# Structured Illumination Microscopy Image Processing using Deep Learning

James Hughes
Supervised by Dr Edward Ward

## Project Report

MPhil, Data Intensive Science

Department of Physics & Department of Chemical Engineering and Biotechnology
University of Cambridge
United Kingdom
28th June 2024

# Acknowledgements

Firstly I would like to thank my supervisor, Dr Edward Ward, for all of his support over the course of this project. The project involved a lot of concepts from microscopy and image processing that were very new to me, but Dr Ward made it clear from early on in the project that this would not be a problem, and was quick to provide reading materials to help me get to grips with the subject. Having been trained in Mathematics during my time as an undergraduate, the chance to go into the laboratory and capture real microscope images that were later used in the work was incredibly exciting. Dr Ward was eager to provide this opportunity and welcomed me to the Chemical Engineering and Biotechnology (CEB) Department and his research group.

I also had the opportunity to attend some of the Laser Analytics Group (LAG) lab meetings, where I had the privilege of learning about some of the world-leading research being undertaken by the group. Later, I shared details about my own project in two presentations to the group. I would like to thank all of the members of the LAG for welcoming me, listening to my presentations and providing great feedback. In particular I wish to thank Professor Clemens Kaminski for his helpful suggestions and words of encouragement.

I would also like to thank Jeremy Wilkinson, Esther Gray, and Emilio Luz-Ricca. I had very insightful conversations about the project with all of them that helped me see the work in a new light.

Lastly, I would like to thank my parents for being a continual source of support and strength throughout my education, in particular for encouraging me to make the most of every opportunity that comes my way.

**Abstract**

Structured illumination microscopy (SIM) produces images whose resolution exceeds the Abbe diffraction limit imposed on widefield images. However, SIM imaging of dynamic cellular processes is restricted by phototoxicity effects, which limit the maximum duration of such time-lapses. In 2023, Li et al. developed a 'two-step denoising' approach to SIM image processing, which enables greatly reducing the illumination intensity of the microscope, and in turn using deep learning to recover the lost signal in the image. Firstly, this project presents a data processing pipeline which implements their method using PyTorch. This pipeline is documented, modular, and open-source, enabling researchers to apply the method to different datasets, or develop extensions to the work. Secondly, this project investigates the reproducibility of this method, by analysing its performance on two datasets: microtubule images acquired using a 2D SIM microscope and synthetic 3D SIM imagery simulated by using data from the Visible Human Project as ground-truth. Results indicate that although the first-step reconstructions can improve the fidelity compared to the low SNR inputs, this is potentially dependent on the variety of the biological structures present in the training data. Moreover, while the performance of the full two-step denoising method produces images qualitatively close to ground-truth, with noticeably reduced reconstruction artefacts compared to the raw and first-step reconstructions, there is little quantitative evidence that the second step increases image fidelity, calling into question the reliability of this second network.

# Contents

# 1 Introduction

Fluorescence microscopy is an essential tool for microbiologists, enabling them to view complex biological phenomena unfolding at the sub-cellular level. Fluorescent dyes are employed to attach to specific organic compounds, which then release photons in response to illumination from a laser at a suitable wavelength, producing images that highlight specific structures of interest to researchers. As a type of optical microscopy, the resolution of these systems is limited by the effects of diffraction. This limit was quantified by Abbe [1] in 1873 as a minimal resolvable distance between two points,

$$d = \frac{\lambda}{2NA}$$

where $\lambda$ refers to the emittance wavelength, and $NA$ refers to the numerical aperture, a property of the optical system and the imaging medium. This resolution limit is summarised by the optical transfer function $O(\vec{k})$ of the microscope, which describes the set of spatial frequencies of the sample structure that can be captured by the optical system, and to what extent they are attenuated in the resulting image (in frequency space). Axial resolution of optical microscopes is typically much worse than their lateral resolution. This fact is evidenced by the optical transfer function's omission of most k-vectors that lie along and near to the z-axis, a phenomenon referred to as the 'missing-cone problem'. This is further compounded in practice with issues such as spherical abberation. This represents a serious obstacle to researchers attempting to view cell dynamics in greater detail.

Structured Illumination Microscopy (SIM) is a technique that combines a specialised microscope set-up, alongside computational processing of the acquired images, in order to surpass the classical Abbe diffraction limit. The theoretical foundations of the technique were first established in 2008 [2], but since then there have been a range of improvements made to the technique [citations]. While SIM does not necessarily provide the greatest improvements in resolution compared to other methods such as confocal, it has other advantages for researchers interested specifically in capturing imagery of dynamic biological processes over extended periods. This relates primarily to the issue of phototoxicity effects. Every time a fluorescence microscopy image of a cell sample is taken, the cell itself is bleached and damaged in the process. This is particularly troublesome when one wishes to view dynamic processes in live cells, because the very process of imaging has an effect on

the process being captured, thereby limiting the duration of imagery that can be obtained that is faithful to the true process. SIM offers a trade-off between resolution improvements and low photo-toxicity effects.

