Label2label: Using deep learning and dual-labelling to retrieve cellular structures in fluorescence images

Lisa Sophie Kölln[1-3,*], Omar Salem[2,3], Jessica Valli[4], Carsten Gram Hansen[2,3,†], Gail McConnell[1,†]

[1] University of Strathclyde, Department of Physics, Glasgow, UK

[2] University of Edinburgh, Centre for Inflammation Research, Edinburgh, UK

[3] University of Edinburgh, Institute for Regeneration and Repair, Edinburgh, UK

[4] Edinburgh Super Resolution Imaging Consortium, Institute of Biological Chemistry, Biophysics and Bioengineering, Heriot-Watt University, Edinburgh, UK

[*] corresponding author: lisa.kolln@strath.ac.uk

[†] last author

Key words: image content filter, fluorescence microscopy, antibody-labelling, deep learning, convolutional neural networks, content-aware image restoration, noise2noise, cellular structures, focal adhesions, cytoskeleton, microtubules

Abstract

Spatial localisation of proteins dictates cellular function. Hence, visualisation of precise protein distribution is essential to obtain in-depth mechanistic insights into protein roles during cellular homeostasis, dynamic cellular processes, and dysfunction during disease. Labelling and staining of cells with protein specific antibodies is therefore a central and widely used technique in cell biology. However, unspecific binding, or cytoplasmic signals originating from the antibodies, make the distinction of the fluorescence signal from cellular structures challenging. Here we report a new image restoration method for images of cellular structures, using dual-labelling and deep learning, without requiring clean ground truth data. We name this method label2label (L2L). In L2L, a convolutional neural network (CNN) is trained with noisy fluorescence image pairs of two non-identical labels that target the same protein of interest. We show that a trained network acts as a content filter of label-specific artefacts and cytosolic content in images of the actin cytoskeleton, focal adhesions and microtubules, while the contrast of structural signal, which correlates in the images of two labels, is enhanced. We use an established CNN that was previously applied for content-aware image restoration, and show that the implementation of a multi-scale structural similarity loss function increases the performance of the network as content filter for images of cellular structures.

Introduction

In recent years, deep learning has increasingly been used for image processing in cell biology [1]. Specifically, convolutional neural networks (CNNs) with, or closely related to, a U-Net architecture [2] are employed for various machine learning tasks, from protein detection in transmission microscopy images [3,4], to image segmentation of single cells [5] or cellular structures such as the nucleus [6]. CNNs have also been used in medical applications, for example, for cell density estimation in corneal endothelial specular microscopy images [7] or tumour classification of urothelial cell carcinoma in pathology [8].

Fluorescence microscopy is a commonly used technique in cell biology to determine the spatial distribution and abundance of specific proteins in cells. It relies on the use of highly specific labels to visualise the cellular components of interest. Fluorescent labels can be chemical stains, antibody (AB) labelling or molecular labelling, where cells are genetically altered to express fluorescent proteins *in vitro* [9]. Here, we mainly focus on immunofluorescence labelling in fixed cells.

Recently, artificial neural networks were used for content-aware image restoration (CARE) of fluorescence images with low photon counts. By training a U-Net with image pairs that were generated at different laser powers, or different exposure times, respectively, it was shown that a trained CNN was able to successfully restore clean images of cell structures like the cell membrane and the nucleus [10]. Moreover, the CSBDeep Framework was used to enhance the image resolution in images that were detected below the Nyquist sampling frequency [10]. In another approach to improve cell image data, a generative adversarial network (GAN) was utilised to enhance the image resolution, using total internal reflection microscopy (TIRF) and stimulated emission depletion (STED) microscopy images as training benchmark [11].

Unsupervised deep learning methods have also emerged to restore noisy fluorescence images, for example, *noise2void* to reject camera shot noise [12], *DivNoising*, based on a variational autoencoder that is trained to predict a distribution of restored images based on a noise model [13], and *noise2noise* (N2N) [14]. For N2N, a neural network is trained with corrupted image pairs of the same sample. Here, due to the statistical nature of how loss is minimised by a neural network during training with a deviation-minimising loss function, uncorrelated image signal like Gaussian, Poisson or Bernoulli noise is rejected while correlating sample signal is recovered (see Loss Functions for Training a Convolutional Neural Network) [14]. N2N was used, for example, as pre-processing step for cryo-electron microscopy images of biological specimen [15] and in single-molecule localisation microscopy for denoising of low frame density images of LifeAct in *Drosophila* follicular epithelium [16].

While these image restoration tools increase image contrast by up-sampling or by correcting for image corruptions caused by the imaging system, all mentioned methods do not correct for inherent artefacts in the sample itself which, in immunofluorescence microscopy, can originate from the cell label or labelling protocol. Unspecific labelling by a stain or antibody binding, as well as internalisation or residue dyes after the sample preparation, can significantly limit the inherent image contrast of the cell component of interest [17]. The performance of antibodies or stains can vary for different reasons. For instance, specific epitopes, that are the binding site of a monomeric antibody, can be altered in the target protein by the fixation step, effectively changing the location or accessibility to the antibody [18]. Also, unspecific antibody binding can occur caused by attractive intermolecular interaction - like van der Waals or hydrogen bonding interaction -, or by binding to proteins with similar epitope, which overall can result in an underlying background signal in images of cells [19]. While super-resolution microscopy techniques can generate images with decreased out-of-focus fluorescence signal, these techniques are often limited by small fields of view, they may suffer from increased photo-bleaching, involve more laborious specimen preparation protocols and, most importantly, they do not prohibit the detection of signal in cell specimens that originates from unspecific binding [20].

Label2label

Here we propose a new application of deep learning for the image restoration of fluorescence images of cellular structures, with the aim to reduce diffuse, structure-less fluorescence signal in the cells caused by the labelling. We call this method *label2label* (L2L). For L2L, a CNN is trained with image pairs of cells that were dual-labelled for the same distinct cellular structure of interest, for example, the cytoskeleton, microtubules or focal adhesions. L2L utilises the varying performance of antibodies and stains in fluorescence microscopy. We hypothesized that a CNN trained with two images of a cellular structure that originate from two non-identical labels and therefore exhibit sample differences would act as a content filter - where fluorescence signals that systematically vary in the images are rejected, while correlating, cellular structures are restored. Consequently, L2L is similar to N2N. In both methods a network is trained without clean benchmark images, while in L2L differences between the training input and benchmark images are not only caused by dynamic image corruptions like noise, but also inherent sample (=label) differences between training input and output.

L2L is also different to restoration methods with a *noise2clean* approach, like CARE, where a CNN is trained with images of the same label that were taken under different imaging conditions [10]. While image pairs for *noise2clean* can also be generated with one microscope, for example, by varying the exposure time or image size, for imaging of fixed cell specimen, this approach is often not practical. For instance, to obtain training data for the restoration of confocal images of fixed cells, image pairs

with different exposure time, or frame-averaging respectively, would need to be taken which is time-consuming, prone to photo-bleaching and stage drift. Also, as in N2N, label-specific artefacts are present in both the training input and output, and therefore they cannot be corrected.

