First Year Report
PhD in Image Analysis

Image analysis using deep learning for optical super-resolution microscopy of living samples

Charles Nicklas Christensen (cnc39)
charles.n.chr@gmail.com

Cambridge, 2019

Laser Analytics Group
Department of Chemical Engineering and Biotechnology
University of Cambridge

West Cambridge Site
Philippa Fawcett Drive
Cambridge CB3 0AS, UK
www.ceb.cam.ac.uk

Computational Biology
Department of Computer Science and Technology
University of Cambridge

West Cambridge Site
William Gates Building
15 JJ Thomson Ave
Cambridge, CB3 0FD, UK
www.cl.cam.ac.uk

Abstract

Image restoration using deep learning for facilitating novel applications in optical microscopy is investigated. One such application is high-speed imaging of living samples with minimal photon budgets to reduce phototoxicity and preserve sample integrity. The simultaneous requirements of low-intensity excitation and high imaging speed necessitate the development of specialised denoising methods. Several different neural network architectures are implemented and test, some of which inspired by recent state-of-the-art single-image super-resolution architectures. The gain in frame rate when employing deep neural networks is estimated from a study of the noise dependency of the exposure time of experimentally acquired images. It is found that the best performing model is able to increase the frame rate by a factor of about 15 if a short exposure time is chosen. Transfer learning is attempted with a model trained on a separate diverse benchmarking dataset evaluated on a dataset that has been experimentally acquired.

Another application that is investigated is related to the endoplasmic reticulum (ER). Structural defects of the tubules in the peripheral domain of ER has been linked to neurological diseases such as Alzheimer's disease, and thus characterising the shape of ER is an important problem. A vital step the characterise the shape is to perform a image segmentation. This is more difficult when the imaging conditions are extreme such as in live-cell imaging, again due to limited photon budget or a requirement of fast acquisition rate. A model that is similar to that used for denoising is found to work well for segmenting experimentally degraded ER images even when trained on images with synthetic degradation.

Preface

This report is an outcome of the first year of my PhD project at University of Cambridge. The PhD project follows a one year Master of Research in Sensor Technologies and Applications under the CDT (Centre for Doctoral Training) for Sensor Technologies and Applications. I hold a M.Sc. in mathematical modelling and a B.Sc. in Physics and Nanotechnology from the Technical University of Denmark. Several of my elective courses have been in photonics including experimental optics and optical microscopy. Given my mixed background, I am happy to be able to work on a PhD topic that combines mathematical modelling in the form of machine learning with the physics of microscopy. The report at hand primarily reflects the former, however, as it has been the focus in the first year. I expect the remainder of the PhD project to be more balanced either by involving more direct experimental work or by exploring and taking advantage of more theoretical aspects of optical imaging.

I was accepted to present at the conference Focus on Microscopy (FOM) in London at April 17. Several of the results shown in this report stem from the preparation for this conference presentation, since it was prioritised rather highly and forced me to focus on obtaining concrete, presentable results. For that reason some of the results shown in this report are perhaps not very fundamental and may tend to be slightly example-based rather than quantitative due to the demonstrative purposes they were intended for. I believe a more statistics oriented and quantitative discussion of results would be needed to make any of the findings publishable, which is something that I hope to address going forward.

Cambridge, June 14, 2019

Charles Nicklas Christensen (cnc39)
charles.n.chr@gmail.com

Acknowledgements

The main host group is the Laser Analytics Group (LAG) lead by Professor Clemens Kaminski, whom is also my first supervisor in the project. I have appreciated his guidance and enthusiasm regarding possibilities of the research. My second supervisor is Professor Pietro Lió of the Computational Biology Group. Pietro always has lots of ideas for new directions and I find his input very useful.

I would like to thank Dr Katharina Scherer, whom is a PostDoc in LAG that has helped me with imaging and sample preparation, which has led to a supervised dataset of actin using wide-field fluorescent microscopy described further in this report.

Dr Edward Ward, also a PostDoc in LAG, deserves my gratitude for several discussions and helping out with imaging too. Some of these discussions with Edward have given ideas for future work – in particular regarding a machine learning-based reconstruction method for Structured Illumination Microscopy (SIM).

Finally, I am thankful to Dr Meng Lu whom is a PostDoc in the Molecular Neuroscience Group, a group that works very closely with LAG, for providing data for the segmentation problem described in this report and showing great interest in the prospects of this application.

Contents

CHAPTER 1

Introduction

1.1 Image analysis with deep learning

Deep learning methods have become state-of-the-art in virtually all low-level computer vision tasks. Within the field of microscopy the range of applications include:

- **Restoration and enhancement.** – improving degraded images by e.g. denoising (removing noise), super-resolution (increasing resolution), inpainting (filling out blanks), artefact removal and deblurring (making images less blurry) [1, 2, 3, 4, 5, 6].
- **Segmentation** – partitioning an image into respective parts each corresponding to a different type of object [7, 8, 9].
- **Reconstruction** – taking raw microscopy data and combining it into meaningful or super-resolution images. In the literature there are several cases of deep learning-based reconstruction methods for stochastic optical reconstruction microscopy (STORM) [10, 11, 12] as well as Fourier ptychography [13, 14]
- **Sample classification** – determining what is seen in the image, i.e. which classes from a set of possibilities are present in the image [15, 16].

The primary focus in this PhD project so far has been the restoration, specifically super-resolution and denoising. However, towards the end of this report, other deep learning-based applications are presented.

To approach the two separate problems of super-resolving images and denoising them in a more fundamental way, the problem of performing the restoration when receiving only a single image as input is considered. The multi-image problem is also relevant for many applications, and while it may not be an easy problem in general, it is easier to come up with good solutions by leveraging the redundancy of data. For denoising, the single-image premise necessitates that a representation of the noise source is implicitly learned to then account for the noise that appears in one individual image, whereas in a multi-image scenario the image variations of the same object could simply be averaged to cancel out the noise (assuming zero-mean noise). Likewise for super-resolution tasks, a deep representation of the typical shapes of naturally occurring objects must be learned to accurately enlarge features, whereas it might be much easier to extrapolate from trends observed in the image variations of the same object.

In this report, the topic of single-image super-resolution will first be discussed and some results are shown, which feed into the more thorough treatment of denoising by

drawing some parallels. Rather than giving an overall literature review of the entire field of restoration, the literature is discussed for each of those topics in their respective parts.

1.2 Aims and objectives

The goal of the PhD project is to find novel and robust ways to perform image analysis of microscopy images using machine learning. A concrete objective has been to design, implement and train models to handle very low signal-to-noise ratios that can facilitate low-light imaging, which is relevant for several areas of live-cell imaging discussed later. As a follow-up goal of this, one such well-performing model would be used in conjunction with a high-speed optical microscope in the host group to study a biological problem that has eluded other imaging techniques.

Since many of the models are very broadly applicable, the sole focus of denoising for low-light imaging would be an unnecessary restriction, and therefore other applications such as super-resolution, segmentation and reconstruction with similar kinds of neural network architectures will also be investigated. At the moment of writing, another concrete aim is to build a versatile segmentation model to use for analysing the endoplasmic reticulum imaged under various imaging conditions. Finally, reconstruction of SIM images using a neural network is something that will be aimed for, but given the premature research into this so far, it is uncertain whether this is realistically achievable. A more elaborate account of concrete goals within the second year of the PhD is given in chapter 6.

CHAPTER 2

Super-resolution

A lot of work has been done in the field of single-image super-resolution. Traditional computational approaches are primarily based on interpolation schemes such as (in order of increasing performance) bilinear, bicubic or Lanczos [17]. Bicubic interpolation is very often used as an efficient approximation of Lanczos resampling, for instance when resizing in Microsoft Word or Adobe Photoshop, and bicubic upscaling can thus arguably be considered the de facto standard of image rescaling and it shall be used as a baseline in the following.

In recent years the literature of SR has split into two directions: one dealing with achieving the best possible reconstruction errors and another focused on producing the most perceptually pleasing and convincing (referred to as having low perceptual loss) output to a human observer. The reason those two directions are not reconcilable is that the reconstruction errors typical defined by the mean squared error tend to give optimal solutions that do not contain high frequency content but rather appear somewhat blurred or washed out when compared to an original. For the recovered image to contain high frequency features it is necessary to artificially generate features that are not at all in the low-resolution input. This is in general achieved using Generative Adversarial Networks (GANs) [18] with an approach pioneered in [19]. The state-of-the-art methods include SRGAN [19], SRFeat [20], ESRGAN [21] and EnhanceNet [22] – all of which are based on

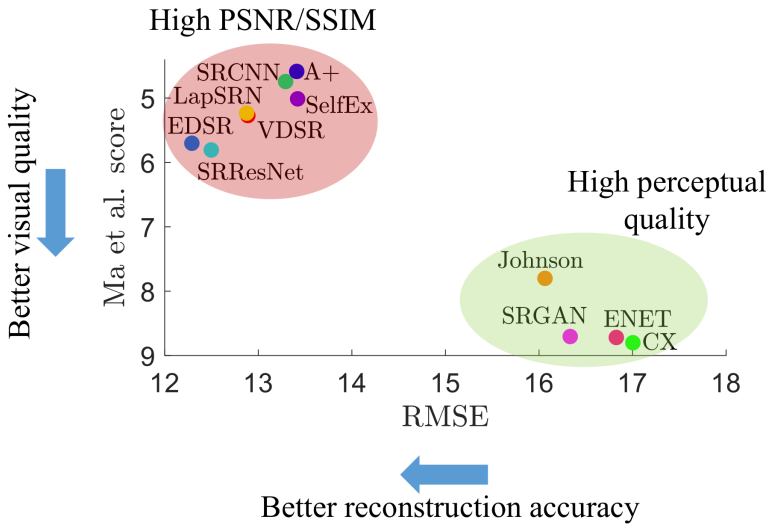

Figure 2.1: Trade-off between reconstruction error and perceptual loss for state-of-the-art methods [26]. Note that models from publications in 2018 are not included.

a GAN. The state-of-the-art methods for achieving the best reconstruction error include SRResNet [19], EDSR [23] and the recent EPSR [24] and RCAN [25].

The trade-off between reconstruction error and perceptual loss of various state-of-the-art methods is summarised in figure 2.1. A comparison of the two currently best performing models in each camp, RCAN and ESRGAN, are shown on figure 2.2. The texture clearly looks more realistic and of higher fidelity in the output of the ESRGAN, while the performance score – peak signal-to-noise ratio (PSNR) in units of dB (introduced formally in chapter 3) – is higher for RCAN. This is because several of those strands of hair in the ESRGAN are simply made up, which should be evident when comparing closely to the high-resolution (HR) ground truth image.

If the purpose of a super-resolved image is to be used for analysis in quantitative research, then it does not seem appropriate to distort the image data, i.e. generate high frequency features, to make it look more realistic, because in the end the user likely prefers to be confident about what the image shows rather than having an artificially realistic image. Therefore this PhD project has so far focused on methods that obtain minimal reconstruction errors; the methods in the red ellipsis of figure 2.1.

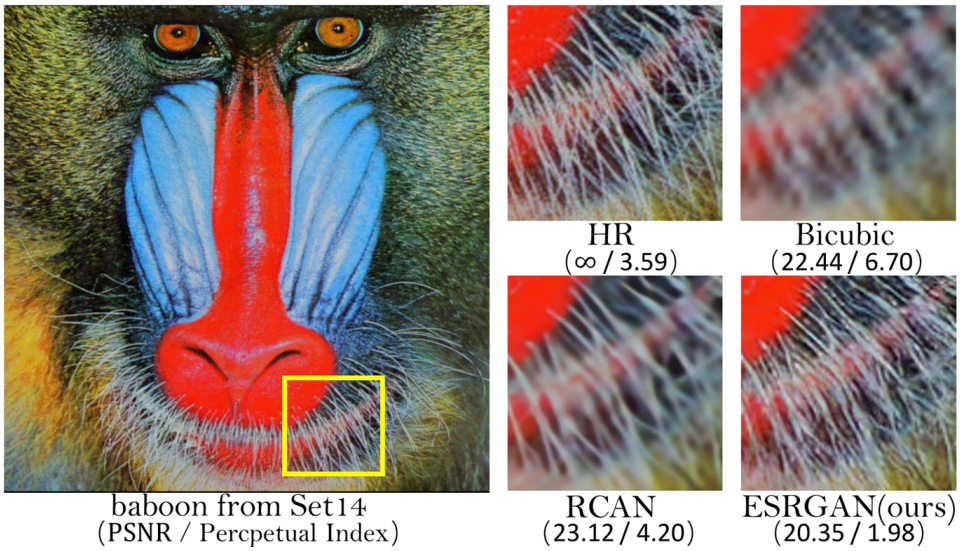

Figure 2.2: Example output of ESRGAN, which uses a Generative Adversarial Network (GAN) architecture to distort the input image to approximate the high-frequency textures. Image credit [21].