The paper by Li et al. [3] explores augmenting the SIM image processing pipeline with deep-learning techniques to improve this trade-off. Their research explores multiple ways in which hardware and computation can be used to improve the resolution of SIM imaging. This project investigates their 'two-step denoising method'. In this method, the illumination intensity of the SIM system is set to around 10 times lower than usual, in order to mitigate phototoxicity effects. In turn, they train two networks to denoise the acquired and reconstructed images, in order to compensate for the noise introduce by the low illumination dose and reclaim lost image resolution.

This project aims to present a full pipeline that implements their method. The tools developed in the repository aim to make this software accessible to other research groups looking to apply it to their own data, with minimal work required for set-up, and compatibility with common tools used for SIM image processing. Moreover, by adopting an open-source ethos, this project should enable the pipeline to be extended upon easily. The second main objective of this work is to study the reproducibility of the results claimed in the original research. In particular, Li et al. assert that this method:

- mitigates the presence of artefacts in the reconstructions of low SNR acquisitions,
- improves the resolution of SIM imaging, particularly the axial resolution, and,
- increases the fidelity of reconstructed images by up to 3.63dB (PSNR) on average.

This work sets out to apply the method to both microtubule data collected from a 2D SIM system, as well as synthetically generated 3D SIM data, and compare the resulting reconstructions.

# 2 Methods

## 2.1 SIM Reconstruction process

Structured Illumination Microscopy stands in contrast to the conventional approach of using a uniform illumination to produce a micrograph image.

Instead, SIM microscopes usually employ a spatial light modulator (SLM) to produce a striped illumination pattern, whose spacing is close to the Abbe diffraction limit of resolution. When the light illuminates the sample causing it to fluoresce, the excitation pattern's spatial frequencies interfere with the high spatial frequencies of the structures in the sample, causing information to be exposed as lower frequency features in the resulting image [2]. Figure 1 demonstrates this effect with Moiré fringes, an interference pattern with lower spatial frequency than the two patterns that generate it.

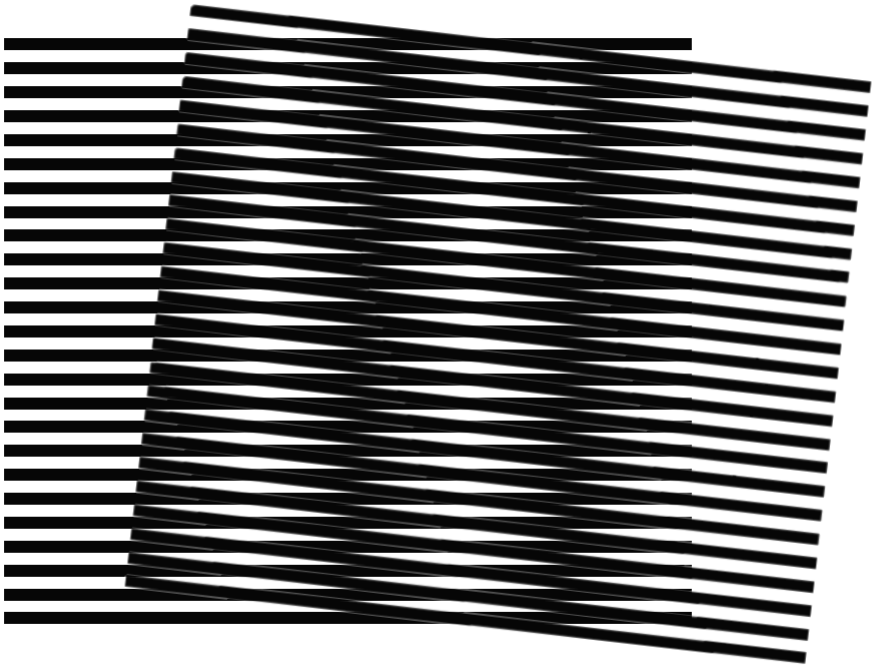

Figure 1: Moiré Fringes

In order to correctly interpret this interference effect, and reconstruct a super-resolved image, multiple images need to be acquired from the microscope and analysed in the Fourier domain [2]. When reconstructed properly, there will be an improvement of lateral resolution in the direction of the k-vector of the pattern. Therefore, when acquiring images for 2D SIM, it is almost always 3 groups of images that are acquired, using patterns whose orientations are angled at multiples of $2\pi/3$ radians, to obtain an near-isotropic

improvement in *lateral* resolution, and the same is true in 3D SIM. Within these groups, the images are acquired with the illumination pattern having a different phase each time, typically with a constant offset between phases.