For L2L training, the images for the training are acquired simultaneously of dual-labelled cells, under quasi-identical imaging conditions (see Methods: Imaging). The images obtained using one fluorescent cell label are selected as training input, while the images from the other fluorescent label are used as benchmark, using images of the label that yields a higher contrast of the cellular structure for the latter. We show that after L2L training, a CNN does not only restore images with reduced noise, but also acts as a content filter for unwanted image signal. For the trainings, we selected the CSBDeep framework for the neural network training that was previously used for image restoration of noisy or under-sampled images of cells [10].

We studied L2L as an image restoration method of cellular structures on image pairs of cells that were labelled for the actin cytoskeleton, focal adhesions and the tubulin network. Actin is a conserved protein in eukaryotic cells that plays a major role in cellular functions like cell migration, cell motility or sustaining the cell shape [21,22]. Actins exist in two states: as globular (G-) actin in its monomeric form, or polymerised as filamentous (F-) actin [23]. There is evidence that different actin isomers vary in function and localisation; they can appear, for example, in stress fibres, circular bundles, cell-cell contacts or the cell cortex [24]. Focal adhesions (FAs) are cellular membrane-associated multi-protein component biomechanical structures. FAs are integral in the ability for most cells to sense and respond to the extracellular matrix and physical changes in the cellular microenvironment, FAs thereby function as central cellular contact points with the extracellular environment, that relay and transmit dynamic external information as internal cellular signalling events [25,26]. Microtubules are fundamental cytoskeletal polymeric structures in all eukaryotic cells that, amongst others, play a part in cell transport, cellular signalling via cilia and cell division [27]. The inhibition as well as the promotion of microtubule assembly has been shown to promote mitotic arrest which makes microtubules a prominent target in cancer therapy research [28].

Loss functions for training a convolutional neural network

We studied how the loss function impacts the restoration performance of the trained network with the purpose of retrieving structures with specific shape. In general, a CNN, which can be described as a function g_ϑ with all its model parameters ϑ, is trained to minimise the error between two images based on a loss function L:

$$\underset{\vartheta}{\operatorname{argmin}} \sum_{t=1}^{T} L(g_\vartheta(x_t), y_t)\,, \qquad [1]$$

with $g_\vartheta(x_t)$ as the predicted image for the input x_t, y_t as the benchmark and T as the number of input-benchmark image pairs that are used for the training [14].

The most commonly used deviation-minimising estimators are the least absolute deviation loss function L_1 (or *LAD*) and the least square deviation loss function L_2 (or *LSD*) [14,29]:

$$L_1(\hat{x}_t, y_t) = \frac{1}{N}\sum_{n=1}^{N}\left|\hat{y}_{t,n} - y_{t,n}\right|, \quad [2]$$

$$L_2(\hat{x}_t, y_t) = \frac{1}{N}\sum_{n=1}^{N}\left(\hat{y}_{t,n} - y_{t,n}\right)^2, \quad [3]$$

where $\hat{y}_t$ (=$g_\vartheta(x_t)$) is the predicted image and N is the number of pixels in the image.

Since L_2 is minimal if it equals the mean value of the observations, it was previously used for N2N training in cases where the image corruption resembles, for example, Gaussian noise that, on average, is zero [14]. L_1 is the loss function of the CSBDeep framework in default configuration (for non-probabilistic training). With both, L_1 and L_2, the network calculates the average loss on a pixel-to-pixel basis which limits the ability of the network to consider 2-dimensional correlating signal of structures in an image. As a result, predicted images from a CNN trained with L_1 or L_2 can be of low quality to a human observer [29].

A multi-scale structural similarity (MS-SSIM) index was proposed to compare structural signal between two images that takes the properties of the human visual system into account [30,31]. The M-scale SSIM for a pixel p is calculated as follows, after making assumptions as outlined in [31]:

$$MS - SSIM(p) = (l_M(p))^{\gamma_M} \cdot \prod_{j=1}^{M}(c_j(p))^{\gamma_j} \cdot (s_j(p))^{\gamma_j}, \quad [4]$$

with the exponent γ_j as the weight for the individual scale j ($\sum_{j=1}^{M}\gamma_j = 1$), and measures that compare the luminance (l_M), the contrast (c_j) and structure (s_j) of two images. These measures are functions of the local statistics: the sample mean $\mu_{\hat{y}_p/y_p}$, the sample variance $\sigma^2_{\hat{y}_p/y_p}$ and the sample correlation coefficient $\sigma^2_{\hat{y}_p y_p}$ between two image patches $\hat{y}_p$ and y_p that are the local neighbourhood of a centre pixel p [30]. They are described as:

$$l(\hat{x}_p, y_p) = \frac{2\mu_{\hat{y}_p}\mu_{y_p} + (K_1R)^2}{\mu^2_{\hat{y}_p} + \mu^2_{y_p} + (K_1R)^2},\ c(\hat{x}_p, y_p) = \frac{2\sigma_{\hat{y}_p}\sigma_{y_p} + (K_2R)^2}{\sigma^2_{\hat{y}_p} + \sigma^2_{y_p} + (K_2R)^2},\ s(\hat{x}_p, y_p) = \frac{2\sigma_{\hat{y}_p y_p} + (K_2R)^2}{2\sigma_{\hat{y}_p}\sigma_{y_p} + (K_2R)^2}. \quad [5\text{-}7]$$

Here, K_1 and K_2 are small constants that are generally set to 0.01 and 0.03 respectively, and R is the dynamic range of the image (for example, 255 in an 8-bit image). Most commonly, the local statistics of a pixel in two images is calculated using a Gaussian weighting function with a window size of 11 and a standard deviation of 1.5 (as in [30]). To calculate the MS-SSIM index, if M>1, a low-pass filter is applied to the image patches after each iteration, followed by down-sampling by a factor of 2. This approach makes the MS-SSIM index sensitive to differing viewing conditions, such as differing

perceived resolution from the point of observation. In [31], to calculate a 5S-SSIM index (M=5), the weights γ_j were experimentally determined to be (0.0448, 0.2856, 0.3001, 0.2363, 0.1333) for a representative image example.