2.1 Datasets

Models have been tested with different popular benchmarking datasets such as ImageNet [27], DIV2K (DIVerse 2K resolution high quality images) [28] and BSD (Berkeley Segmentation Dataset) [29]. But in the interest of training and evaluating models on relevant data, microscopy image data acquired from members of the host group (using structured

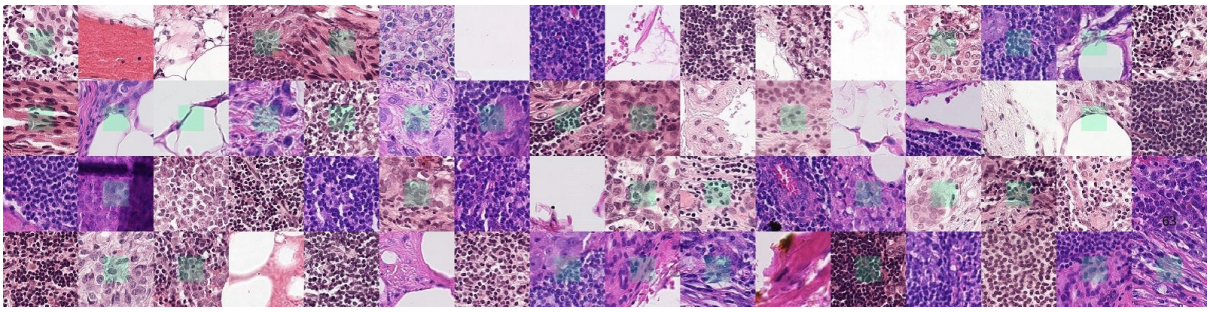

Figure 2.3: The PatchCamelyon (PCam) benchmarking dataset available on GitHub [30].

illumination microscopy and light sheet fluorescence microscopy) has also been considered and will be discussed in chapter 3. However, due to the necessity of a very large quantity of diverse training samples when training deep models, the data from the host group is not currently enough. To ensure that trained models generalise better, a large bioimage dataset called PatchCamelyon (PCam) [30] is used for the moment. PCam consists of brightfield microscopy images of lymph nodes from histology. A random sample of images from the dataset can be seen on figure 2.3.

2.2 Single-image super-resolution

Four different models of the ones previously mentioned have mainly been tested: SRResNet, SRGAN, EDSR and RCAN. The method RCAN is a recently proposed model that is found to have very good reconstruction performance but also is quite computationally demanding. Each of the four models are trained on 32000 images from the PCam dataset for at least 30 epochs (training iterations of the entire dataset) using ADAM [31] as an optimiser with a learning rate of 1e-5. The trained models are then evaluated on 50 independent test images also from PCam. The average performance scores measured by the two metrics PSNR and structural similarity index (SSIM) [32], both formally introduced in chapter 3, can be seen in table 2.1.

PCam	bicubic	SRResNet	SRGAN	EDSR	RCAN	HR
PSNR	17.38	18.59	18.60	18.76	**19.43**	∞
SSIM	0.433	0.616	0.614	0.610	**0.657**	1

Table 2.1: Comparison of different methods for 16x image upscaling and the original high-resolution, HR, on the benchmark dataset PCam. The methods are bicubic interpolation, SRResNet [19], SRGAN [19], EDSR [23] and RCAN [25]. RCAN is observed to have the best performance on the test set in terms of the metrics PSNR [dB] and SSIM.

Two examples of an input image recovered by bicubic upscaling and compared to a prediction from the trained RCAN model can be seen on figure 2.4.

A direct comparison of an individual output of the four tested models is shown on figure 2.5. The SRGAN is found to perform very similarly to SRResNet, which is because

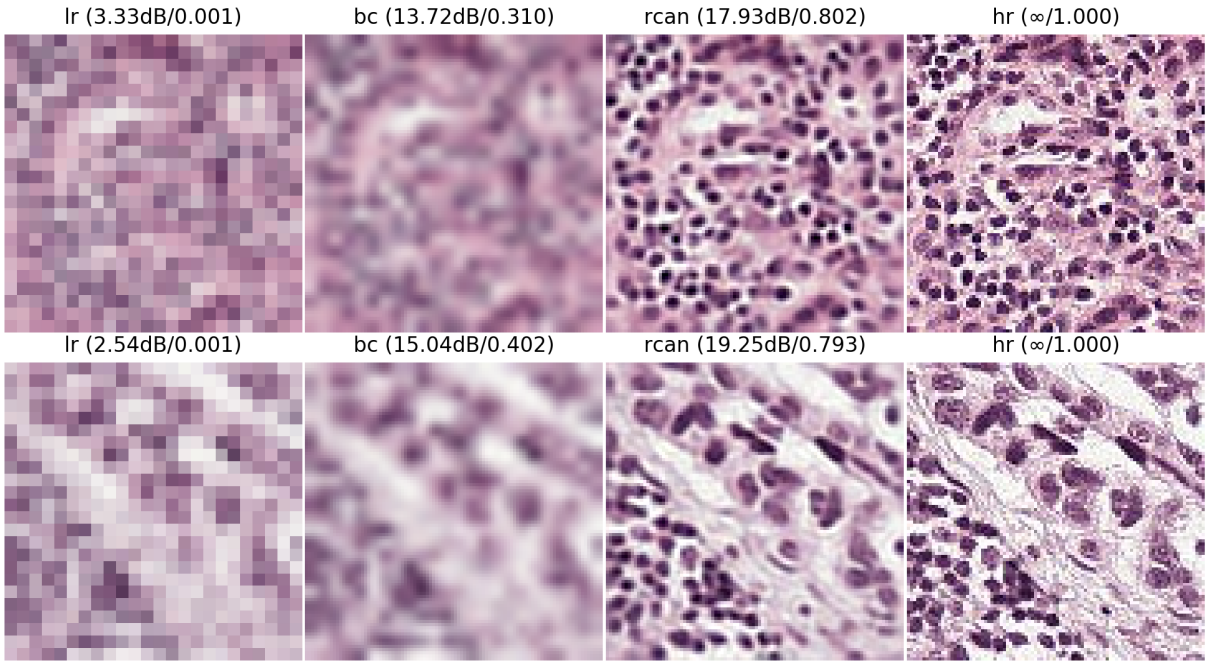

Figure 2.4: Two different test images that are both 4x super-resolved (in each dimension, so 16x in terms of pixel count) from a 24x24 pixel input image to a 96x96 output. From left to right: input low-resolution image (shown here upscaled with simple repetition), image upscaled by bicubic interpolation, the model prediction and the unseen high-resolution image.

they are based on the same model, but have different loss functions. The very similar performance indicates that the loss function of SRGAN has not been configured properly in the test to allow enough distortion for the GAN to really shine. The EDSR model is an improvement of SRResNet and as expected we do see a somewhat better performance in table 2.1. The RCAN model performs significantly better than the other methods, which is impressive given that it was the model trained for the least number of epochs, namely 30 vs 100 for the others (a limit of 10 hours of computation on the ComputerLab GPU cluster was set, and the training of RCAN did not advance further in that time window).

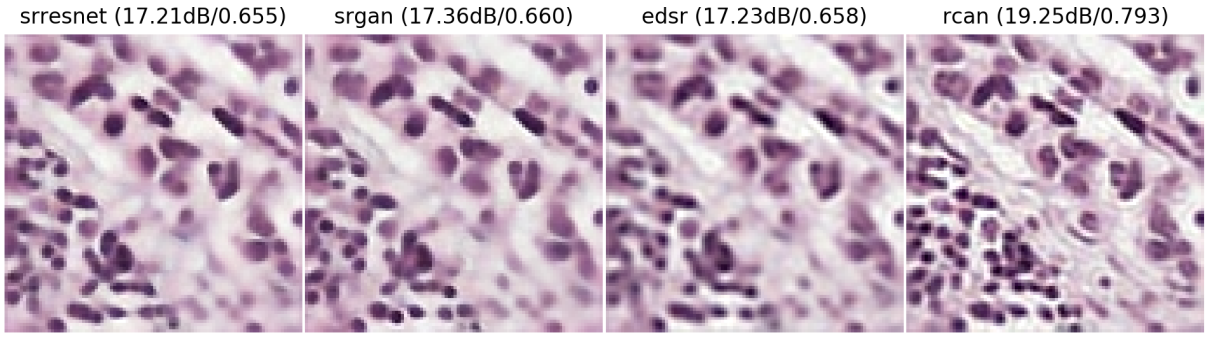

Figure 2.5: Comparison of predictions from state-of-the-art learning-based SR methods for 4x upscaling in each dimension. For the original high-resolution image see figure 2.4.

CHAPTER 3

Denoising

This chapter reports on work done on denoising of microscopy images in the setting of low-light imaging. As motivation for doing low-light imaging, it is worth to consider some existing use cases that operate with a very limited photon budget. Firstly, there is imaging of highly dynamical systems. To resolve objects that are moving fast without producing motion blur, a fast acquisition rate is required, which implies that a short exposure time must be used. One such example is imaging of the endoplasmic reticulum to investigate peristaltic flow of luminal proteins, which has been achieved via structured illumination microscopy (SIM) at 40 frames per second (FPS) [33]. Another example is the monitoring of a beating heart in zebrafish, a process that requires an even higher frame rate of 100 FPS to be temporally well-resolved, which was achieved with light sheet flurescent microscopy (LSFM) [34].

Another category of imaging with a limited photon budget is long-term imaging, where illumination must be kept minimal to reduce photodamage. Long-term imaging is important for studying developmental biology. Since LSFM is a gentle imaging method, it has proved particularly suitable for long-term imaging. The technique has been used to investigate organ morphogenesis in drosophilia over a period of 20 hours [35].

Finally, it may also be that a three-dimensional dataset of a living sample is desired. Even if the sample system may not be very dynamic, volumetric imaging may well require a fast acquisition rate, since several planes have to captured to construct a volume of the sample (a z-stack). Fast volumetric imaging has for instance enabled a detailed study of mitosis, the separation of chromosomes into two new nuclei, by resolving the complex three-dimensional structure of two chromosomes as they split at 1 volume per second with each volume consisting of 200 planes [36].

All of these current use cases can be realised with LSFM. The question this section attempts to bring some light on is whether the technology can be pushed further? If one desired hundreds of frames per second of a dynamic system, or imaging of a developing organism for days instead of hours or many volumes per second of moving three-dimensional structures, the only remedy would be to have higher frame rate and less photo-toxicity. This in turn requires shorter exposure times and less excitation power. Thus, a trade off with quality is necessary, which means one must accept large amounts of noise. Analysis of the extent to which deep learning for image restoration can improve on these limitations is presented.

3.1 Related work

The literature on image denoising has traditionally been based around local averaging approaches, such as the application of a Gaussian smoothing filter [37, 38]. Other local filter methods include least mean squares filter [39], anisotropic filters [40] and in the frequency domain; Wiener filters [41] and wavelet thresholding methods [42].

Local methods are computationally light, but have obvious limitations. First of all, the averaging often involved in local methods introduces blur, which is a degradation by itself, rendering features to be less defined. Secondly they do not perform well for high noise levels, since the correlations between neighbouring pixels deteriorate [43].

Non-local filters solve some of these problems by using self-similarity of natural images beyond neighbouring pixels [43]. The first method to propose this is the non-local means method [37], in which patches are restored by weighted averaging of all other patches in an image. Since then a number of improvements have been proposed such as invariance to patches that are rotated or mirrored with respect to each other [44], and improved computational efficiency, automated parameter tuning and extension to 3D image stacks [45]. Although the non-local filters are better at high noise levels, they will typically lead to artefacts like over-smoothing [43].

Another category of denoising methods that are distinct to the ones previously mentioned is learning-based methods. The first learning-based methods to become a trend in denoising were sparse dictionary learning methods that attempt to find sparse representations of the input data in the form of linear combinations. The methods perform denoising by expressing an image patch in the denoised image as a linear combination of other patches in a trained redundant dictionary consisting of a large number of patches obtained from an image dataset [43]. An example of this type of method is the K-SVD method that uses K-clustering with singular value decomposition [43, 46].

More recently with the emergence of deep learning supervised learning has taken over, and several end-to-end convolutional neural networks (CNNs) have been proposed for denoising. These will be discussed briefly in the following section.

3.1.1 Denoising using deep learning

A pioneering deep CNN for image analysis is the U-Net [7]. The model was originally intended for segmentation, but it has seen use for restoration tasks such as inpainting [5] and in particular denoising [3, 6].

The central idea of U-Net is that an input image is taken through convolution layers at different resolutions, see figure 3.1, while employing so-called skip connections at every resolution. After the image is passed through three convolution layers, a pooling operation is used to lower the resolution and this sequence of convolution layers and pooling repeats until a certain number of levels is reached. The pooling operation can be defined in different ways, but typically max-pooling is used with a window width of 2, meaning that each 2x2 pixel segment is reduced to the maximum value of the pixels in the patch. The lower resolution at deeper levels greatly saves computational load, and consequently

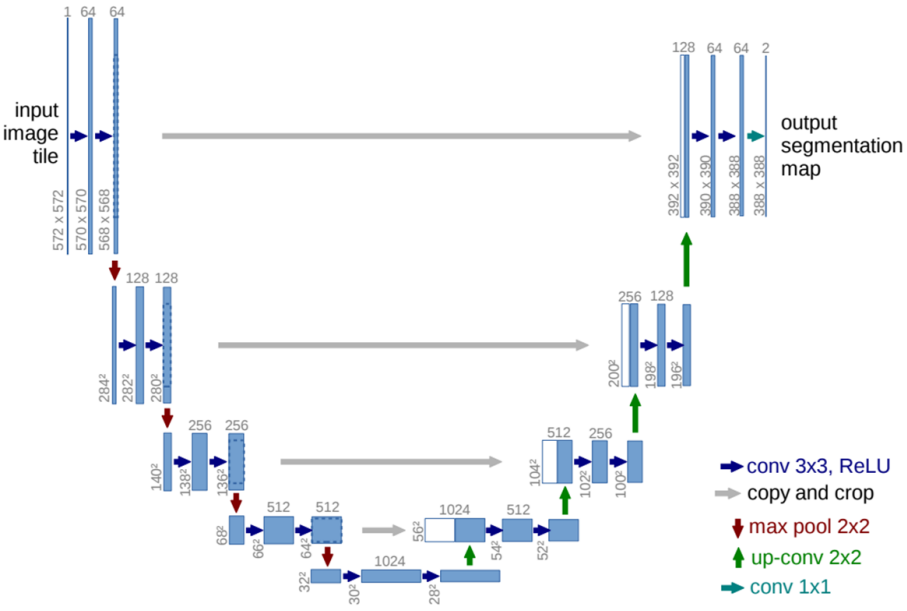

Figure 3.1: Convolutional neural network based on the U-Net architecture [7].

the number of filters in convolution layers can be increased without causing the training time to increase significantly.

