Improving axial resolution in SIM using deep learning - Supplementary materials

Boland, Miguel A.; Cohen, Edward A. K.; Flaxman, Seth R.; Neil, Mark A. A.

Department of Mathematics
Imperial College
South Kensington Campus
180 Queen's Gate
London SW7 2RH
United Kingdom
1. Supplementary material A1 - Measuring FWHM

As detailed in the main paper, the FWHM is measured from the autocorrelation function using the methodology below. The example details the estimation of the axial (Z) resolution; lateral resolution can be calculated by summing over the Z axis in Equation 1.3, selecting the zero-padded row of matrix F in Equation 1.5 which contains the peak value for the matrix and decomposing the series into two Gaussian functions in Equation 1.6.

Let $I$ be a 3D microscopy $X \times Y \times Z$ image stack of dimensions $(N_X, N_Y, N_Z)$  

(1.1) Let $P = |FFT_3(I)|^2$ where FFT3 is a 3D fast-Fourier transform. (1.2) Summing $P$ across the $X$ and $Y$ dimensions gives vector $S$ with $k$th element

$$S_k = \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} P_{i,j,k}. \quad (1.3)$$

We then pad vector $S$ using the following rule

$$S_k = \begin{cases} S_k & \text{if } k < \frac{N_Z}{2} \\ 0 & \text{if } \frac{N_Z}{2} < k < \frac{N_Z}{2} + p \text{ where } p \text{ corresponds to the width of 0-padding.} \\ S_{k-p} & \text{if } k > \frac{N_Z}{2} + p. \end{cases} \quad (1.4)$$

We then inverse FFT this vector

$$F = FFT_{shift}(IFT(S)). \quad (1.5)$$

and decompose it into two Gaussian functions $g_{acf}(\alpha, \mu, \sigma)$ and $g_{ccf}(\alpha, \mu, \sigma)$, for the auto-correlation and cross-correlation function, respectively, via a least squares fit

$$F = g_{acf}(\alpha_{acf}, \mu_{acf}, \sigma_{acf}) + g_{ccf}(\alpha_{ccf}, \mu_{ccf}, \sigma_{ccf}). \quad (1.6)$$

Constraints are placed that $\mu_{acf} \approx \mu_{ccf}$ and $\sigma_{acf} < \sigma_{ccf}$. The FWHM is then measured on $g_{acf}(\alpha_{acf}, \mu_{acf}, \sigma_{acf})$. 
2. Supplementary material A2 - Y/Z cross sections of output images

![Reference HR](image1)
![SIM reconstruction](image2)
![RCAN](image3)

Figure 1: Y/Z cross-sections of a simulated chromatin structure

3. Supplementary material A3 - Effect of obscuring a frame from the RCAN's input

The charts below demonstrate the effect of pinning any individual SIM frame to 0 when inputting a chunk of 3 SIM reconstructions into the RCAN. This demonstrates the usage of all available input frames to reconstruct the output image, as the lowest MSE and best SSIM are only achieved when no frame is obscured.

4. Supplementary material A4 - Example modelling of autocorrelation function as sum of two gaussians

Figure 3 illustrates modelling of the inverse FFT of an image stack’s power spectrum, as described in Section A1. Gauss1 is modelled as the autocorrelation function, whilst Gauss2 models the cross-correlation function; the sum of both model the measured function well as demonstrated by the Gauss1+Gauss2 series.

5. Supplementary material A5 - Network architecture & optimisations

The RCAN architecture was optimised for 3D SIM reconstruction of our test datasets across a series of parameters. These are:
Figure 2: Evaluating the effect of masking one of 21 frames in an image chunk, taken from a simulated chromatin structure & processed by our model.