The (multi-scale) SSIM exhibits values between (-1, 1) where 0 implies no structural similarity and -1/1 a negative/positive correlation between the image signals. Since a network aims to minimise loss during training, a multi-scale SSIM loss function $L_{MS\text{-}SSIM}$ follows [29]:

$$L_{MS-SSIM}(\hat{y}_t, y_t) = 1 - \frac{1}{N}\sum_{n=1}^{N} MS - SSIM(\hat{y}_{t,n}, y_{t,n}) \,. \qquad [8]$$

We hypothesized that a neural network will learn to filter out non-structural signal after L2L training with appropriate image data, and that the performance of the network would improve for this task when trained with a MS-SSIM loss function. We conducted trainings with the same image data for each cellular structure, using a L_1, L_{SSIM} (M=1), $L_{3S\text{-}SSIM}$ (M=3) or $L_{5S\text{-}SSIM}$ (M=5) loss function. The results of the trainings were compared to the denoising method N2N for which the same pre-processing steps and network settings were used, but two images of the same label were used as training input and benchmark instead. While N2N, so far, has only been used in cell biology to denoise images and not to filter out image content, we wondered if the implementation of a MS-SSIM loss function would increase the contrast of cellular structures in restored images after N2N training, making it a competitor to L2L. Our aim was to compare two image restoration methods that both can be implemented to increase contrast in confocal image data of fixed cells.

Results

Label2label to filter out de-polymerised β-actin in images of HeLa cells

To study the impact of L2L training on images of actin, image pairs of HeLa cells (N=68) were generated that were fixed 6h after plating, and dual-labelled with the monoclonal anti-β-actin antibody AC-15 and a phalloidin-ATTO-565 stain (see Methods: Imaging). **Figure 1A** shows an example immunofluorescence (IF) image pair of AC-15 and the phalloidin stain in a HeLa cell. Images shown in this work were excluded from the training. While the phalloidin stain labels almost exclusively the actin filaments, the monoclonal anti-β-actin antibody also binds to cytosolic protein which leads to an increased underlying fluorescence signal in the cell body, as well as high intensity punctate regions where larger quantities of de-polymerised actin are accumulated. The resulting difference in image contrast between both labels is quantifiable by calculating the mean Michelson and RMS contrast for the images of each label. To calculate these, we first applied a Gaussian filter with sigma=2 to the images and normalised the images to their 1st and 99.9th percentile. Then we derived the sample intensities based on the assumption that the 10% brightest pixels in the image represented the sample

[32]. For the images obtained from the labelling by the antibody AC-15 and the phalloidin stain, the mean Michelson contrast values of 0.44±0.08 and 0.96±0.04, as well as mean RMS contrast values of 0.15±0.03 and 0.19±0.01 were calculated for the generated images (N=68). Consequently, images of AC-15 were used as training input and images of the phalloidin stain as benchmark for L2L training to train the CNN to filter out non-structural signal originating from de-polymerised actin.

To evaluate the performance of the network after L2L training, the average peak signal-to-noise-ratio (PSNR) and normalised root-mean square error (NRMSE), as well as the average 5S-SSIM, 3S-SSIM and SSIM indices were calculated between the predicted images for AC-15 and the corresponding images of phalloidin, dependent on the loss function that was used for the training, and using the validation image patches that were excluded from the training data (see Methods: Evaluation).

	PSNR (dB)	**NRMSE**	**5S-SSIM**	**3S-SSIM**	**SSIM**
Input	16.77	0.149	0.444	0.339	0.156
L_1	19.88	0.105	0.641	0.538	0.322
$L_{5S\text{-}SSIM}$	**20.23**	**0.101**	**0.679**	0.573	0.348
$L_{3S\text{-}SSIM}$	20.21	0.102	0.678	**0.575**	**0.351**
L_{SSIM}	20.05	0.104	0.667	0.569	0.348

Table 1 - Evaluation of L2L for images of actin.
Evaluation after training a CNN for L2L with image pairs of HeLa cells (N=68) that were dual-labelled with an anti-actin antibody and a phalloidin stain, dependent on the loss function that was used for the training. The metrics were calculated between the cell image of AC-15, or predicted image for AC-15, respectively, and the benchmark image (phalloidin), using 13,927 image patch pairs that were excluded from the training. The best value for each metric is shown in bold.

All calculated metrics show that the correlation between the predicted images for AC-15 after L2L training and the benchmark (phalloidin) increases compared to the input (see **Table 1**). Here, the network performance increased after replacing a standard L_1 loss function with a (multi-scale) SSIM loss function. In comparison, the $L_{5S\text{-}SSIM}$ loss function narrowly yielded the best PSNR and NRMSE values.

Figure 1B shows the restored image for AC-15 in a HeLa cell (see **Figure 1A**), after L2L training with a $L_{3S\text{-}SSIM}$ loss function. Cytoplasmic signal is reduced throughout the cell body by the CNN, while the relative signal of filamentous actin labelled with AC-15 is increased. In **Figure 1C**, the results are shown for two selected regions of interest for both labels, including an average that was derived from imaging 20 frames of the respective field of view, the predicted image after N2N training with a $L_{3S\text{-}SSIM}$ loss function for each label, and the RMS map between the original and predicted image with L2L. For L2L training, only images of AC-15 were used as training input, but the trained network also decreased non-filamentous content when restoring images of phalloidin. For N2N training, for each label, the