The skip connections in U-Net that are seen as the horizontal lines in the diagram of figure 3.1 pass intermediate results to subsequent layers. These connections prove to be crucial for a robust convergence during training by avoiding the vanishing gradient problem. This has been further investigated in [47], in which another deep CNN was proposed that has also been used for denoising [3], and it was shown that the skip connections lead to restoration performance gains.

The original U-Net architecture has five levels in this way, but it can be customised for better or worse, such as the more light architecture in [3] with only four levels and fewer convolution filters in the convolution layers. This lighter architecture is referred to as UNet-N2N (N2N meaning noise2noise after [3]) for the remainder of the chapter. For the numerical experiments presented later, one of the models will be this one. Another variant that will be considered is a customised, heavier version of U-Net with six levels of resolution and more than double the number of filters at the lowest level when compared to the original architecture. This model will be referred to as UNet-60M, since it has about 60 million trainable parameters, whereas UNet-N2N and the original architecture has about 1 million and 13 million parameters, respectively.

3.2 Leveraging super-resolution architectures

While the U-Net architecture provides efficiency and robustness, one might wonder how much the restored output suffers by having the majority of computations done on down-sampled versions of the input image. The architectures employed in super-resolution

neural networks tend to be different. As argued in section 2.2 the field of single-image super-resolution (SR) research has seen more activity than that of denoising. The state-of-the-art methods perform very well when trained on microscopy images as indicated by figure 2.4 and 2.5.

These results have motivated an experimentation in this PhD project of customising the super-resolution models to perform denoising rather than upsampling.

The state-of-the-art SR architectures generally do not have downsampling between layers [4, 19, 23, 25], however they alleviate training by following the structure of residual networks as first introduced with ResNets [48] with image classification in mind and later repurposed for restoration with SRResNet [19]. Residual networks use the previously mentioned skip connections but more rigorously by having the shortcuts after every few stacked convolution layers, which then constitutes the residual building block that can be repeated many times. The residual networks allow for training of very deep networks, which was demonstrated in [48] with an appropriately termed "aggressively deep model" consisting of 1202 layers that was trained with no optimisation difficulty, although such networks have a large risk of suffering for overfitting thus needing careful regularisation.

The design idea of residual networks was taken one step further in Enhanced Deep Residual Networks (EDSR) [23] by proposing a modified residual building block called ResBlock, which was found to be superior to the previously proposed and more directly adapted ResNet model called SRResNet [19]. A diagram showing the EDSR network can be seen on figure 3.2, which includes a block that has simply been coined "Flexible" since it varies with the different purposes it has succesfully been customised for during the research of this PhD project.

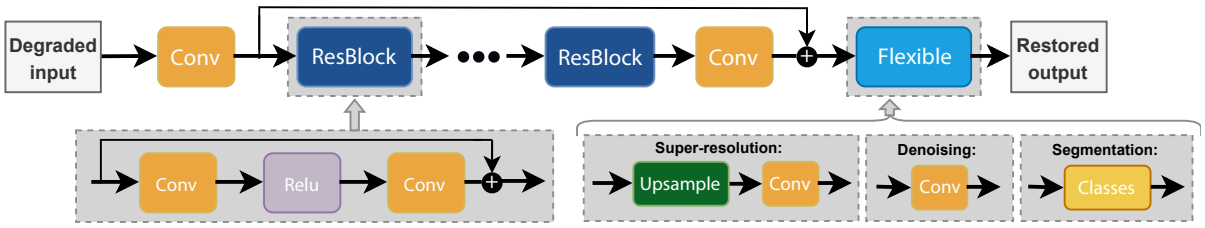

Figure 3.2: EDSR model

Yet another improvement in this class of network architectures was made with Residual Channel Attention Network (RCAN) [25], which augments the ResBlock with two more convolution layers and a global pooling operation. These extra layers are combined with the layers in the standard ResBlock by a skip connection that does not combine old and new results by addition but by multiplication, which the author argues allows the model to learn to adaptively rescale channel-wise features by considering the interdependencies among channels. RCAN also adds more regular skip connections in a design that is coined Residual in Residual (RIR), where a preset number of ResBlocks are considered a group, and a long skip connection is made from the beginning of the group to the end, whereas the skip connection of each ResBlock only passes by a few convolution layers. The author believes this allows abundant low-frequency information to be easily passed

on, thus enabling the main network to focus on learning to restore high-frequency information. The benefits of not only having shoer skip connections but multiple long skip connections is supported by an earlier study [49].

The *Flexible* block of figure 3.2 is set to be a single convolution layer for denoising tasks, since the image resolution of input and output is equal.Thus the only thing required by this final convolution layer to produce the model output is to take the numerous channels of the ResBlocks feature maps and reduce them to the desired number of output colour channels, i.e. 1 for grayscale or 3 for RGB, which is achieved by configuring the convolution layer to have the same number of input channels as the feature maps while only applying 1 or 3 trainable filters. For super-resolution the upsampling is ideally done with a fractional convolution layer (also known as a transposed convolution layer) that has a stride, e.g. a stride of 2 in a fractional convolution layer will give a double resolution. Alternatively a so-called pixel shuffle operation can be used, which is another way to perform sub-pixel convolution with fractional strides. The cheaper way to do upsampling without any extra trainable parameters is to interpolate image, such as bicubic interpolation, before feeding the result to the final convolution layer.

3.3 Supervised training dataset via variable exposure time

To assess how much these methods potentially can improve low-light imaging operation, a simple investigation into how noise depends on exposure time was first conducted. As a metric to quantify degradation the structural similarity index (SSIM) can be used, which takes on values from 0 to 1 and is defined as

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{3.1}$$

where x and y are two images of equal size, parameters $\mu_{x|y}$ are the means of the images, parameters $\sigma_{x|y}$ are the variances of the images and σ_{xy} is the covariance of the two images. The constants c_1 and c_2 are included to stabilise the division with weak denominator – by default they are set to small values.

A wide-field fluorescent microscope was used to acquire several hundreds of images of different parts of a fixed sample of actin with both short and long exposure times, from 5 ms to 200 ms.

This dataset set consists of pairwise low and high quality images that could be used as noisy inputs and clean targets to train a model in a supervised manner. However, the dataset is not large nor diverse enough for training a deep model such that it becomes generalised – overfitting is very likely to happen unless thousands of diverse training samples are available.

For this reason the PCam dataset as introduced in section 2.1 is used for training. The aim is to use the original dataset to obtain statistics for the noise that can be used to degrade the PCam dataset to estimate improvements in acquisition rate and to obtain a generalised model that can even denoise the original dataset.

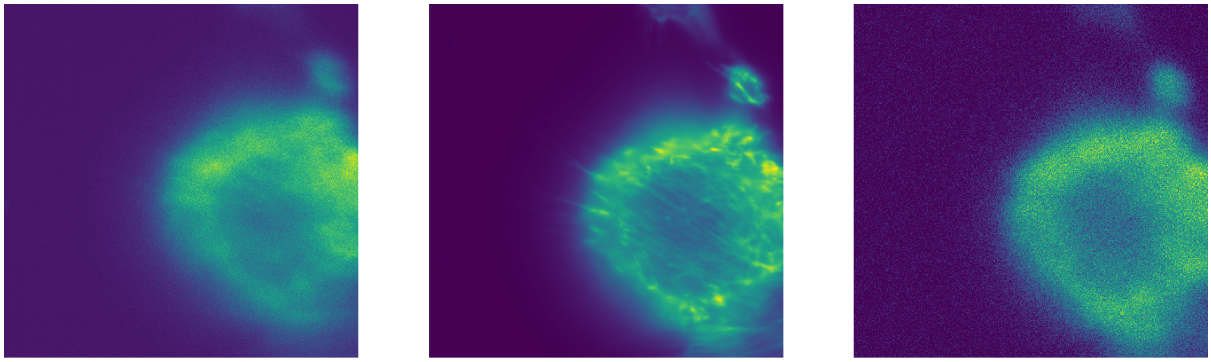

Figure 3.3: Approximating degradation with synthetic noise model. Experimentally degraded 5 ms exposure time image (left), high quality 200 ms exposure time image (centre) and synthetically degraded 5 ms exposure time image with matching SSIM (right).

In order to use the PCam dataset for training, it is desirable to have a good approximation of the noise sources governing the original data. A noise model that includes Gaussian and Poisson noise, which is introduced more formally in section 4.1, is tuned to match the experimentally acquired data as closely as possible. On figure 3.3 an example image in the acquired dataset in its low and high quality versions are shown with an approximately matching synthetically noised version.

3.4 Quantifying potential gains in acquisition speed

An example of a section of the sample in the acquired dataset with different exposure times are shown in figure 3.4. Using the image with exposure time of 200 ms as a reference, it is seen as expected that the SSIM approaches zero as the exposure time becomes shorter.

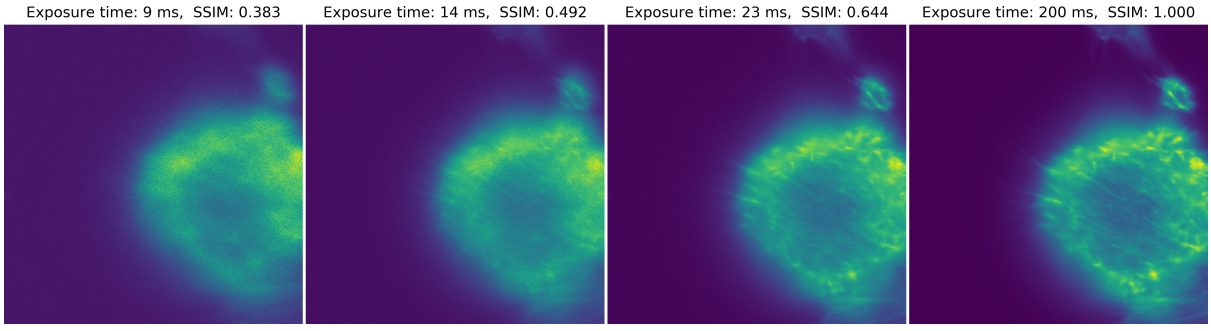

Figure 3.4: Image quality as a function of exposure time.

The tendency of the SSIM as a function of the exposure time when averaged over all images in the original dataset can be seen to the left in figure 3.5. A fit to a two-term exponential function $\mathrm{SSIM}_{\mathrm{fit}}(\tau) = a \cdot \exp(b\tau) + c \cdot \exp(d\tau)$, where τ is exposure time,

enables the relationship to be described analytically. The fit is also plotted on the figure and is seen to closely follow the trend.

When degrading the high quality images with the noise model described above with different levels of noise, a similar tendency appears as can be seen on the right of figure 3.5. An analytical description using a fit of a two-term exponential function is obtained again, i.e. $\mathrm{SSIM}_{\mathrm{fit}}(\eta) = a \cdot \exp(b\eta) + c \cdot \exp(d\eta)$. The two functions for SSIM can now be set equal to correlate exposure time and noise level, i.e. $\mathrm{SSIM}_{\mathrm{fit}}(\tau) = \mathrm{SSIM}_{\mathrm{fit}}(\eta) \Rightarrow \tau(\eta)$, such that equivalent exposure times for a given noise level can be estimated.

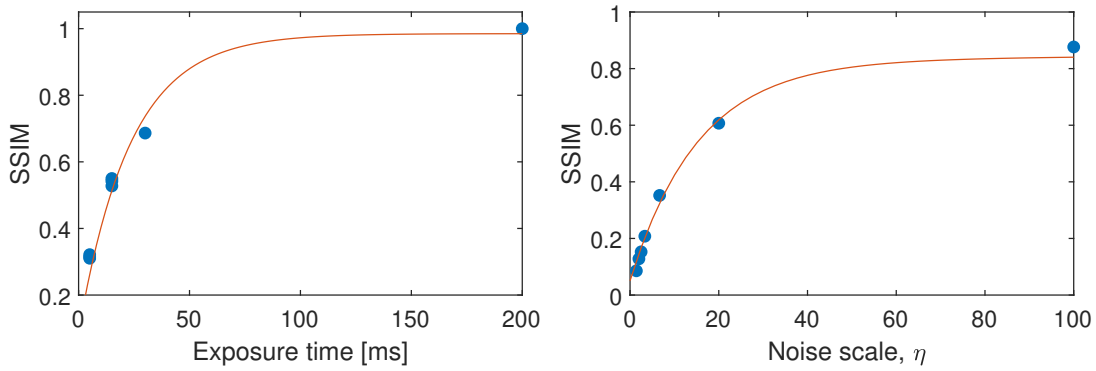

Figure 3.5: Correlating exposure time and noise level via structural similarity index. A two-term exponential function is fitted to each series to provide an analytical description of the tendency.

3.5 Implementation, performance and results

Being able to approximate the effect of noise under low-light conditions, the PCam dataset is used for training with synthetically degraded inputs corresponding to 5 ms exposure. The models implemented are the modified super-resolution models EDSR and RCAN as well as three variants of U-Net: the original architecture, a lighter architecture referred to as UNet-N2N and a heavier customised architecture referred to UNet-60M.

The number of trainable parameters and the memory usage of each model is depicted on figure 3.6. The SR models have a relatively low number of parameters of about 1 million, whereas the original and heavy U-Net models have many times more. This is clearly reflected by the memory usage for storing the parameters. However the memory usage during a feeding an input forward through the network, as well as backpropagating the resulting error, is actually lower for the U-Net models, because most computations for these models are done on downsampled versions of the input.