The reconstruction involves six key steps:

1. parameter estimation,
2. fourier transform,
3. band separation,
4. Wiener filtering,
5. apodization, and,
6. inverse Fourier transform.

Parameter estimation is primarily concerned with the position of the illumination pattern, including the phase, angle, and modulation depth. This is more accurate than measuring these quantities in the physical system which would require high degree of care and precision. Then the image is converted to the frequency domain.

Denoting the image intensity by $D(\vec{r})$, the pattern k-vector and phase by $\vec{p}$, $\phi_n$, the modulation depth by $a_m$, the density of the fluorescent substance as $S(\vec{r})$ and the point-spread function by $H(\vec{r})$, we see that the effect of the optical system on the 'true' ground-truth structure $S$ is to multiply it with the excitation pattern, and then convolve with the point-spread function:

$$D_n(\vec{r}) = \sum_{m=-M}^{M} S(\vec{r}) a_m \exp(im(2\pi\vec{p}\cdot\vec{r} + \phi_n)) \otimes H(\vec{r})$$

Utilising the Convolution Theorem this becomes

$$\tilde{D}_n(\vec{k}) = \sum_{m=-M}^{M} \exp(im\phi_n) a_m \tilde{S}(\vec{k} - m\vec{p})\tilde{O}(\vec{k})$$

In turn, with sufficient acquired images at different phases, namely M, this can be used to solve a fully determined set of linear equations for

$$\tilde{S}(\vec{k} - m\vec{p})\tilde{O}(\vec{k}) \qquad m = -M, \ldots, M-1, M$$

This constitutes the band separation step, and explains why 2D SIM uses 3 sets of 3 images, while 3D SIM uses 3 sets of 5 images; the number of different phases used in imaging must correspond to the number of delta peaks that represent the illumination pattern in Fourier space, in order to set up a fully-determined system of linear equations [4].

The steps of Wiener filtering and apodization are used to combine the separated bands of the image, ...

The parameters used for the reconstructions are shown in Table 1

| Parameter | 2D Dataset (1) | 2D Dataset (2) | 3D Dataset |
|---|---|---|---|
| *NA* | 1.1 | 1.1 | 1.12 |
| Pixel width (nm) | 107 | 107 | 50 |
| Wavelength (nm) | 488 | 561 | 464 |
| OTF param. | 0.15 | 0.15 | 0.15 |
| APO cutoff | 1.59 | 1.68 | 1.82 |
| APO bend | 1.0 | 1.0 | 1.0 |
| Wiener parameter | 0.05 | 0.05 | 0.05 |
| RL Iterations | 5 | 5 | 5 |

Table 1: Parameters used to reconstruct the images in fairSIM.

## 2.2 Data

In the original work, Li et al. acquired pairs of high and low SNR images from a 3D SIM system, in order to train the networks. This project takes a slightly different approach, simulating the increased image noise from a lower the illumination intensity in-silico. This makes the acquisition of the training data much faster, and avoids the need for image-pair registration, along with the errors that this could induce. A low SNR image is simulated from the ground-truth high SNR image on a pixel-by-pixel basis: a pixel whose value is $N$ in the high SNR image is set to a random draw of a Poisson random variable whose rate parameter is $N/s$, where $s$ is some chosen scale factor constant across all pixels and images. In both datasets this scale factor is set as 20.

The method is applied to a dataset of 2D SIM images in the first instance, in which the fluorescent dye marks microtubules within the cells. This contains more high spatial frequency content that can be resolved by SIM than

an earlier dataset acquired for the project, in which viruses and the cell membrane of the host cell were highlighted. The samples were illuminated with visible light at 488nm and 561nm.

In the second case, the Visible Human Dataset[1] is used to generate synthetically acquired 3DSIM micrographs. This dataset was released in 1994 [cite] and provides images of human cadavers prepared as a series of thousands of thin cross sections. The work uses the 70mm photographs of the female body dataset, image 2000 through to 2383. While this imagery does not capture *microscopic* biological structures, those biological structures present are complex enough to yield an approximation to the image features one might expect from a typical SIM micrograph of a cell. These images 300 , RGB 2

These images were downloaded, cropped into 256x256 squares 3 and stacked into image volumes of size (128, 256, 256)

This was generated with the help of another student's code. The code was modified to include only the most essential parts, and to generate a 3D volume of SIM acquisition stacks (15 each)

Talk about modification with focal length stuff etc. Maybe even the limitations of this data (is it representative? is the physical system faithfully simulated? More for discussion section perhaps)

Splitting up of data.

## 2.3 RCAN

In both instances, the denoising models were implemented using the residual channel attention network (RCAN) architecture, which first emerged in the computer vision literature in 2018 [5].