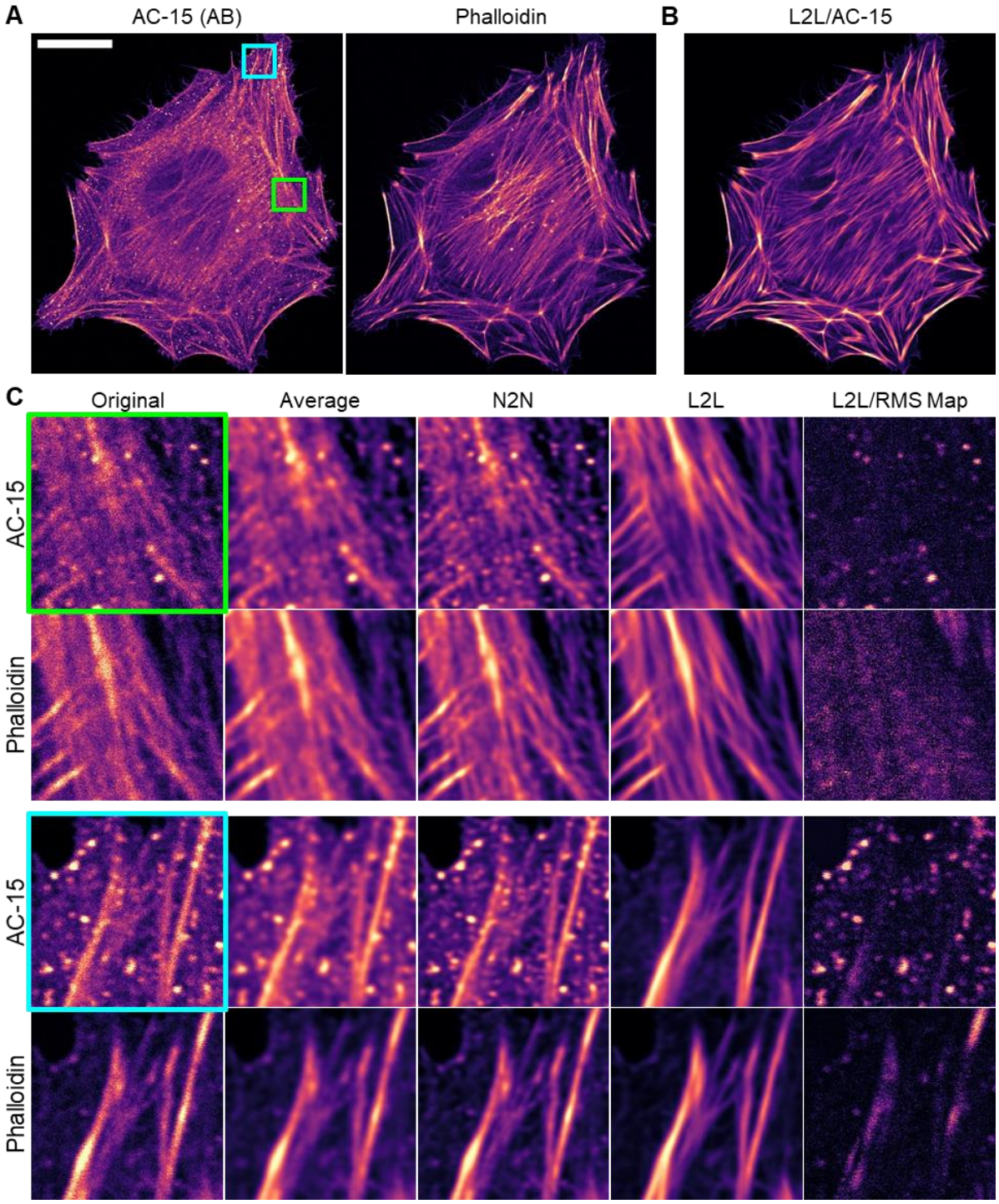


Figure 1 - Qualitative label2label (L2L) and noise2noise (N2N) results for images of actin.
(**A**) *Image pair of a HeLa cell that was dual-labelled with the anti-β-actin antibody AC-15 and a phalloidin stain which was excluded from the CNN trainings (scale bar = 20 µm).* (**B**) *Reconstructed image of AC-15 by a CNN after L2L training with images of AC-15 and phalloidin as training input and benchmark, respectively, using a* $L_{3S\text{-}SSIM}$ *loss function.* (**C**) *Original and processed images of both labels for two ROIs (from left-to-right): raw image, 20-frame average, restored images after N2N and L2L training with a* $L_{3S\text{-}SSIM}$ *loss function, and the RMS map between the original and the predicted image of the network after L2L training (image dimensions: 8.1 µm x 8.1 µm).*

network was trained with two images of the respective label. While the actin cytoskeleton appeared sharper in restored images after N2N training with $L_{3S\text{-}SSIM}$ compared to a 20-frame average image, predictably, a network trained with two images of the same label did not distinguish between

filamentous structures and protein accumulations in images of HeLa cells that were labelled with the antibody (AC-15). In **Supplementary Information (SI) Figure 1**, the qualitative results for both N2N and L2L are shown for AC-15 after training with a L_1, $L_{5S\text{-}SSIM}$, $L_{3S\text{-}SSIM}$ and L_{SSIM} loss function. The corresponding images of phalloidin which were used as benchmark during L2L training are shown in panel C, while, for the fair assessment of N2N results, high frame average images of AC-15 are shown in panel A that exhibit reduced noise. In comparison, for both methods, using a loss function that only assesses the least absolute deviation (L_1) led to the most conservative prediction results, where, in L2L, de-polymerised actin was filtered out by the network, but actin filaments appeared blurry. On the other hand, cellular structures in restored images after training the CNN with a single-scale SSIM loss function (L_{SSIM}) appeared sharp with a higher likelihood of appearing wrong predictions. These "hallucination effects" were predominately noticeable in the results for N2N, but also for L2L. With N2N, using a SSIM loss function, filaments that appeared with low contrast in the raw images of AC-15 were restored with intensity fluctuations along the structure, which was not observed for L2L. N2N and L2L both decrease image noise.

Label2label for cytosolic signal reduction in cell images of PAXILLIN

To study L2L as content filter method for cytoplasmic signal in images of focal adhesion markers, the CNN was trained with image pairs of fixed MeT5A (N=47), HeLa (N=17) and U2OS (N=13) cells that were dual-labelled with two monoclonal antibodies for the focal adhesion molecule PAXILLIN (PXN) (5H11 raised in mouse and Y113 raised in rabbit). As expected, the raw IF images of both antibodies show correlating focal adhesion structures in the cells, but also a diffuse signal throughout the cell body that differs in relative intensity to the focal adhesion signal between the two labels (see **Figure 2A**). For all cell lines we observed the same trend; the images of the antibody Y113 exhibited a higher focal adhesion-to-cytosolic signal ratio compared to 5H11. Since all cell lines were also imaged under the same imaging conditions, we used all indistinguishable as training data. For the cell images of 5H11 and Y113 (N=77), mean Michelson contrast values of 0.57±0.07 and 0.99±0.01, and mean RMS contrast values of 0.15±0.01 and 0.18±0.02 were calculated, respectively [32].

To determine if the cytoplasmic signal in the images of both PXN antibodies originates from clone-dependent binding to protein in the cytosol or unspecific binding, shRNA mediated *PXN* knockdown MeT5A cells were generated. The knockdown was confirmed via a qPCR analysis that showed a knockdown efficiency of >70% on mRNA level and also substantial lower protein levels confirmed by Western blotting (see **SI Figure 2A-B**). For both antibodies, the immunofluorescence signal of PXN was overall reduced in images of fixed shPXN cells in the FAs and in the cytoplasm (see **SI Figure 2C**). This