Training is done on 30000 training samples from the PCam dataset using the 5 ms synthetically noised inputs. The learning rate is initially set to 10^{-4}, and is halved for every 5 epochs. After every epoch the peak signal-to-noise ratio (PSNR) is evaluated on

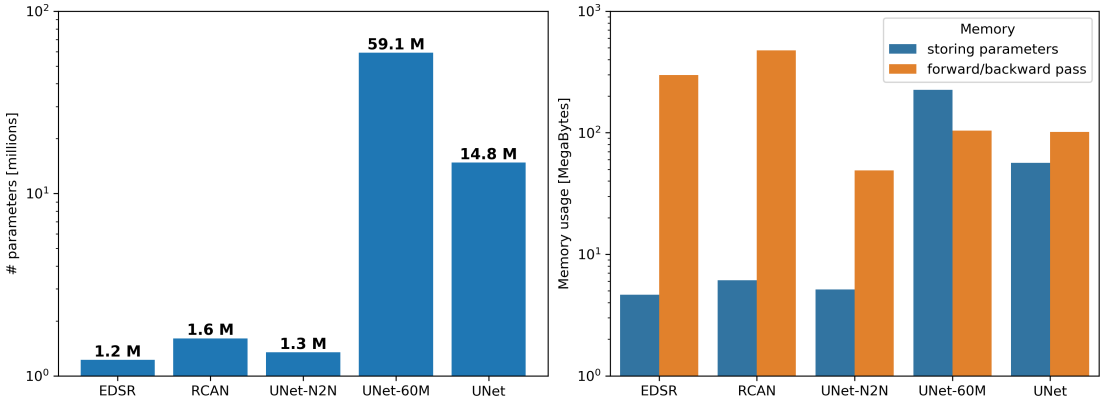

Figure 3.6: Number of trainable parameters in models (left) and their respective memory consumption (right).

a separate test set of 100 images. The PSNR is defined as

$$\mathrm{PSNR}(x, y) = 10 \log_{10} \left(\frac{1}{\mathrm{MSE}} \right), \quad \mathrm{MSE} = \frac{1}{N} \sum_{i=0}^{N-1} [x(i) - y(i)], \tag{3.2}$$

for two images x and y each consisting of N pixels that are assumed to be represented by floating point numbers ranging from 0 to 1. The convergence of the PSNR during 30 epochs of training can be seen on figure 3.7. Both SR models are seen to perform well. EDSR tends to match the more parameter-heavy U-Net models. The much heavier U-Net model does not add much in performance. Although it is more computationally expensive, RCAN performs significantly better at around 0.5 dB higher PSNR compared to the U-Net models.

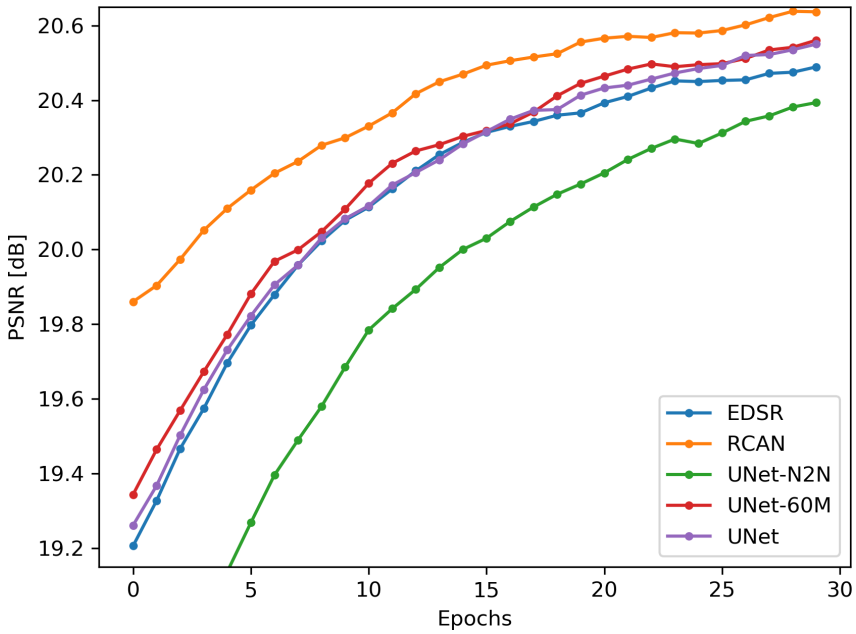

Figure 3.7: Performance of models on test set during training.

In terms of training time it is clear that the U-Net models save time by exploiting downsampling. Training for 30 epochs on a GPU cluster the U-Net models takes 1000, 1500 and 2400 seconds for UNet-N2N, UNet and UNet-60M, respectively. On the other hand EDSR and RCAN take 4000 and 7000 seconds, respectively. RCAN thus takes the most time, but it also performs best while having a relatively low number of parameters.

An example of a restoration output comparing RCAN and U-Net are shown on figure 3.8 with a smoothed version as reference. Both model outputs are clearly great improvements from the input resembling the target, the unseen ground truth, closely. As expected from the test results during training, the RCAN model performs better – again with a PSNR at about 0.5 dB higher than for U-Net. Looking more closely at the two model outputs reveals significant differences in the details of certain features – note green disks in figure. Some features are not recovered at all by the U-Net model, while others appear more washed out.

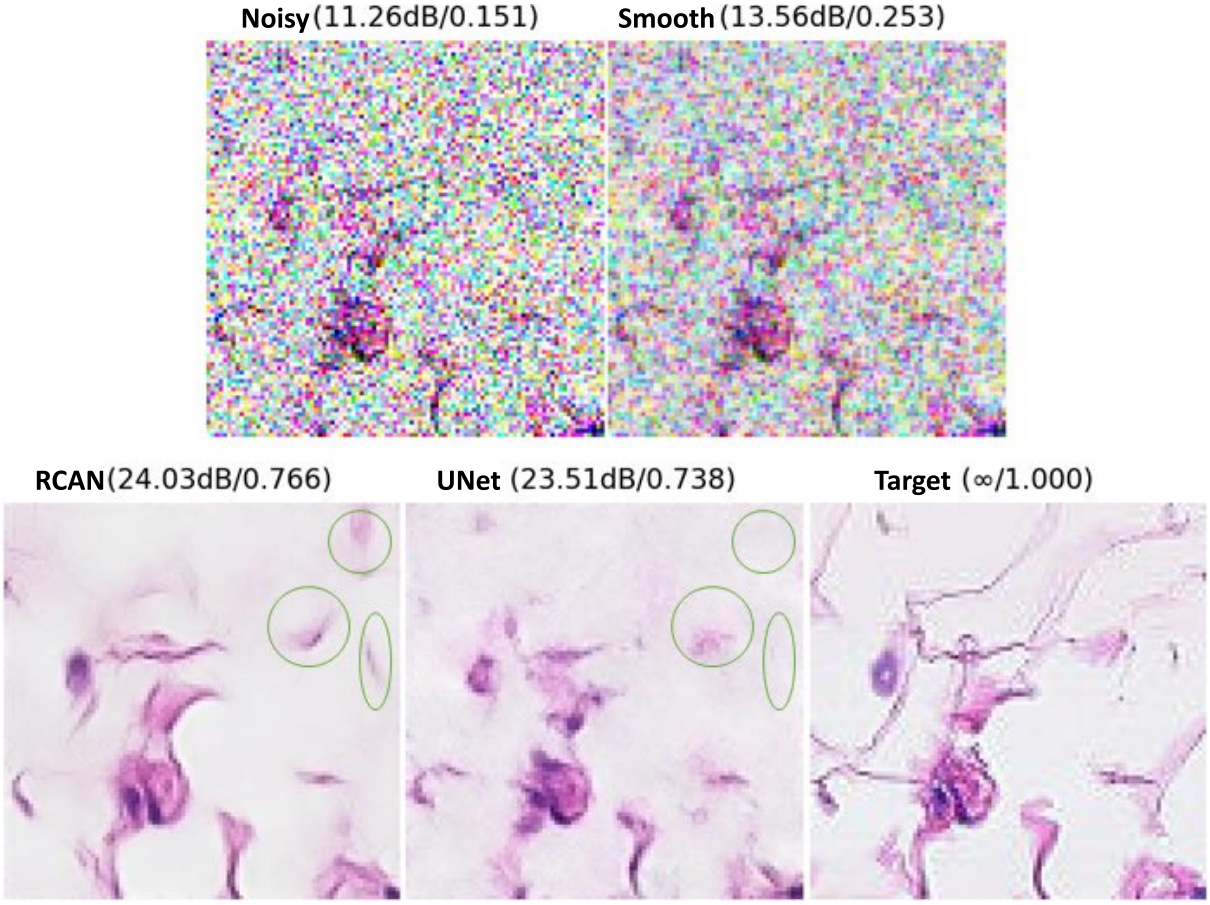

Figure 3.8: Example outputs from models. The output of RCAN has a better PSNR score, and considering the features within the green disks, it is evident that the U-Net model has not managed to resolve the same details as the RCAN model.

Based on this example, the image quality of the input image can be converted into its estimated equivalent exposure time of 2 ms, whereas the restored output from the RCAN model corresponds to 33 ms exposure time. This equals a 15 times higher frame

rate if one was aiming for the quality at ≈30 ms exposure time but operating at 2 ms exposure time. A similar figure of 15 times improvement is obtained when evaluating on an ensemble.

3.6 Usefulness for quantitative analysis

Going beyond looking at the quality scores another means of validation could be to consider the benefits for the purpose of performing quantitative analysis. A simple task to quantify an image of a biological sample could be to count the occurrence of some objects in a frame. The PCam dataset consists of a large part of images of cell nuclei. Nuclei are easy to count with a blob detection algorithm. A standard difference of Gaussians blob detection algorithm was used to count nuclei at different exposure times.

For a random sample at 200 ms, 178 nuclei was detected in this way. As shown on figure 3.9 the number of detected nuclei goes down as the exposure time becomes shorter, since the degradation due to noise starts to corrupt the shape of nuclei. For this random sample, the count is 127 nuclei at 7 ms exposure time, meaning that almost 30 % of the nuclei now fail to be detected. When evaluated on an ensemble it is found that this misclassification rate is about 21 % for the given noise level.

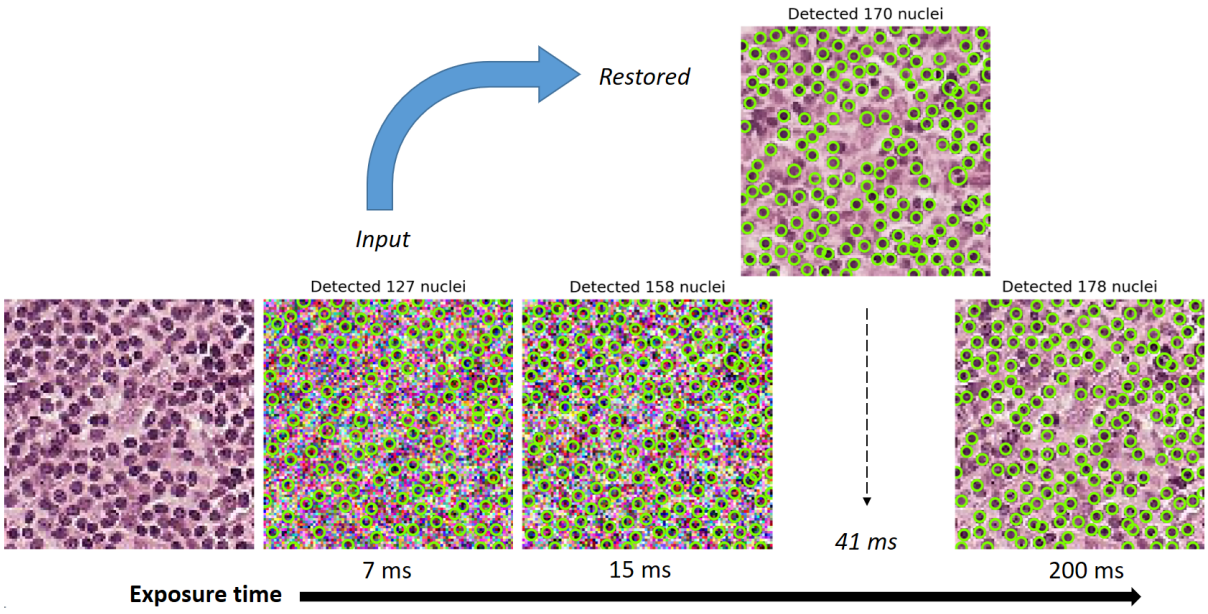

Figure 3.9

However the restored image of this example yields a count of 170, meaning that the misclassification is just 4 %. The estimated equivalent exposure time of the restored image is 41 ms up from 7 ms. Hence, the improvement of the misclassification rate is found to be as significant as that of the frame rate, indicating that the restoration not only improves quality scores but also has a more tangible practical benefit.

3.7 Transfer learning

The results reported so far in this section indicate promising improvements for low-light imaging when using deep neural networks. However, it is clear that a good training dataset is needed to achieve this level of performance, which begs the question how applicable the methods are in general. More specifically, if a model trained on a benchmarking dataset, might it still perform well on completely distinct datasets?

This is in general referred to as transfer learning, and it is something that might be very valuable if a training dataset is hard to acquire. Using images from the original dataset described previously, one can try to apply the trained model on them. An example can be seen on figure 3.10, where the output still seems quite well restored when compared to the high quality, 200 ms exposure time version.

