observer D1*, extended by a multi-photon excitation laser. Exemplary image pairs for each network are shown in Figure 5 and Figure 6. The output of the network can be improved by reapplying it a second time. For this purpose, an image already processed by the network is entered
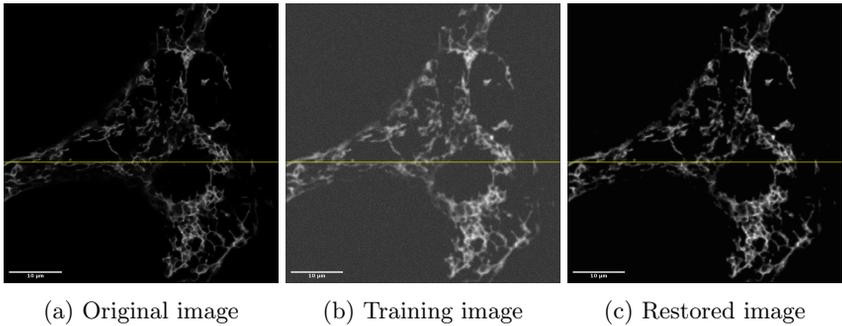
(a) Original image   (b) Training image   (c) Restored image

Fig. 2: The image sequence shows an exemplary application of the NN for mitchondria of a HeLa cell. The first image (2a) shows the original ground truth image. Digital addition of noise and blurring results in the corresponding training image (2b). Via a *Fiji* plugin, the NN can be applied and calculates the restored image (2c). The scale bar is set to 10 μm.

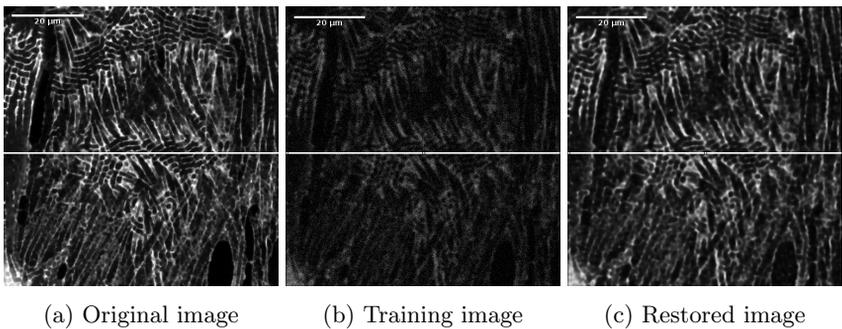(a) Original image   (b) Training image   (c) Restored image

Fig. 3: The image sequence shows training pair as well as the restored image of cardiomyocytes. A significant reduction in intensity is applied to the original image, which means that the characteristic structure of the muscle cells is poorly resolved (3b). The image (3c) is obtained by applying the trained CARE network to the training image. The scale bar is set to 20 μm.
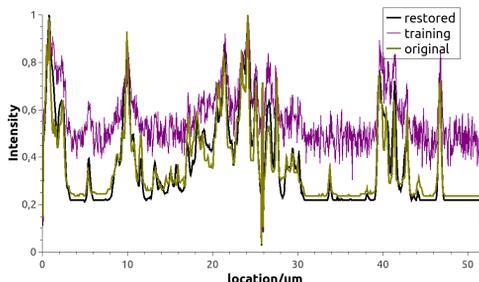
(a) Intensity profiles along lines in Fig. 2.
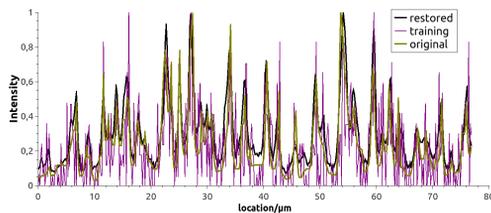
(b) Intensity profiles along lines in Fig. 3.

Fig. 4: Comparison of intensities in original, training and restored image along the respective marked line in the image. In Figure 4a, images from the mitochondria dataset are used. For Figure 4b we use restored images of cardiomyocytes in Figure 3.

as input. A second application leads to another substantial improvement. This is shown in the image sequence in Figure 5. Application of the network to an already improved image (Fig. 5b), further enhances the image contrast (Fig. 5c). We also observe an increase in structural detail in the image of cardiomyocytes, when applying the network (Fig. 6a–6b). The effect of a second application is considerably smaller (Fig. 6c) than for the image of mitochondria.

(a) Input   (b) first application   (c) second application
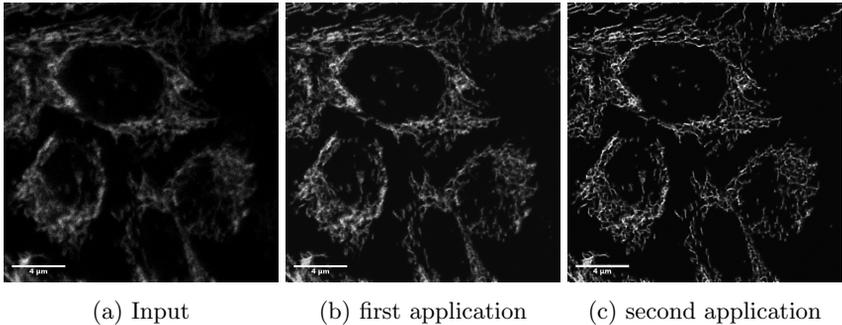
Fig. 5: The trained network is applied twice to the input image in Figure 5a. The image was not included in the training set of the network. The output shows a clear increase in image quality. The scale bar is set to 4 μm.