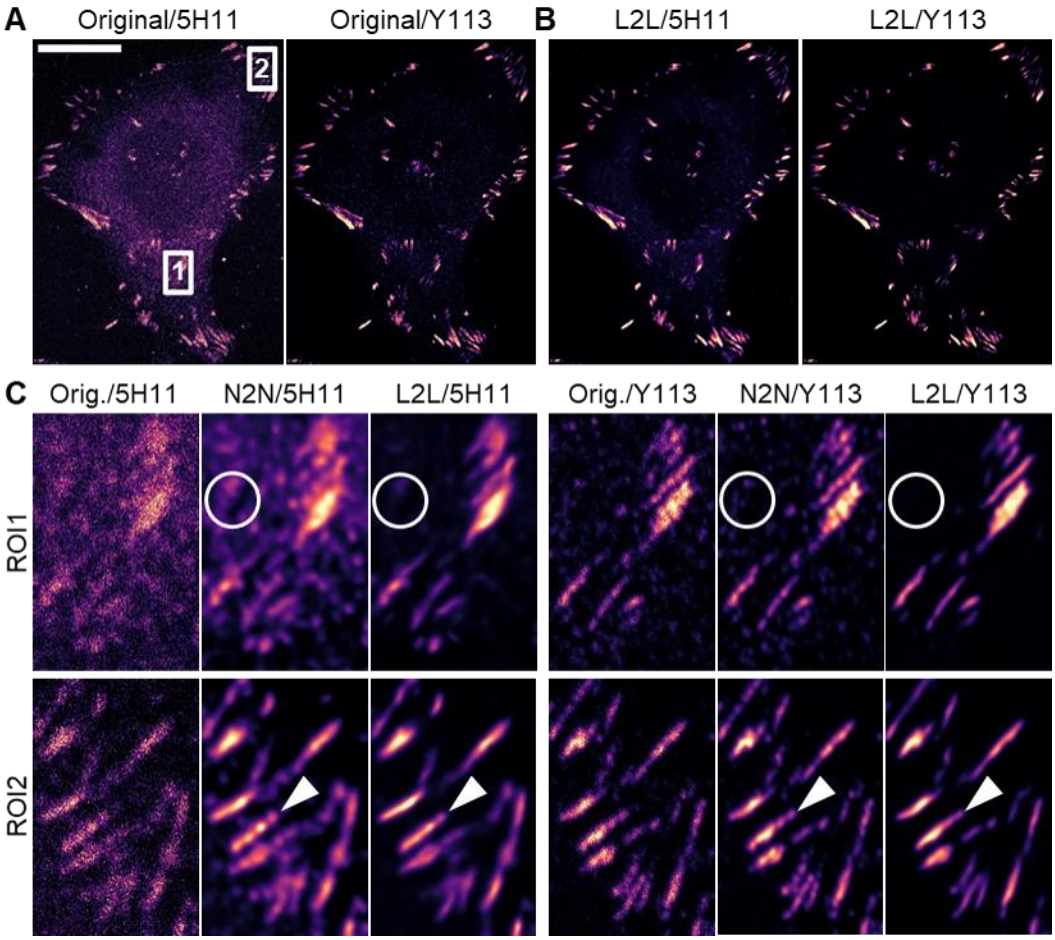


Figure 2 - Qualitative label2label (L2L) and noise2noise (N2N) results for images of PXN.
(**A**) *Image pair of a HeLa cell that was dual-labelled with the anti-PXN antibodies 5H11 and Y113 which was excluded from the CNN trainings (scale bar = 20 µm).* (**B**) *Reconstructed images of both ABs by a CNN after L2L training with images of 5H11 and Y113 as training input and benchmark, respectively, using a L_1 loss function.* (**C**) *Comparison of predicted images for (left) 5H11 and (right) Y113 in a HeLa cell after N2N and L2L training with a L_1 loss function for two ROIs. White circles highlight areas in the predicted images where cytosolic content is successfully filtered out by a CNN after L2L training; white arrows point to an example focal adhesion that was in-painted along the structure in the L2L results (image dimensions: 9.6 µm x 6.2 µm).*

confirms that the cytoplasmic signal that is observed in cell images of both clones is largely not caused by unspecific binding. Instead, the relatively high cytosolic signal in cell images of the antibody 5H11 when compared to results for Y113 indicates a differing accessibility of the respective binding sites for both clones of protein that is cytosolic or bound to focal adhesions, and that cytosolic PXN likely functions as a readily available replenishable buffer for FA localised PXN.

For L2L training, images of cells that were labelled with the antibody 5H11 were used as training input, and images of Y113 were used as benchmark. In **Table 2**, the average PSNR and NRMSE values, as well as the average 5S-SSIM, 3S-SSIM and SSIM indices are listed after L2L training with different training loss functions, calculated for image patches that were excluded from the training. In line with previous results, the evaluation indicates that the highest network performance is achieved with a $L_{5S\text{-}SSIM}$ loss function.

	PSNR (dB)	NRMSE	5S-SSIM	3S-SSIM	SSIM
Input	18.73	0.118	0.455	0.337	0.152
L_1	19.78	0.107	0.465	0.339	0.135
$L_{5S\text{-}SSIM}$	**21.27**	**0.091**	**0.605**	**0.497**	0.329
$L_{3S\text{-}SSIM}$	21.2	0.092	0.6	0.496	**0.33**
L_{SSIM}	21.13	0.093	0.592	0.489	0.328

Table 2 - Evaluation of L2L for images of PXN.
Evaluation after training a CNN for L2L with image pairs of MeT5A (N=47), HeLa (N=17) and U2OS (N=13) cells that were dual-labelled with the anti-PXN antibodies 5H11 and Y113, dependent on the loss function that was used for the training. The metrics were calculated between the cell image of 5H11, or predicted image for 5H11, respectively, and the benchmark image (Y113), using 15,770 image patch pairs that were excluded from the training. The best value for each metric is shown in bold.

After L2L training with a L_1 loss function, cytosolic PXN signal is clearly reduced in restored images for both anti-PXN antibodies compared to the raw images, while FA structures are recovered (see **Figure 2A-B**). Again, the network could be used for cell images of both antibodies, although only cell images of 5H11 were used for the training input. For images of both antibodies, a network that was trained with L2L filters out cytosolic content, which is not the case after N2N training (see white circles in **Figure 2C**). Contrary to the results of the evaluation (see **Table 2**), subjectively, the L_1 loss function was better suited to filter cytosolic signal in this dataset. For N2N, using a MS-SSIM loss function trained the network to artificially accentuate intensity fluctuations in cytosolic content in the restored images that increased when decreasing the number of iterations in the MS-SSIM loss function (see **SI Figure 3**). While similar artefacts were much less observable for L2L, a network trained with a MS-SSIM loss function increasingly (with decreasing M) recovered cytosolic protein signal as small, low-intensity structures. This is likely because signal of protein accumulations in the cell cytoplasm were present in images of both the training input and benchmark, albeit with differing intensity. The selective recovery of high-intensity cytosolic content and rejection of low-intensity cytosolic content led to hallucination effects. Using a L_1 loss function, the trained network restored images more conservatively, where remaining cytosolic content in the restored images appeared blurry against the recovered sharp FA structures. In addition, a structural in-painting effect was observed after L2L training, where sample-intrinsic intensity fluctuations along focal adhesion structures in the detected images were in-painted by the trained network (see white arrows in **Figure 2C**). As expected, this effect was not observed after N2N training with two sequential images for the respective label.