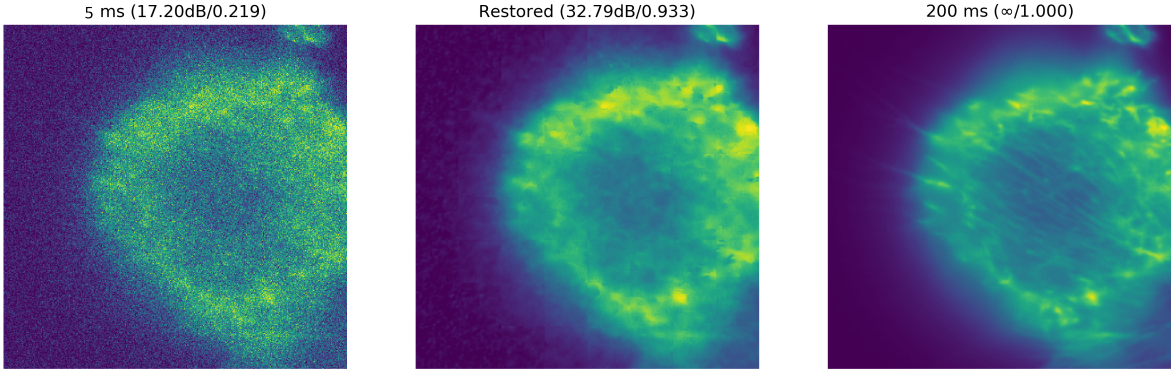

Figure 3.10: Attempt at transfer learning by applying the trained model to the original image dataset. The training dataset consisted of entirely different sample types acquired with another microscopy technique.

The restored image quality has an equivalent exposure time of 65 ms, while the input was 5 ms, almost achieving the frame rate improvement found for the PCam dataset of about 15 times. Looking closely at the output however, it is seen that while some features are now resolved, the stripes that are characteristic of actin is not properly recovered. This is likely because the training dataset was not of actin, but rather lots of nuclei images, which makes the model good at recovering spots but not necessarily stripes.

CHAPTER 4
Segmentation of image data of the endoplasmic reticulum

In this section some of the previously described neural network architectures are modified to perform image segmentation and applied to images of the endoplasmic reticulum (ER). The ER is known to be a highly dynamic environment [50] with processes such as the peristaltic flow of luminal proteins [33, 51] and fluctuations of the shape [52].

With the advent of super-resolution microscopy, the structure of the ER is well-known today. The image of the ER in figure 4.1 shows the major structural domains of the ER, including the nuclear envelope, sheets and peripheral tubules [52]. Mutations in ER-shaping proteins can lead to morphological defects, and many of these proteins have been linked to the pathology of human diseases [52]. One example is reticulan that structurally shapes the ER tubules in the peripheral domain and has been found to be involved with Alzheimer's disease [53].

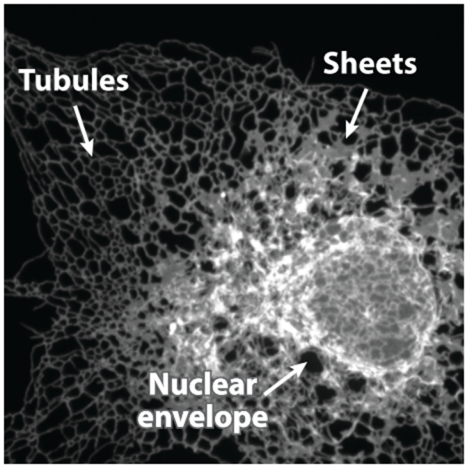

Figure 4.1: Layout of the domains in the endoplasmic reticulum. Image credit [52].

In this chapter the peripheral tubules will be considered with the aim of segmentation. Being able to accurately distinguish between what constitutes the ER and what does not, enables detailed analysis of the shape of the ER and its dynamics.

4.1 Training data

For images with high signal-to-noise ratio it is very easy to perform a binary segmentation of the endoplasmic reticulum simply by pixel intensity thresholding. The difficulty arises

when the image data is so degraded that the tubular structure of the ER is no longer intact, such that a thresholding approach would yield disconnected segmented network. Ideally the segmentation model should learn how to reconstruct the network structure of the ER, thus filling out blanks between parts that are likely to be connected.

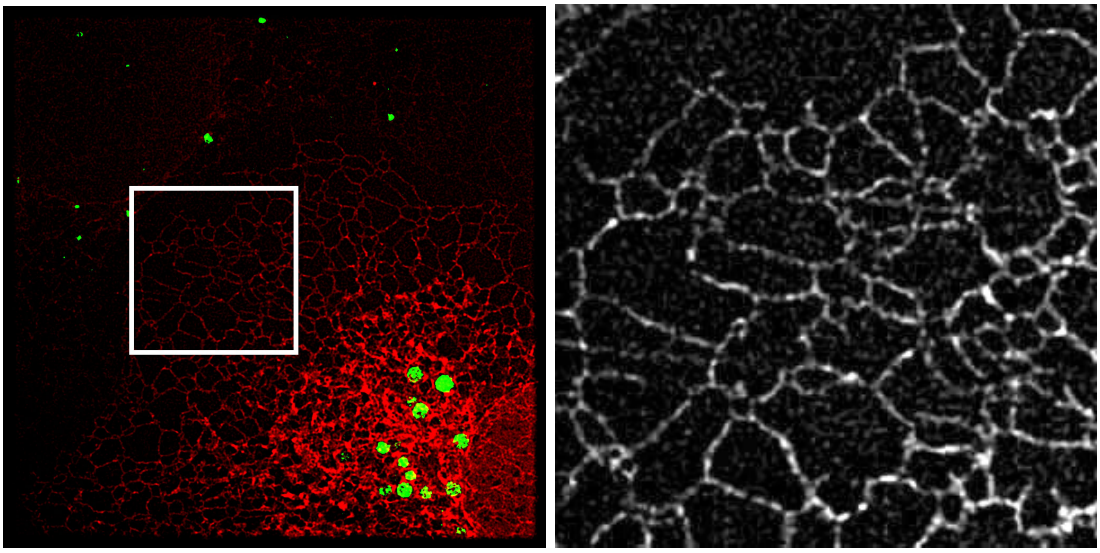

Figure 4.2: Example of a low-quality experimental image that are not useful for training, but can be used for testing the trained model.

One might generate training data in the same way as described in section 3.3 with corresponding pairs of low quality and high quality images of the same sample. One complication with this approach is that it may not be possible to even acquire images of particularly high quality, since the living samples of interest are dynamic and fragile. A way around this could simply be to make the sample static, or perhaps use simulated samples as ground truth.

All the ER images that have been available for training, provided by group member Lu Meng, are of living samples and the quality varies a lot from image to image depending on the capture settings and photo-bleaching. In addition to this there are no duplicate

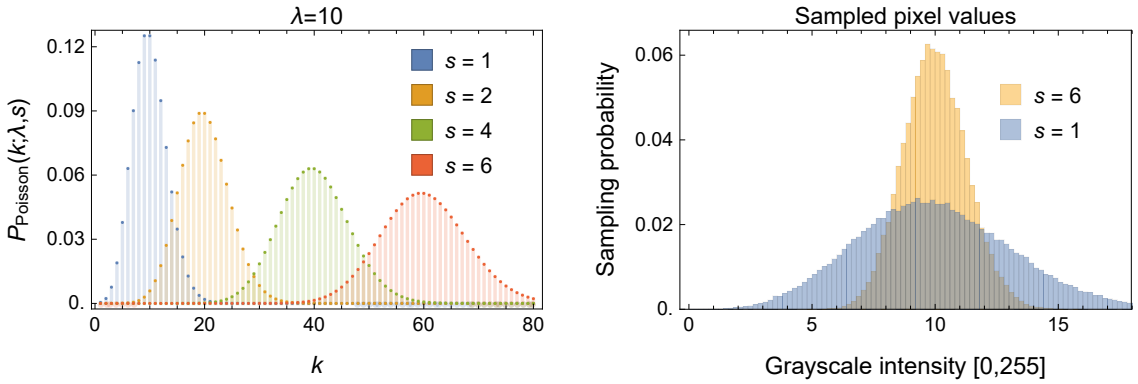

Figure 4.3: Comparisons of different values for the parameter of s in $P_{\text{Poisson}}(k;\lambda,s)$, eq. (4.2).

images that could provide the pairwise low-quality and high-quality training data. This necessitates an approach using synthetic degradation to facilitate the supervised learning. Since some of the provided images are of low quality to begin with, those images can be used for testing purposes, while the cleaner images can be used for training. To get a supervised dataset, the clean images can be used as ground truth while the inputs can be chosen as corresponding synthetically degraded images.

Although the degradation is synthetic the aim will still be for the trained network to be able to segment the original low quality damage, which involves learning to reconstruct the network structure of the ER under high noise levels. For this to work well on the real test set, the synthetic degradation should be as realistic as possible by including the noise sources that are known to affect experimentally acquired images. Considering an example of a low quality experimental image shown on figure 4.2, it is clear that noise is very prevalent, so much that the tubules of the ER appear to become disjointed at some points. In addition to this the fluorescence intensity is also not consistent over the image, presumably due to photobleaching, causing the signal-to-noise ratio to vary significantly. This behaviour is present in several of the experimental images, and it will be approximated as a radially decreasing brightness. This decreasing brightness is assumed to originate from a central point and follow a two-dimensional Gaussian distribution. Both Gaussian and Poisson noise is expected to be present in the experimental data [54], and thus both noise sources are included in the simulated noise model. The characteristic

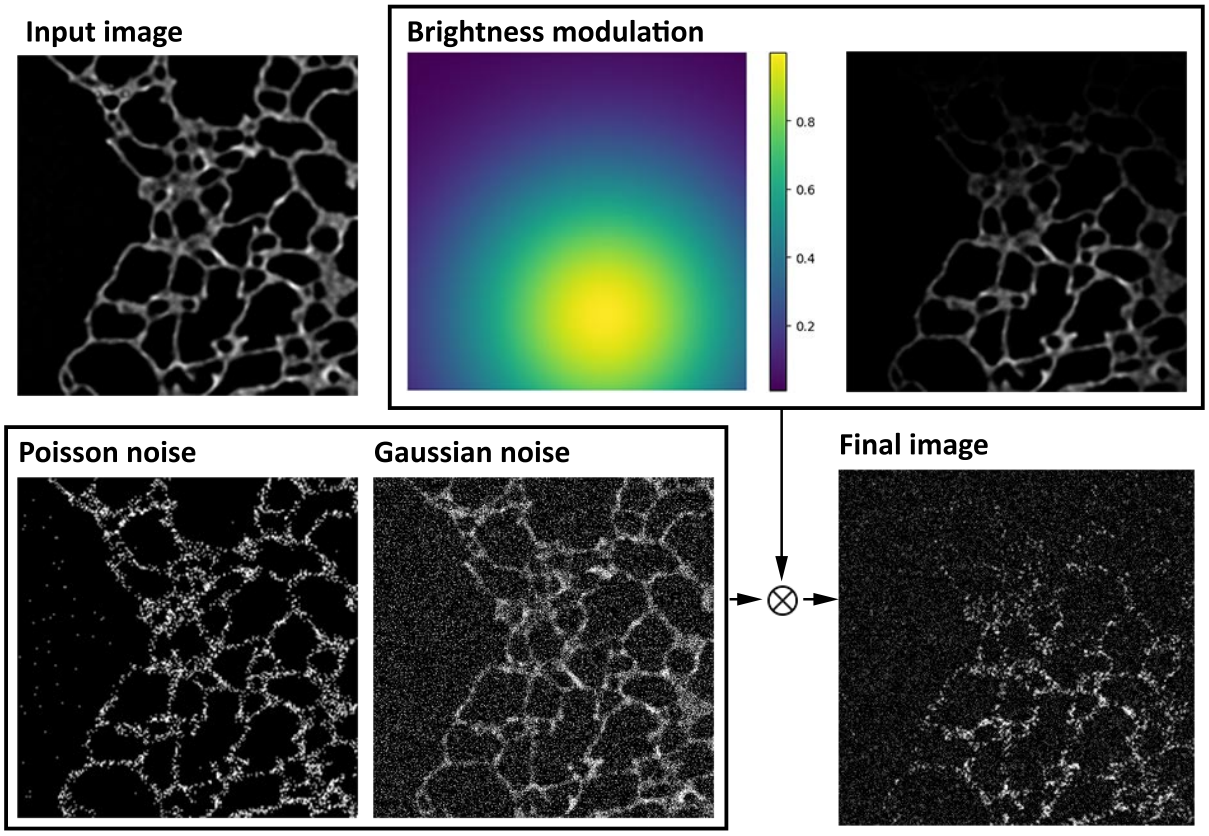

Figure 4.4

of each distribution is hard to identify from the images alone, and thus an assumption has to be made about the parameters of the distributions. To make all the assumptions a bit less arbitrary, the degradation parameters are not set to any single set of values but rather take on random values from broad ranges for every generated synthetic image. The brightness modulated image is given by the function

$$I'[x,y] = I[x,y] \exp\left(-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)\right), \tag{4.1}$$

where $I(x,y)$ is the original image as a function of pixel row and column (x,y) and the randomly generated origin (x_0, y_0) that is somewhere within the bounds of the image, and σ_x and σ_y are the standard deviations of the kernel.

The Gaussian and Poisson noise is generated by sampling the following probability distributions with respective random variables x and k

$$P_{\text{Gaussian}}(x;\sigma,\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad P_{\text{Poisson}}(k;\lambda,s) = \frac{(\lambda s)^k \exp(-\lambda s)}{k!}, \tag{4.2}$$

where for μ is the mean of the Gaussian noise, assumed to be zero, σ is the standard deviation for the Gaussian noise, while λ is the expectation value of the Poisson distribution set to the pixel values of the non-degraded image and s is a parameter controlling the amount of Poisson noise. Given samples from the probability distributions, say x_{Gaussian} and x_{Poisson}, respectively, the resulting noisy images are generated by

$$I_{\text{Gaussian}}[x,y] = I[x,y] + x_{\text{Gaussian}}(\sigma), \quad I_{\text{Poisson}}[x,y] = \frac{x_{\text{Poisson}}(I[x,y],s)}{s}. \tag{4.3}$$

It may not be clear at first how the parameter s affects the resulting Poisson noise given how it enters both (4.2) and (4.3). In (4.2) the s parameter appears as a scaling factor of λ, thus changing the effective expectation value, while the division by s in (4.3) brings the expectation value of the noise back to the value of λ. However, although the final expectation value is unaffected by s, the variance of the samples are affected. This behaviour is shown on figure 4.3, where it is clear that higher values of s lead to less uncertainty in the final sampling distribution, in spite of the probability density function on the left side becoming more broad.