Figure 3: Decomposing an image stack’s inverse FFT of the power spectrum into the sum of two Gaussians

- Number of residual groups / blocks
- Kernel filter size
- Chunk size (number of concurrent SIM reconstructions)

As a multi-parameter grid-search optimisation would be computationally intractable, sequential optimisation was performed in the above order. Networks were trained for 50 epochs using the same dataset used in the final model’s training scheme (see Methodology section of main document).
(a) Network depth

Network depth was optimised across possible network configurations which would fit in the memory of the GPU available (Nvidia GTX 2080 TI). This ranges from networks of a single block/group up to networks of 12 groups of 3 blocks, or 3 groups of 12 blocks. Evaluation of these networks on the withheld dataset yielded the results seen in Figure 4, and default parameters were used for the kernel size (3x3) and chunk processing (3 SIM images).

Changes in the topology of skip connections in the network (i.e. change in the number of blocks and groups) has little effect on the structural similarity, MSE, or lateral and axial resolution of the network given a sufficient network depth; this is visible in the curved edge of the charts, where networks with approximately similar depth have near-indistinguishable performance. A configuration of 12 groups of 3 blocks was therefore selected due to its marginal performance improvement across all metrics, though other similarly complex configurations may be more suitable for other training data.

(b) Kernel filter size

Using the previously determined network architecture and a chunk size of 3, the same training scheme was applied with varying kernel filter sizes. Padding was adjusted to each kernel size so as to maintain the overall image size. As seen in Figure 5, kernel sizes larger than 5 result in a
complete failure of the reconstruction, demonstrated by large increases in SSIM and MSE. Of the remaining kernel sizes evaluated, a size of 3 yielded the best mean MSE, lateral resolution and axial resolution with an insignificantly different SSIM; this was therefore selected.

![Box plots showing SSIM, MSE, lateral resolution, and axial resolution for different kernel sizes.](image)

**Figure 5:** Effect on kernel filter size on network performance

(c) Chunk size - parallel SIM reconstructions

The concurrent processing of neighbouring SIM images was assumed to allow some axial inference between reconstructions, which would be reflected in improved performance metrics. An initial evaluation (see Figure 6) over 50 epochs demonstrated a degradation of performance as we increased the number of concurrently reconstructed images; it was also noted that the network complexity increased, thereby hampering the training speed of the network.

Repeating the experiment for 100 epochs (see Figure 8) demonstrated that a chunk size of 3 yielded an improvement of axial resolution (the main goal of this project) with non-significant degradation of SSIM, MSE and lateral resolution. We could further hypothesise that the degradation in performance would be minimised as the number of epochs is increased, and therefore selected a chunk size of 3 as the optimal chunk size for our dataset.
Figure 6: Effect of chunk size on network performance (50 epochs)
Figure 7: Effect of chunk size on network performance (100 epochs)
6. Supplementary material A6 - Training charts

Figure 8: An example training log, demonstrating that over-fitting is not occurring as the validation loss does not consistently increase.
7. Supplementary material A7 - Sub-resolution test cases

The following two cases demonstrate the reconstruction of objects whose axial separation is smaller than the theoretical axial resolution of SIM but above the resolution of our RCAN model. Raw SIM image stacks were generated using the same methodology as in our main experiment; these contain either 2 points axially separated by 900 nm, or 3 densely sampled planes (1 µm × 1 µm, 10 000 points per µm²) separated by 900 nm. The stacks were then processed using our trained RCAN model and SIM reconstruction, the images were normalised to the range [0,1] and pixel values below 0.001 were removed to mitigate noise in the output and remove negative (imaginary) values from the images.

![Sub-resolution test cases](image)

(a) 2 axially separated points  
(b) Reconstruction of an axial grating

Figure 9: Mean pixel values along Z axis for sub-resolution test cases

As we can see in Figure 9a and Figure 9b, the SIM reconstruction fails to resolve the data into separate objects, whilst the RCAN reconstruction identifies all peaks in both cases.

Repeating this analysis (as shown in Figure 10b over a range of grating separations demonstrates a significant improvement in contrast between sim and rcan reconstructions, though the contrast for lower resolution features could be improved in future work.
(a) Mean contrast between gratings at varying distances

(b) X/Z cross-sections of 3 gratings axially separated by 1500nm.

Figure 10