Label2label to restore microtubule structures

Lastly, L2L was studied as restoration method for fluorescence images of microtubules. Confocal image pairs of fixed MeT5A cells (N=51) were generated that were dual-labelled with two monoclonal antibodies against α-tubulin (DM1A raised in mouse and YOL1/34 raised in rat). The image contrast differed for both antibodies: for the antibody DM1A, a mean Michelson contrast of 0.45±0.14 and a

291 mean RMS contrast of 0.13±0.03 were calculated; for YOL1/34 the contrast values were 0.9±0.05 and
292 0.17±0.02, respectively [32]. In the cell images of YOL1/34 we observed intensity fluctuations along
293 the microtubules that might be caused by selective binding to specific binding sites of polymerised
294 tubulin or in-homogeneous binding, which was not observable for the clone DM1A (see **Figure 3A-B**
295 and **SI Figure 4C**). The microtubules appeared overall sharper in the cell images of YOL1/34, compared
296 to DM1A, with a lower, observable "haze" in the cytoplasm that was likely caused by unspecific
297 binding, binding to cytosolic tubulin and/or out-of-focus signal. Notably, YOL1/34 was conjugated to
298 FITC that emits light around 518 nm, while DM1A was labelled with the secondary antibody Alexa 633
299 and, as a consequence, image differences of both microtubule markers might also originate from a
300 slightly different image resolution for both fluorophores.

	PSNR (dB)	NRMSE	5S-SSIM	3S-SSIM	SSIM
Input	17.2	0.14	0.54	0.359	0.135
L_1	16.88	0.146	0.415	0.271	0.103
$L_{5S\text{-}SSIM}$	**18.37**	**0.123**	**0.66**	**0.475**	**0.203**
$L_{3S\text{-}SSIM}$	18.23	0.125	0.652	0.469	0.2
L_{SSIM}	18.03	0.128	0.648	0.468	0.199

301 **Table 3 - Evaluation of L2L for images of tubulin.**
302 *Evaluation after training a CNN for L2L with image pairs of MeT5A cells (N=51) that were dual-labelled with the*
303 *anti-α-tubulin antibodies DM1A and YOL1/34, dependent on the loss function that was used for the training. The*
304 *metrics were calculated between the cell image of DM1A, or predicted image for DM1A, respectively, and the*
305 *benchmark image (YOL1/34), using 10,445 image patch pairs that were excluded from the training. The best value*
306 *for each metric is shown in bold.*

307 For L2L training, cell images of DM1A and YOL1/34 were used as training input and benchmark,
308 respectively. The loss function-dependent evaluation of the performance of a trained CNN indicated
309 that the correlation between reconstructed and benchmark image was highest after training with a
310 $L_{5S\text{-}SSIM}$ loss function (see **Table 3**). The results showed a decrease in correlation between the restored
311 image of DM1A and the benchmark (YOL1/34) after L2L training with a L_1 loss function.

312 The predicted images of the network for each antibody image after L2L training of a CNN with a $L_{3S\text{-}$
313 $SSIM}$ loss function are shown in **Figure 3B**. The trained network decreased image noise upon restoration
314 of DM1A, and microtubules appeared sharp against an overall reduced background in the cell body
315 (see **Figure 3B** (*left*, *bottom*)). However, the qualitative results for cell images of the benchmark
316 (YOL1/34) indicated an introduction of artefacts: the microtubule diameter varied along the structure,
317 along with visible intensity fluctuations (see **Figure 3B** (*right*, *bottom*)).

318 Microtubule structures are insufficiently resolved with standard confocal microscopy, especially if
319 densely packed [33]. To better assess the restoration performance of the network after N2N and L2L
320 training, we took confocal-STED image pairs of the antibody DM1A in MeT5A cells, using the same
321 sample, but a different imaging setup (see Methods: Imaging). With STED microscopy, a resolution