The sequence of degradation steps and their respective effects are shown on figure 4.4. The shown input image is a randomly selected sample among the high-quality samples. The mean and variance of the Gaussian kernel for the brightness modulation is randomly generated. The s parameter for the Poisson noise and the variance, σ, for the Gaussian noise, are also randomly generated. The noise images are simply added and then modulated via multiplication with the Gaussian kernel (4.1).

An example of an image pair, where the degraded image is output from the degradation model, is shown on figure 4.5 with the corresponding binary ground truth segmentation image. The ground truth is obtained by thresholding of pixel intensity of the non-degraded image that is already of high-quality compared to the low-quality example of figure 4.2, which is why thresholding works reasonably well.

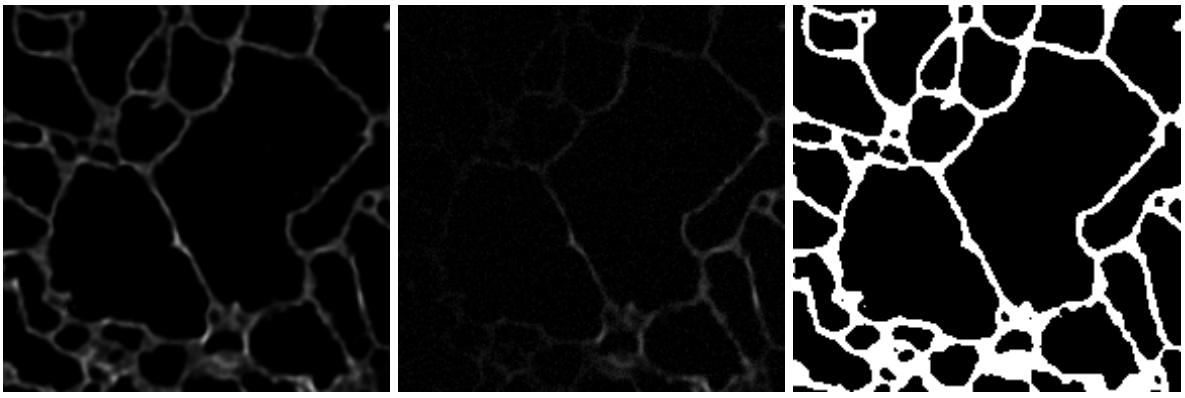

Figure 4.5: Original image and image pair in training dataset. (Left) original relatively high-quality image, (centre) synthetically degraded input image, (right) binarised segmentation image based on thresholding of the original image used as ground truth for training.

4.2 End-to-end CNN segmentation model

The training data set is generated as described in the previous section from 11 relatively high quality 512x512 pixel ER images. A 100 randomly located 192x192 pixel subimages are drawn from each of these source images, where each subimage sample is degraded with a randomly generated set of degradation parameters. In the end a total of 1100 training pairs are available for training. To further enrich the dataset a few data augmentation transformations are applied before feeding a sample into the model, namely any combination of a rotation by 90 degrees, horizontal flip and vertical flip, in total 8 different possible transformations that will make the training dataset slightly more capable.

A separate high-quality source image is reserved for a test set, thus providing 100 subimage samples for testing. Finally a low-quality image akin to that of figure 4.2 is also randomly sampled but not synthetically degraded for a proper validation of the model's functioning.

The neural network architecture that was found to perform the best is a customised version of the super-resolution model RCAN that also was found to work very well when customised to do denoising in section 3. For the model to be able to do segmentation rather than super-resolution, the final block of the diagram in figure 3.2 (also note that the ResBlock is slightly different for RCAN as described in chapter 3) must be a convolution layer that only outputs two channels, one channel for the probability of each of the segmentation classes, i.e. ER or background, being in a given pixel. For simplicity the convolution layer is set to have a kernel of size 1x1, which ensures the output image has the same dimensions as the input image without applying any padding. The convolution operation is then only over the same pixel across all feature maps (the existing filter channels that are input to the convolution layer).

The model is trained for 40 epochs each epoch iterating through all samples in the training dataset with a batch size of 20. The learning rate was initialised to 10^{-4} and halved every 10 epochs. For every epoch the model was evaluated on the test set and the typical performance metrics were calculated, namely the peak signal-to-noise ratio

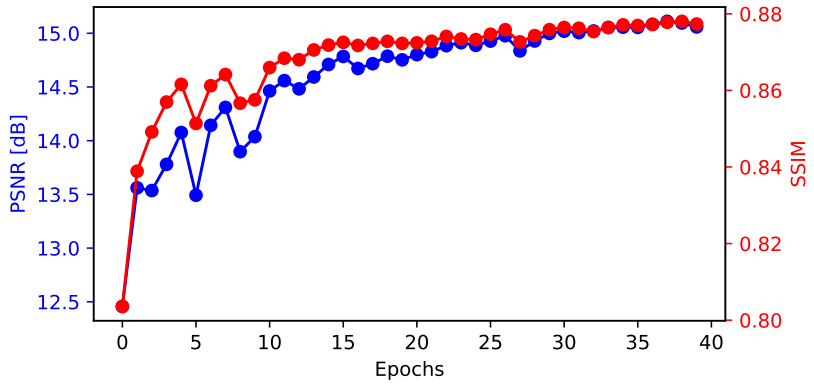

Figure 4.6: Convergence of peak signal-to-noise ratio and structural similarity index during training when model is evaluated on a test set.

(PSNR) and the structural similarity index (SSIM). A convergence plot of these metrics is shown on figure 4.6. As is generally the case the metrics are seen to be highly correlated given the coinciding local extrema. After about 30 epochs the improvements in performance start to be marginal. Within the first 10 epochs there seem to be significant fluctuations, which could indicate that the initial learning rate is too high, meaning the step size in each update is so large that overshooting the local minima becomes a risk. But since the learning rate is set to decrease every 10 epochs, the convergence eventually becomes more stable.

4.3 Results

A first validation is to test the trained network on a synthetically degraded image that is separate from the images in the training dataset. This will test whether the trained network has learned to restore and segment images based on the same type of synthetic degradation, and not only have memorised the specific structures in the ER images of the training dataset.

It can be useful to consider outputs based on the test set during training to see how the model learns. Test results at an early stage after just 3 epochs are shown on figure 4.7. It is clear that the model has not yet learned how to deal with the degradation given how those regions of the ER that are still relatively clear to the human eye just turn up as background in the model output. Furthermore, the region in the left part of the image that is well-lit, near the centre of the Gaussian kernel responsible for the brightness modulation, is not resolved properly in the output. The tubules are very broad and do not appear to have the fine structure seen in the ground truth. This is indicative of neighbouring weights not yet being sufficiently distinctive due to the low number of updates.

After 20 epochs the training has converged much better according to figure 4.6, and indeed the test outputs are found to resemble the ground truths far better, see figure 4.8. Even in the presence of degradation, the model is able to do a segmentation map that is

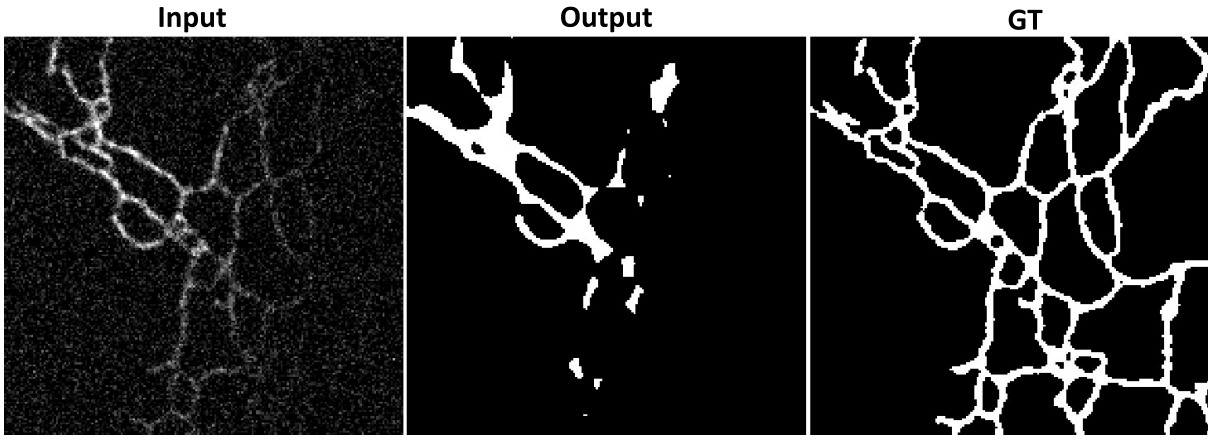

Figure 4.7: Premature test results after training for only 3 epochs.

almost identical to the ground truth showing nearly the same structural resolution and only a few blanks in the top left corner due to the effects of modulating the brightness.

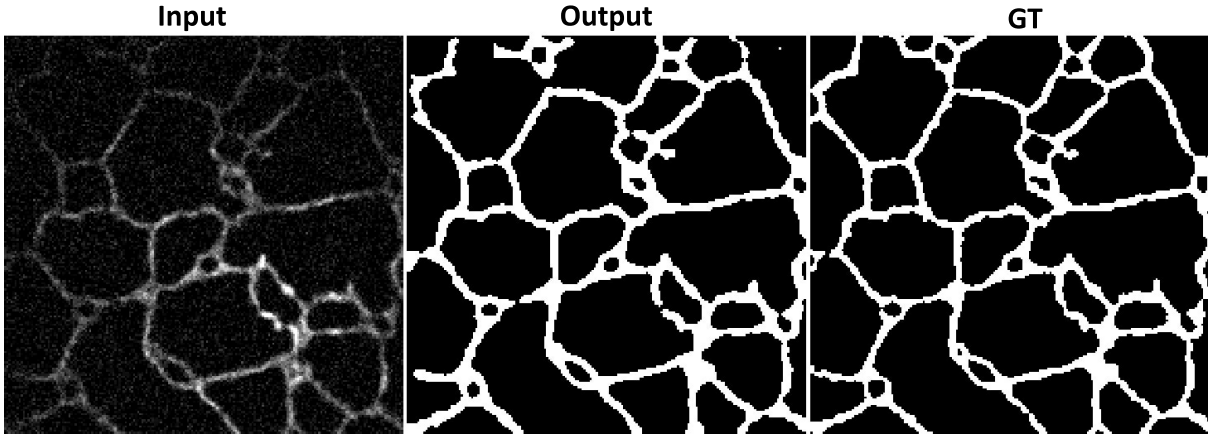

Figure 4.8: Test results after 22 epochs. In spite of the degradation the output closely resembles the ground truth.

Due to the permanent loss of information from the degradation, there are limits to how well the ground truth can be recovered. After 40 epochs where the training is expected to have converged essentially as well as it can for the chosen model, test results still occasionally turn up with significant blanks. One such example is shown in figure 4.9. It is clear from this example that improvements could be made to the model. Even though the information in the dim region on the right may be unrecoverable, the small isolated patches that do occur in the model output are likely to do more harm than good for any further analysis. Since this is undesirable, the model should ideally be modified to reject isolated islands of pixel. This could potentially be remedied relatively easily by customising the loss function used in training to penalise the presence of disconnected patches.

The final validation that will determine whether the model is useful in practice is to test on images that are experimentally degraded. This will test whether the network has

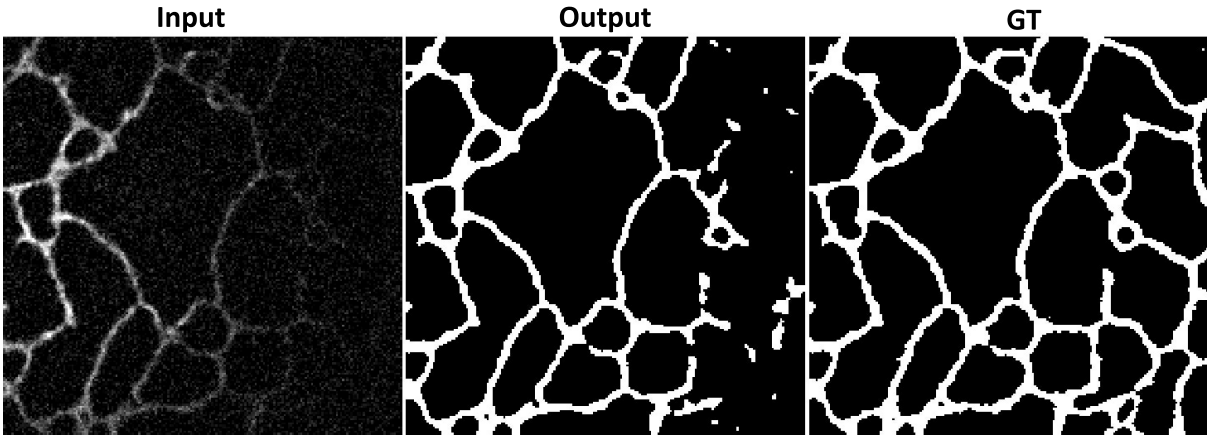

Figure 4.9: Test results after 40 epochs. Although the model is able to recover the ground truth in most of the image, the right side is so degraded that little can be done.

generalised so well that it is able to recognise the typical structure of ER and reconstruct it even when type of degraded input images have never been seen before by the network. These test images are fed directly into the model without any preprocessing in the form of the synthetic degradation. An example of an experimentally degraded input and the corresponding output can be seen in the top row of figure 4.10. The network structure of the ER is clear in the segmentation with only relatively few patches that are disconnected from the network that should ideally either have been rejected or connected via inpainting (content-aware restoration whereby blanks in the image are filled out) in the most realistic way. For comparison a manually fine-tuned thresholded segmentation image is also generated as well as the segmentation result from a built-in Fiji plugin called WEKA – see bottom two rows in the first column of figure 4.10.