(a) Input   (b) first application   (c) second application
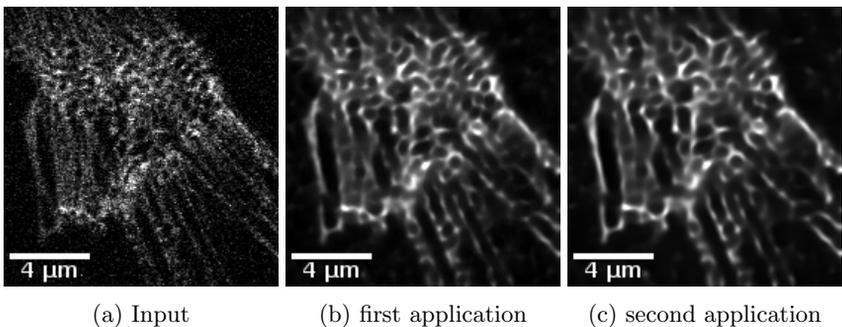
Fig. 6: The trained network is applied twice to the input image in Figure 6a. The image was not included in the training set of the network. The output shows a clear increase in image quality. Unfortunately the noisy regions are causing artifacts. This problem might be resolved by increasing the dataset size.

6   Limitations

Despite an improvement of the image quality was reached, several limitations can be identified. Especially on the smaller dataset of cardiomyocyt images, the CARE restorations show artefacts. Structures with comparable low intensity are sometimes omitted in restoration. For example, after applying the network to the cardiomyocyte image in Figure 6, the structures are now displayed continuously. However, small structures in the order of $1\mu m$ have not been restored and are difficult to detect. Furthermore, image noise (in particular in the background) is not fully removed. These problems are not that apparent on the larger mitochondria dataset. This suggest, that image restoration can be further improved by increasing the dataset or adding different distortions to the training data. This may be a topic of future work.

7   Conclusion

The application of the open source project CSBDeep on our own multi-photon and confocal microscopic recordings has led to promising results. With the provided Python package, the pre-constructed CNN architecture can be trained and tested on own pictures. The generation of training images has turned out to be the main challenge. An efficient solution is the generation of semi-synthetic training data to reach a sufficient number of training images from a small amount of original data. For this purpose, high-resolution images of a confocal microscope were digitally distorted. The approach allows the variation of intensity, noise and blur of the training images. By choosing appropriate parameters, the limitations of a fluorescence microscope can be simulated digitally.

The performance of the network is analyzed based on the training history, the intensity analysis of training, ground truth images and the restoration quality of unknown images. Despite the comparable low quantity of training images, the networks are converging in training. The application to known and unknown images results in restored images, which show an appealing increase in image quality. This is supported by the quantitative evaluation of the intensity profiles of selected images and the training graphs.

It is true that deterministic algorithms for denoising and deconvolution of images already exist. Based on our prior experimentation with such algorithms, by using the self-learning neural networks we experienced a reduction in processing time and better restoration quality for the specific structure, the network was trained on. Additionally, the network can be used to remove fixed image distortions like a pattern of repeating fibre cores in fibre-based imaging methods ,making its application far more flexible than deterministic algorithms. Therefore, CARE is a great, easy-to-use restoration tool, which can cope with small datasets and still provide reasonable improvement of image quality.

References

1. Gael Moneron and Stefan W Hell. Two-photon excitation sted microscopy. *Optics express*, 17(17):14567–14573, 2009.
2. Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, 15(12):1090–1097, 2018.
3. Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.
4. Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems*, pages 253–256. IEEE, 2010.
5. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
6. Kalina L Tosheva, Yue Yuan, Pedro Matos Pereira, Siân Culley, and Ricardo Henriques. Between life and death: strategies to reduce phototoxicity in super-resolution microscopy. *Journal of Physics D: Applied Physics*, 53(16):163001, 2020.
7. Alex Kiepas, Elena Voorand, Firas Mubaid, Peter M Siegel, and Claire M Brown. Optimizing live-cell fluorescence imaging conditions to minimize phototoxicity. *Journal of Cell Science*, 133(4), 2020.
8. Guy M Hagen, Justin Bendesky, Rosa Machado, Tram-Anh Nguyen, Tanmay Kumar, and Jonathan Ventura. Fluorescence microscopy datasets for training deep neural networks. *bioRxiv*, 2020.
9. Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void – learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.
10. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
11. Brendan P Lucey, Walter A Nelson-Rees, and Grover M Hutchins. Henrietta lacks, hela cells, and cell culture contamination. *Archives of pathology & laboratory medicine*, 133(9):1463–1467, 2009.
12. Johannes Schindelin, Curtis T Rueden, Mark C Hiner, and Kevin W Eliceiri. The imagej ecosystem: An open platform for biomedical image analysis. *Molecular reproduction and development*, 82(7-8):518–529, 2015.