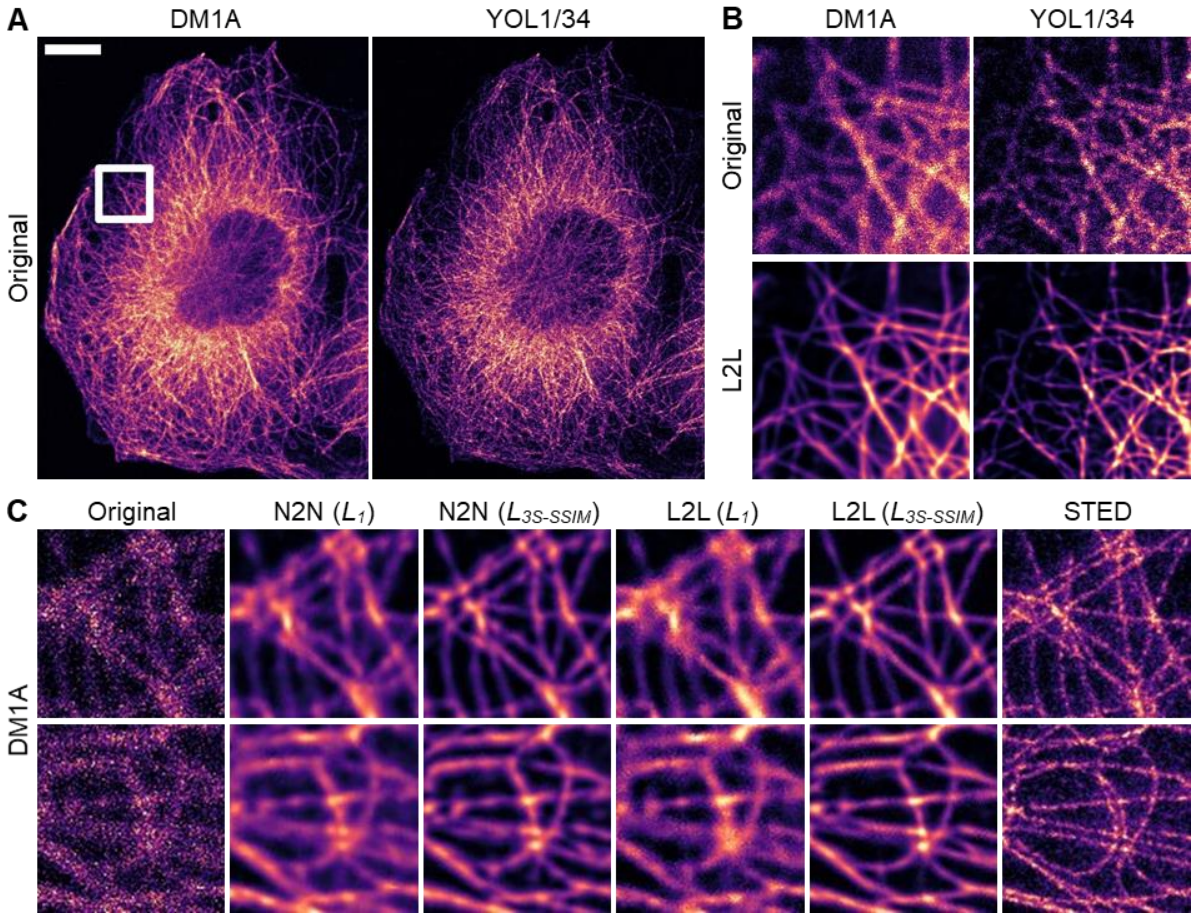


Figure 3 - Qualitative label2label (L2L) and noise2noise (N2N) results for images of tubulin.
(A) *Image pair of a MeT5A cell that was dual-labelled with the anti-tubulin antibodies DM1A and YOL1/34 which was excluded from the CNN trainings (scale bar = 10 µm).* (**B**) *Reconstructed ROI of both ABs by a CNN after L2L with images of DM1A and YOL1/34 as training input and benchmark, respectively, after using a $L_{3S\text{-}SSIM}$ loss function (image dimensions: 9 µm x 9 µm).* (**C**) *Reconstructed cell images of DM1A after N2N and L2L training (as described for* (**B**)*) with a L_1 or $L_{3S\text{-}SSIM}$ loss function, compared to the STED image (image dimensions: 2.3 x 2.3 µm).*

enhancement by a factor of circa 3 was achieved compared to standard confocal microscopy. **Figure 3C** shows the confocal images of the antibody DM1A in MeT5A cells, as well as the restorations of the confocal image after N2N and L2L training with a L_1 or $L_{3S\text{-}SSIM}$ loss function, compared to the STED image. For comparison, the prediction results of true representative input images are shown in the **SI Figure 4**. We found that the restoration quality slightly decreased for confocal images taken with another imaging setup, but the loss function-dependent trend for both methods was the same: using $L_{3S\text{-}SSIM}$ instead of L_1 for the trainings increased the contrast and sharpness of microtubules in the images for both methods (N2N and L2L). For all training settings the restoration success was dependent on the tubule density in the image. The closer microtubules were packed in the cell, the less likely was the successful recovery of separate structures by the CNN.

Testing the robustness of label2label

To evaluate the robustness of L2L on the actin dataset (N=68), the PXN dataset (N=77) and the tubulin dataset (N=51) and study the impact of the total number of image pairs that are used to generate the

training data for L2L training on the network performance, we conducted repeated cross-validations. For the actin and PXN dataset, we conducted 8-fold cross-validations for a varying total number (N_{tot}) of image pairs (N_{tot}= 8, 16, 32, 64), for the tubulin dataset, we conducted 10-fold cross-validations (N_{tot}=10, 30, 50). We increased the number of repetitions for small N_{tot} (see **Table 5** in Methods: Training the Convolutional Neural Network). The difference to this approach and the default validation in the CSBDeep framework, is, that in the latter the validation is conducted via a train/test split of the training data which is generated from all raw image pairs. This way, image patches generated from an image pair can appear in both, the test and the training data, making it impossible to assess if the network performance is dependent on which raw image pairs are used for the training.

The mean relative change of the NRMSE and 5-scale SSIM index between the input and the restored image (both versus the reference (=benchmark)) are shown in **Figure 4**. Each data point represents the mean value of a cross-validation dependent on N_{tot}. The results for all datasets show a similar trend: even when using a small N_{tot}, the structural similarity to the training benchmark of the restored image is higher compared to the original image. However, restoration performance seems to be dependent on the particular images that are used for the training indicated by the wide distribution of *ΔNRMSE* and *Δ5S-SSIM* values for small N_{tot}. This can either mean that certain images are better suited to train the network, or easier to predict for the network. As expected, on average, the highest performance was achieved using a higher number of image pairs to generate the training data. While the trend of *Δ5S-SSIM* for all datasets indicates that above N_{tot}=30, or 32 respectively, the network performance is relatively consistent, the *ΔNRMSE* improves continuously with increasing N_{tot}.

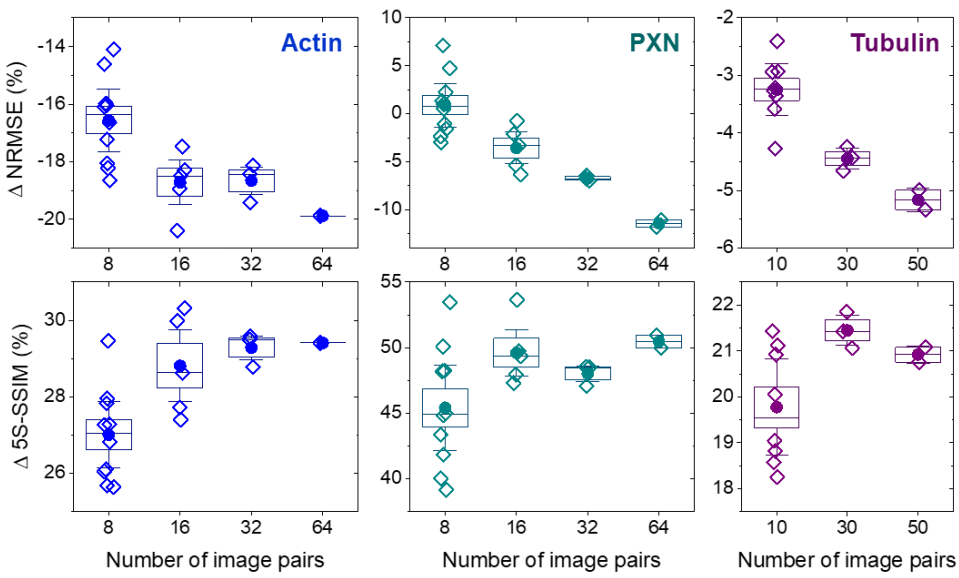


Figure 4 - Repeated cross-validation for L2L training.
The mean relative change between the initial (input versus benchmark) and final (restored versus benchmark) NRMSE and 5-scale SSIM index after L2L training with image pairs of cells that were dual-labelled for actin, PXN and tubulin, dependent on the number of raw image pairs that were used for each cross validation. Each data point is the mean value of an 8-fold (actin, PXN) or 10-fold (tubulin) cross-validation that was repeated for small image pair numbers.

355 ## Discussion

356 We showed that a CNN can be trained as content filter for images of cellular structures, using images
357 of two non-identical cell labels with appropriate uncorrelated signal for the training. Using L2L, label-
358 specific artefacts and undesired cytosolic content can be reduced, which is different to restoration
359 methods that use images of the same label. Here, for images of PXN and actin, we showed that a
360 trained network can be re-used to further improve images that were used as training benchmark. For
361 instance, while cytosolic content was still present in restored images of the anti-PXN antibody 5H11
362 (input), it was almost eradicated completely in restored images of Y113 (benchmark).