The segmentation maps from these alternative methods are seen to have many disconnected tubules. To force some of these tubules to connects in the segmentation map, test outputs were also made with the methods configured to be less conservative – see the second column of the bottom two rows in figure 4.10. For the thresholding approach this means that the cut-off value, for the pixel intensity, for what constitutes background is simply lowered. A 15 % lower cut-off value clearly produces a better connected network structure, but at the price of more isolated patches occurring in the left side of the image in addition to some resolution loss given that all the tubules become significantly broader.

The WEKA plugin uses basic machine learning, by default random forests, to perform the segmentation. The method requires the specification of areas in the input image that correspond to the respective classes, ER and background, after which the classifier can be trained and then run on the entire image. Similarly to the thresholding approach, it is possible to control how conservative the method is towards classifying something as ER. By specifying more dim areas of the ER as part of the example class data that is used for training the classifier, the trained model will become more ”generous” with respect to classifying something as the ER. The areas of the input image used for this example class data can be seen in the miniature images of figure 4.10. The less conservative output is again seen to be have a more well-connected network structure, but with significantly

worse resolution due to the tubules being much broader to the point where the shapes look slightly distorted. However, the presence of isolated patches has become worse, presumably because the selected areas in the example class data are larger, thus allowing the model to filter out regions that are significantly smaller than those selections.

For the example in figure 4.10, the neural network model achieves both an accurate resolution and very few patches, thus not being affected by the trade-off between resolution and number of patches found for the thresholding approach. It is also worth noting that the neural network has not been configured in anyway to deal with the experimental test images, whereas the two other methods were either fine-tuned to achieve the best results or in fact trained on the very image itself, which arguably defeats the purpose of the segmentation tool. As such it is clear that the neural network model is far superior in terms of versatility. For the other examples in the experimental test set the same applies: the output from the neural network model appear clean and consistent, and qualitatively better than the alternative methods, but no experimental ground truth images have been available yet, so a quantitative comparison with performance scores has not been possible.

There is room for improvement, both with regards to the training data but also possibly by modifying the model with a more suitable loss function that penalises disconnected patches, but overall the results are promising given how consistent and versatile the model turned out even with the relatively sparse training data available.

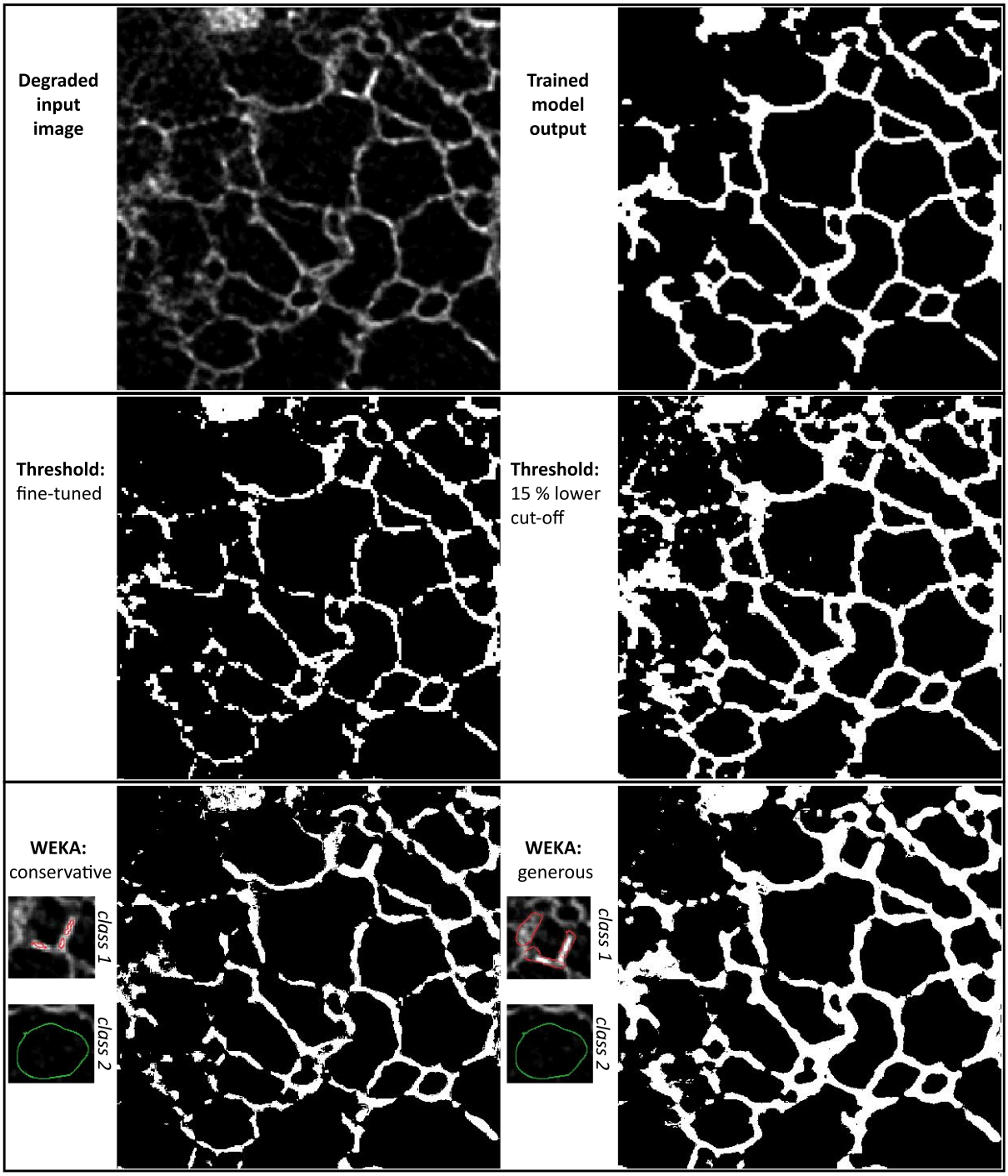

Figure 4.10: Comparison of segmentation maps from different methods made from a test set of experimentally degraded images. The "trained model output" refers to the neural network model that has been implemented and trained on a synthetically degraded training dataset. In spite of this the model is seen to work well even when the synthetic degradation is no longer present. The other methods are more simple: thresholding by pixel value (grayscale intensity) and a plugin in Fiji called WEKA that uses random forests. Both of these other methods have to be manually tweaked for each image that is to be segmented, while the neural network is more versatile and works directly after having been trained on the separate synthetically degraded dataset.

CHAPTER 5

Conclusion

Multiple state-of-the-art models for single-image super-resolution have been implemented, and performance of trained models on biological data is encouraging. Modifying these super-resolution architectures to do denoising is found to yield better performance than the commonly used U-Net models [7, 3, 6]. By obtaining an analytical description of image quality as a function of exposure time based on statistics from experimentally acquired images, it is possible to estimate the equivalent exposure time of a restoration output. It has been found in this way that deep learning may increase acquisition rate by as much as 15 times. Transfer learning has been attempted by training on a completely different dataset, but then evaluating the trained model on an experimentally acquired dataset. Since the noise characteristics of the two datasets are similar, the model still performed relatively well even though it had never seen the type of samples in the experimentally degraded dataset. However, certain features were unable to be recovered when the experimentally degraded images were restored, such as the characteristic stripes of actin. The training dataset consisted primarily of images of cell nuclei, and it seems that the trained model is indeed good at recovering spots. A possible improvement could be to retrain the model with a smaller dataset of images with the sample type of interest. In terms of performance, it might be worth trying to combine the efficiency of the U-Net architectures with the superior performance of the super-resolution inspired network architectures.

Finally, work on segmentation of endoplasmic reticulum (ER) images has also been reported. A network trained on images that were synthetically degraded according to a degradation model, featuring brightness modulation in addition to regular noise sources, has been found capable of segmenting experimentally degraded images. The model outputs binarised segmentation maps with an accurate resolution and very few noisy patches. When compared to traditional methods for segmentation, this model is much more versatile, since no tweaking or configuring was required before applying it to experimental test images and yet the outputs were qualitatively better than outputs from alternative methods. These alternative methods were either fine-tuned to achieve the best results or trained on the very image itself. No experimental ground truth images have been available yet, and in a more rigorous comparison of the model with the alternative methods it would be desirable to have such ground truth images. Directions for improvements could be on both the data side as well as the modelling side. The training dataset used was based on just 11 source images, so a more rich dataset should be possible to acquire. With respect to modelling, the training algorithm of the model could likely be improved, for instance by using a custom loss function that penalises disconnected patches to reduce the occurrence of isolated, noisy patches in the resulting segmentation maps.

CHAPTER 6

Future work

Based on the results so far, the progress in denoising and segmentation is found to be promising enough that more research will be done in those areas. The aim is that both avenues lead to a publication. For that to happen more experimental work is required, by the author or collaborators, which will feed back into achieving better model performance and hopefully facilitate publishable analysis.

Generative Adversarial Network (GAN) models have not yet been properly investigated in this PhD project due to concerns about the amount of distortion they are able to introduce. However, the trade-off between reconstruction error and perceptual quality is starting to be better understood [24], so with appropriate constraints these GAN models could be an interesting avenue for further research.

There are also more novel neural network architectures that have yet to be applied to image restoration, namely graph neural networks, and more specifically graph convolutional networks [55] and graph attention networks [56]. The advantage of a graph is that it is more universal than an image representation, and thus other information could be encoded along with the image data perhaps physical knowledge based on modelling or empirical data such as functional measurements of the sample. Be it with graphs or without, it seems sensible to combine a learning-based method with a model-based approach (i.e. a physical model, not a statistical one), since some optical and biological facts about the imaging process and the sample are known a priori, for instance the optical transfer function (OTF) and the typical noise sources etc. Such graph models or learning/model-based hybrid approaches could be yet another avenue worthy of further investigation.

In addition to the directions for further research into restoration discussed above, the prospect of a new application will also be researched, namely the reconstruction of super-resolution structured illumination microscopy (SIM) images using neural networks. Reconstruction of microscopy images with neural networks is something that is already done for photo-activated localisation microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) [10, 11, 12] as well as Fourier ptychography [13, 14]. However, there is nothing published to the knowledge of the author of this report regarding SIM reconstruction. SIM reconstruction are commonly done through plugins in ImageJ/Fiji such as FairSIM [57], which require the specification of multiple optical parameters although some quantities can normally be estimated, like the optical transfer function. Furthermore, the reconstruction will use a deconvolution algorithm, normally Richardson-Lucy deconvolution or a Wiener filter, which introduces another set of parameters including the Wiener parameter, apodisation cutoff and bend as well as the number of iterations of deconvolving. In the end there will be characteristic SIM artefacts, and a careful tweaking of all the parameters are required to minimise the presence of these

artefacts and enhance the visibility of the sample structure. Furthermore, uncertainty in the optical parameters may also lead to sub-optimal output necessitating further tweaking. Ideally all this tweaking could be alleviated by using a neural network optimised to produce the most accurate results. Given the capability of neural networks to restore degradation as seen in this report, it is likely that avoiding artefacts during a SIM reconstruction would be possible. If trained properly one could imagine that the neural network could become versatile enough to perform a robust and faithful reconstruction when optical parameters are uncertain or even not provided at all, which is sometimes referred to as blind SIM [58].

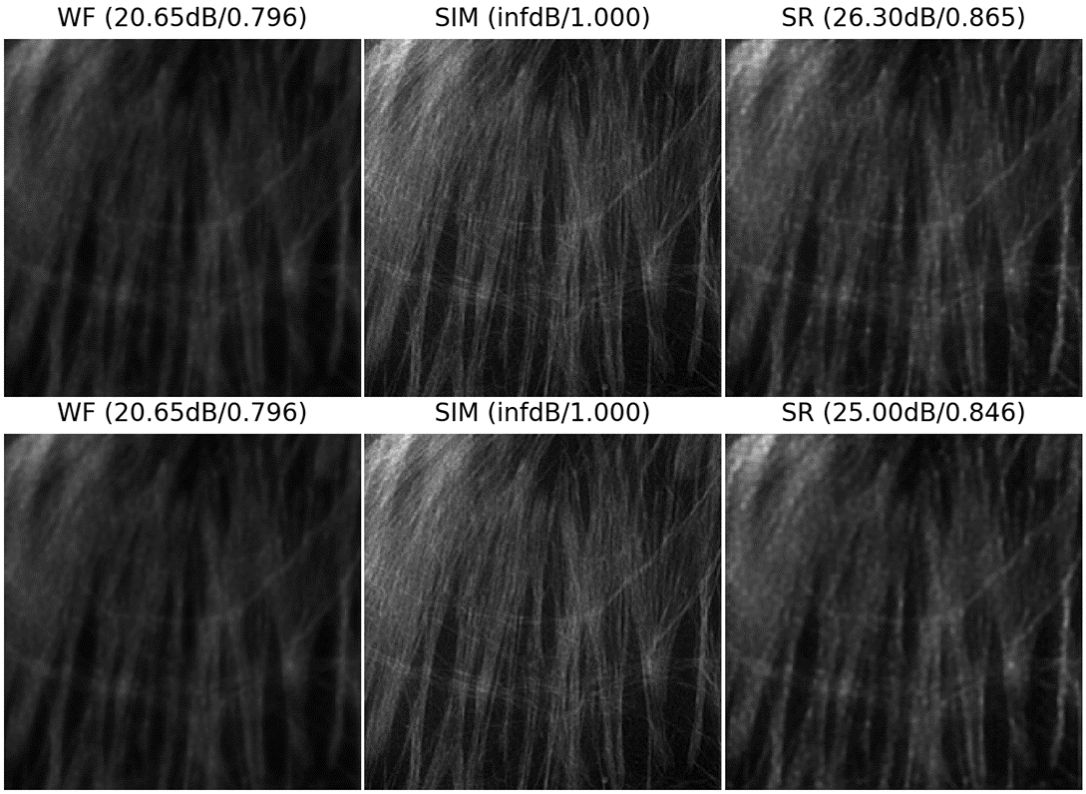

Figure 6.1: Reconstruction of a super-resolved image from 9 raw SIM frames acquired experimentally (top row) and, as a control, from 1 widefield input image (bottom row). First column shows shows the input images in the two different cases, where all channels have been merged by addition, which makes the two equal since the widefield image approximation is exactly the mean of the 9 SIM frames to begin with. The centre column shows the fairSIM reconstructed super-resolved target. Right column shows the model outputs, which are very similar between the two model with only a marginal improvement for the model that is trained on all 9 raw SIM frames.