Li et al. employed a slight variant of the RCAN implemented more recently [6]. In particular, this variant is re-implemented as a denoising model rather than a super-resolution model, so there is no upsampling of the images over the course of the network architecture. However, the code provided alongside this more recent work is written in TensorFlow. In order to make the code compatible with the software available (specifically the versions of CUDA and cuDNN available) on the HPC platform used, this codebase was migrated to PyTorch. This also has the advantage of making the software more accessible to other researchers wishing to develop in PyTorch.

[1] Courtesy of the U.S. National Library of Medicine

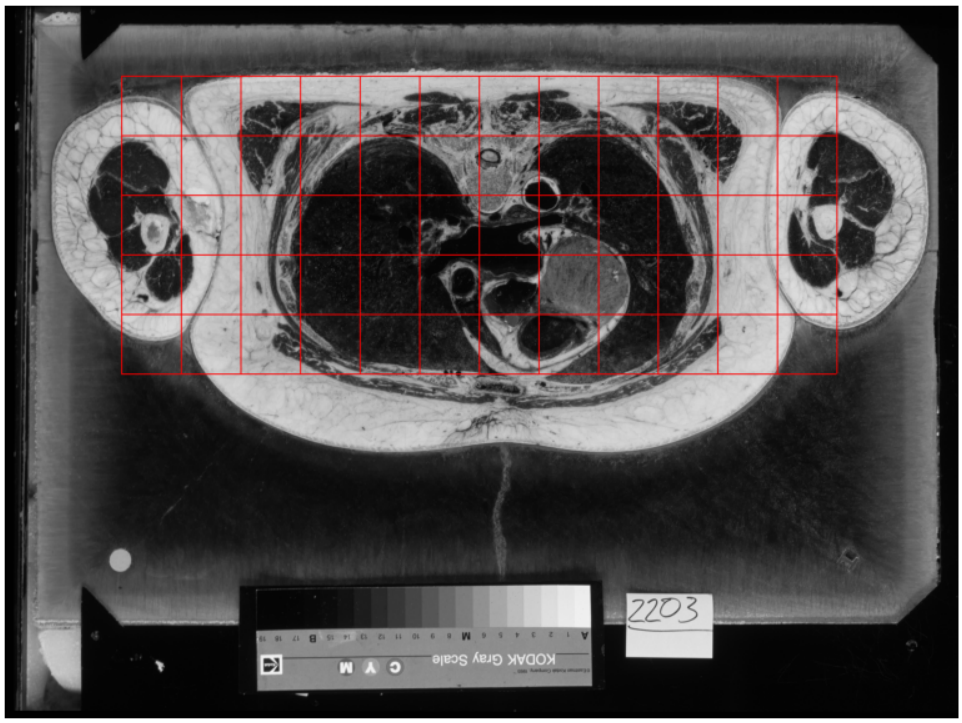


Figure 2: Cropping of Visible Human Dataset images

The
Hyperparams! 3,5 etc.

## 2.4 Pipeline

What they made available

Diagram, steps: Noising, Train-test-Splitting (talked more about above), step1train, preprocessing, reconstruction (including czxy-¿omx(or even paz) and stacking, unstacking, back to czxy), postprocessing, step2 training, post-processing.

What happens during training? Trainloader, data augmentation, regions of interest via intensity, patching/overlapping

Software best practices I/O, documentation, modularity, version control

Using CSD3, hardware, parallel -¿ serial

# 3 Results

Affine rescaling to minimize the MSE to the GT? Discuss the results they found

## 3.1 2D Data

Parameter estimation
- Generalizability
- Tables of results, metrics
- Images
- SIM check results!

## 3.2 3D Data

Axial resolution
- Tables of results, metrics
- Images

# 4 Discussion

Results/conclusions Further work What I learned How I could have improved

# 5 References


[1] E. Abbe, "Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung," *Archiv für Mikroskopische Anatomie*, vol. 9, pp. 413–468, 1873.

[2] G. Mats G.L. *et al.*, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophysical Journal*, vol. 94, no. 12, pp. 4957–4970, 2008. [Online]. Available: https://doi.org/10.1529/biophysj.107.120345

[3] X. Li *et al.*, "Three-dimensional structured illumination microscopy with enhanced axial resolution," *Nature Biotechnology*, vol. 41, pp. 1307–1319, 2023. [Online]. Available: https://doi.org/10.1038/s41587-022-01651-1

[4] C. Karras *et al.*, "Successful optimization of reconstruction parameters in structured illumination microscopy – a practical guide," *Optics Communications*, vol. 436, 2019.

[5] Y. Zhang *et al.*, "Image super-resolution using very deep residual channel attention networks," in *Computer Vision—ECCV*, 07 2018, pp. 294–310.

[6] J. Chen *et al.*, "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes," *Nature Methods*, vol. 18, pp. 678–687, 2021.

# A Statement on the use of auto-generation tools

# B High-Performance Computing Resources

This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).