In this PhD project there has been some work looking into this. The training dataset could simply be generated by simulating the imaging process of a SIM microscope, in which case both the patterned raw frames and the ground truth would be available. A group member of LAG, Edward Ward, has provided code to do just this. There are also several experimental SIM images open-sourced at GitHub related to the fairSIM project,

and these images can as well be used for training by feeding in the raw SIM images and using a fairSIM reconstructed image as target. In this latter approach however, the trained neural network would end up simply reproducing the output of the fairSIM plugin including the artefacts, thus making the simulated images more appealing for training. In spite of that it could be a reasonable sanity check to reproduce the outputs of the fairSIM plugin. This has briefly been attempted in this PhD project, and example outputs can be seen on figure 6.1 for two identical models that are trained differently: one is trained on raw SIM frames as input, 9 frames in total due to 3 angles and 3 shifts in the SIM pattern, while the other is trained on the mean of all raw frames, approximating a widefield image, as input in order to have a control group of results. As can be seen the super-resolved versions, SR, are very similar in terms of PSNR and SSIM. Therefore it is clear that the model does not fully exploit all the frequency information in the raw SIM images when the same output can roughly be generated by training on a single approximate widefield image with all the patterns averaged.

Since this has only been briefly attempted, it is not surprising that it does not work better at this point. But the potential of this application is deemed to be interesting enough that it will be further investigated, and if the results are promising a more thorough study will be made. This avenue is therefore also included in the future plans for this PhD project. An overview of the milestones planned for the year to come are summarised in the Gantt chart of figure 6.2.

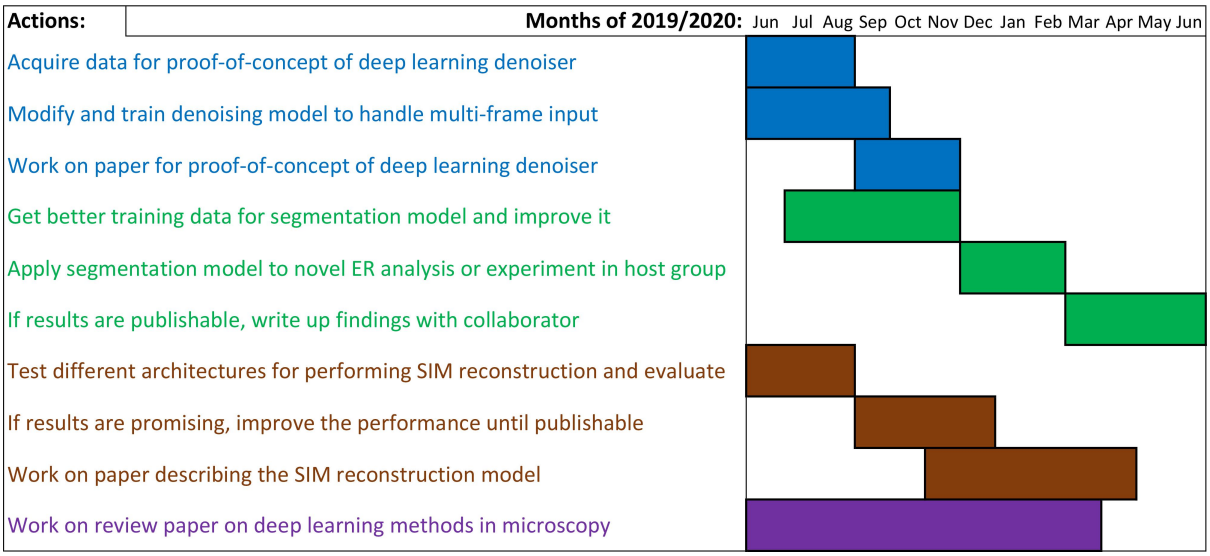

Figure 6.2: Overview of milestones planned for the second year of the PhD.

APPENDIX A

Lecture list

As a 1st Year PhD student it is required to attend at least 50 % of the total number of seminars held in the department over Lent and Michaelmas Terms in order to be registered for the PhD. The lectures that I have attended include:

- Professor Prof Hanssjörg Freund
  Department of Chemical and Biological Engineering
  CRT Friedrich-Alexander-Universität Erlangen-Nürnberg
  2 pm Wednesday 28th November
  Optimal design of catalytic reactors and structured catalysts

- Dr Silvia Vignolini
  Department of Chemistry
  University of Cambridge, UK
  2 pm Wednesday 20th February
  Bio-inspired photonics: from nature to applications

- Professor Ed Louis
  Centre for Genetic Architecture of Complex Traits
  University of Leicester, UK
  2 pm Friday 1st March
  Biotechnology meets Breeding in Baker's Yeast

Bibliography

[1] Kai Zhang, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a fast and flexible solution for CNN-Based image denoising". In: *IEEE Transactions on Image Processing* 27.9 (2018). arXiv: 1710.04026.

[2] Kai Zhang et al. "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising". In: *IEEE Transactions on Image Processing* 26.7 (2017). arXiv: 1608.03981.

[3] Jaakko Lehtinen et al. "Noise2Noise: Learning Image Restoration without Clean Data". In: (2018). arXiv: 1803.04189.

[4] Chao Dong et al. "Image Super-Resolution Using Deep Convolutional Networks". en. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.2 (February 2016).

[5] Victor Lempitsky, Andrea Vedaldi, and Dmitry Ulyanov. *Deep Image Prior*. Technical report. 2018, pages 9446–9454. arXiv: 1711.10925v3.

[6] Martin Weigert et al. "Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy". en. In: (July 2018).

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (May 2015). arXiv: 1505.04597.

[8] Chawin Ounkomol et al. "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy". In: *Nature methods* 15.11 (2018).

[9] Dan Ciresan et al. "Deep neural networks segment neuronal membranes in electron microscopy images". In: *Advances in neural information processing systems*. 2012, pages 2843–2851.

[10] Elias Nehme et al. "Deep-STORM: super-resolution single-molecule microscopy by deep learning". en. In: *Optica* 5.4 (April 2018).

[11] Wei Ouyang et al. "Deep learning massively accelerates super-resolution localization microscopy". In: *Nature Biotechnology* 36.5 (2018).

[12] Nicholas Boyd et al. "DeepLoco: Fast 3D Localization Microscopy Using Neural Networks". en. In: (February 2018).

[13] Armin Kappeler et al. "PtychNet: CNN based Fourier ptychography". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pages 1712–1716.

[14] Jizhou Zhang et al. "Fourier ptychographic microscopy reconstruction with multiscale deep residual network". In: *Optics Express* 27.6 (2019).

[15] Kaiming He et al. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pages 1026–1034.

[16] Oren Z Kraus, Jimmy Lei Ba, and Brendan J Frey. “Classifying and segmenting microscopy images with deep multiple instance learning”. In: *Bioinformatics* 32.12 (2016).

[17] Claude E. Duchon. “Lanczos Filtering in One and Two Dimensions”. In: *Journal of Applied Meteorology* 18.8 (1979).

[18] Ian J Goodfellow et al. *Generative Adversarial Nets*. Technical report.

[19] Christian Ledig et al. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”. In: (September 2016). arXiv: 1609.04802.

[20] Seong Jin Park et al. *SRFeat: Single Image Super-Resolution with Feature Discrimination*. Technical report. 2018, pages 455–471.

[21] Xintao Wang et al. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. Technical report. 2018. arXiv: 1809.00219.

[22] Mehdi S.M. Sajjadi, Bernhard Scholkopf, and Michael Hirsch. *EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis*. Technical report. 2017, pages 4501–4510. arXiv: 1612.07919.

[23] Bee Lim et al. *Enhanced Deep Residual Networks for Single Image Super-Resolution*. Technical report. 2017, pages 1132–1140. arXiv: 1707.02921.

[24] Subeesh Vasu, Nimisha Thekke Madam, and Rajagopalan A. N. *Analyzing Perception-Distortion Tradeoff using Enhanced Perceptual Super-resolution Network*. Technical report. 2018. arXiv: 1811.00344.

[25] Yulun Zhang et al. *Image super-resolution using very deep residual channel attention networks*. Technical report. 2018, pages 294–310. arXiv: 1807.02758.

[26] Yochai Blau et al. *2018 PIRM Challenge on Perceptual Image Super-resolution*. Technical report. 2018. arXiv: 1809.07517.

[27] Olga Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision* 115.3 (2015). arXiv: 1409.0575.

[28] Eirikur Agustsson and Radu Timofte. *NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study*. Technical report. 2017, pages 1122–1131.

[29] D Martin et al. “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics”. In: *Proc. 8th Int’l Conf. Computer Vision*. Volume 2. July 2001, pages 416–423.

[30] Bastiaan S. Veeling et al. “Rotation equivariant CNNs for digital pathology”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11071 LNCS (June 2018). arXiv: 1806.03962.

[31] Diederik P Kingma and Jimmy Lei Ba. *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*. Technical report. arXiv: 1412.6980v9.

[32] Zhou Wang, A C Bovik, and H R Sheikh. "Image quality assessment: From error measurement to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (2004).

[33] David Holcman et al. "Single particle trajectories reveal active endoplasmic reticulum luminal flow". In: *Nature cell biology* 20.10 (2018).

[34] Michaela Mickoleit et al. "High-resolution reconstruction of the beating zebrafish heart". In: *Nature methods* 11.9 (2014).

[35] Lo\"\ic A Royer et al. "Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms". In: *Nature biotechnology* 34.12 (2016).

[36] Thomas A Planchon et al. "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination". In: *Nature methods* 8.5 (2011).

[37] Antoni Buades, Bartomeu Coll, and Jean-Michel J.-M. Morel. "A non-local algorithm for image denoising". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* 2.0 (2005).

[38] Michael Lindenbaum, M Fischer, and A Bruckstein. "On Gabor's contribution to image enhancement". In: *Pattern Recognition* 27.1 (1994).

[39] Simon S Haykin, Bernard Widrow, and Bernard Widrow. *Least-Mean-Square Adaptive Filters*. Volume 31. Wiley Online Library, 2003.

[40] Pietro Perona and Jitendra Malik. "Scale-space and edge detection using anisotropic diffusion". In: *IEEE Transactions on pattern analysis and machine intelligence* 12.7 (1990).

[41] Norbert Wiener and Mass.) Massachusetts Institute of Technology (Cambridge. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. Technology Press, 1950.

[42] David L Donoho. "De-noising by soft-thresholding". In: *IEEE transactions on information theory* 41.3 (1995).

[43] Ling Shao et al. "From Heuristic Optimization to Dictionary Learning : A Review and Comprehensive Comparison of Image Denoising Algorithms". In: *IEEE Transactions on Cybernetics* 44.7 (2014).

[44] Sven Grewenig, Sebastian Zimmer, and Joachim Weickert. "Rotationally invariant similarity measures for nonlocal image denoising". In: *Journal of Visual Communication and Image Representation* 22.2 (2011).

[45] Pierrick Coupé et al. "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images". In: *IEEE transactions on medical imaging* 27.4 (2008).

[46] Michael Elad and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries". In: *IEEE Transactions on Image processing* 15.12 (2006).

[47] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. *Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections*. Technical report. 2016. arXiv: 1606.08921.

[48] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: (December 2015). arXiv: 1512.03385.

[49] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. "Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections". In: (June 2016). arXiv: 1606.08921.

[50] Sergei I Bannykh and William E Balch. "Membrane dynamics at the endoplasmic reticulum–Golgi interface". In: *The Journal of cell biology* 138.1 (1997).

[51] S Nadeem and E N Maraj. "The mathematical analysis for peristaltic flow of nano fluid in a curved channel with compliant walls". In: *Applied Nanoscience* 4.1 (2014).

[52] L M Westrate et al. "Form Follows Function : The Importance of Endoplasmic Reticulum Shape". In: (2015).

[53] Yvonne S Yang and Stephen M Strittmatter. "The reticulons: a family of proteins with diverse functions". In: *Genome biology* 8.12 (2007).

[54] Anna Jezierska et al. "Poisson-Gaussian noise parameter estimation in fluorescence microscopy imaging". In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2012, pages 1663–1666.

[55] Thomas N. Kipf and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks". In: (September 2016). arXiv: 1609.02907.

[56] Petar Velickovi et al. "Graph Attention Networks". In: 2005 (2018). arXiv: arXiv:1710.10903v3.

[57] Marcel Müller et al. "Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ". In: *Nature communications* 7 (2016).

[58] Li-Hao Yeh, Lei Tian, and Laura Waller. "Structured illumination microscopy with unknown patterns and a statistical prior". In: *Biomedical optics express* 8.2 (2017).