363 The network performance as content filter was dependent on the level of correlation between the
364 images of the two labels and the training loss functions. Using a single-scale SSIM loss function for the
365 training increased the likelihood of hallucination effects in restored images after L2L training,
366 especially for the PXN dataset where cytosolic content in the cell body was not entirely without
367 structure and was present in both, training input and output. At the same time, all MS-SSIM loss
368 functions were better suited than a L_1 loss function to train a network to restore sharp cellular
369 structures. We found that by increasing the number of iterations in the MS-SSIM loss function, the
370 results increasingly converged to results obtained with a L_1 loss function. For images of actin and
371 tubulin we determined that using a $L_{3S\text{-}SSIM}$ loss function led qualitatively to the best restored image
372 results. Here, the difference of the to-filter-out signal either sufficiently deviated between the image
373 pairs (actin) or was inhomogeneous within the cell body for both labels (tubulin), allowing the network
374 to clearly distinguish to the cellular structure. For the PXN dataset, a L_1 loss function was more suited
375 to recover FAs only, likely due to the correlating cytosolic signal in both images, albeit at different
376 relative intensity to the FA signal.

377 Thus far, deep-learning based image restoration methods have relied on image pairs of the same label
378 with differing resolution or photon counts as training data. For such training data, our results suggest
379 that powerful loss functions like the MS-SSIM loss function are prone to overestimate structural signal,
380 resulting in increased hallucination effects (see, for example, N2N results in **SI Figure 3B**). In this work,
381 we showed that a MS-SSIM loss function can be successfully implemented to train a CNN for image
382 restoration in cell biology, if the to-filter-out signal differs sufficiently in the training input and
383 benchmark images.

384 We found that metrics like the PSNR and NRMSE were insufficient to forecast which training loss
385 function led to the best qualitative result - especially given that the reference image was not clean,
386 showing the cellular structure only. For instance, the calculated metrics after L2L training with

387 different MS-SSIM loss functions barely deviated for all dataset examples while, qualitatively, the
388 number of iterations set in the MS-SSIM loss function visibly impacted the restored images.

389 The calculated metrics also suggested that a high-scale SSIM index is better suited to detect a
390 correlation in fluorescence images. While the mean SSIM indices that were calculated between input
391 and benchmark images indicated scarcely any correlation, with values around 0.14-0.16 for all
392 datasets (see **Tables 1-3**), the initial correlation between the training data was much higher according
393 to the calculated 5S-SSIM indices (around 0.44-0.54). This indicates that, in general, a high-scale SSIM
394 index is more suited to recognise structural similarity in confocal fluorescence images of cellular
395 structures than the single-scale SSIM, likely because the images are the convolved signal of a sample
396 volume rather than the strict 2D depiction of the sample at a specific section. This observation fits to
397 the theory of the MS-SSIM [31].

398 The results for the repeated cross-validations for a different number of raw image pairs that were
399 used for the training data generation indicated that, even at a small number of image pairs, a network
400 learns to restore an image with increased structural similarity to the respective benchmark during L2L
401 training. However, the evaluation of the cross-validations also showed that the network performance
402 varied for small image pair numbers. In practice, using a high number of raw image pairs (>50) for the
403 training data generation is clearly recommended to generate a more robust neural network.

404 Although clean benchmark data are not required for L2L training, in images of structures that are not
405 sufficiently resolved by the imaging technique, either a posteriori knowledge about the structure or
406 clean reference data are required to assess the qualitative performance of a trained CNN. This was
407 especially noticeable after L2L training with cell images of tubulin, where the fine microtubule network
408 was not fully resolved with confocal microscopy (see **Figure 3C**). This led to erroneous predictions in
409 image areas with a dense microtubule network.

410 In all examples, a CNN that was trained for L2L did not only act as a content filter but also in-painted
411 the respective cellular structures in restored images, reducing intensity fluctuations in the structure
412 itself. This was especially observable for the PXN dataset, where the fluorescence signal of PXN bound
413 to the FAs in the raw images sometimes appeared inhomogeneous along a FA (see **Figure 2C**). This
414 underlines that restored images of cellular structures after L2L training do not represent a true protein
415 distribution, since intensities in the structures are altered (see RMS maps in **Figure 1C**). Consequently,
416 L2L results cannot be used for a quantification of protein distribution in cells. Instead, L2L could be
417 used as image pre-processing step to extract the binary information about the location of a specific
418 structure in a cell image, for example, of the cell membrane or the nucleus. Here, the systematic
419 recovery of specific structure by the network might make L2L superior to conventional image filters.

420 For L2L, training data can be generated in parallel with the same imaging setup using two detectors
421 simultaneously, which makes the images independent of stage drift and sample dynamics, and is time-
422 efficient. L2L could be useful for super-resolution microscopy imaging methods that suffer from photo-
423 bleaching [20], or as a post-processing step in live cell imaging, where high-performance labels for
424 cells with little background signal or internalisation are rare. For that, image pairs for L2L training could
425 be generated after the acquisition of the live cell images, by fixing the cells and labelling with a higher
426 performing antibody that target the same structure. In multiplexed cell imaging, L2L could be applied
427 in cases where antibodies are not selected based on performance but compatibility issues between
428 the species in which they are raised.

429 Conclusion

430 We have presented a new restoration method for cellular structures that utilises the varying
431 performance of labels in immunofluorescence microscopy. We show that by training a CNN that was
432 developed for content-aware image restoration, with images of two non-identical labels that show
433 the same cellular structure but exhibit systematic sample differences, the network learns to selectively
434 restore predominantly the correlating signal in the images. L2L is similar to other unsupervised image
435 restoration methods like *noise2void* and N2N which rely on the convention of the network to under-
436 estimate inherently unpredictable signal. However, with L2L, not only image noise but also specimen
437 artefacts can be corrected in the images when selecting appropriate training data. The ability to
438 correct images for unspecific binding, inhomogeneous labelling or binding to cytosolic protein makes
439 L2L, to our knowledge, unique in comparison to other deep learning-based image restoration methods
440 that are currently used in cell biology.

441 # Methods

442 ## Cell Culture

443 For imaging, the human mesothelial cell line MeT5A (ATCC CRL-9444), the adenocarcinoma cervical
444 cancer cell line HeLa, and the human osteosarcoma cell line U2OS were used. HeLa cells were a gift
445 from Margaret Cunningham and U2OS cells were a gift from Kathryn McIntosh (both Strathclyde
446 Institute of Pharmacy and Biomedical Sciences, Glasgow, UK). MeT5A cells were grown in RPMI-1640
447 medium (Corning), supplemented with 10% (v/v) fetal bovine serum (FBS) (Labtech), 100 µg/ml
448 penicillin-streptomycin (Gibco), 1 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco) and 2 mM
449 HEPES buffer solution (Gibco). HeLa and U2OS cells were grown in DMEM+GlutaMAX medium (Gibco)
450 supplemented with 10% (v/v) FBS and 100 µg/ml penicillin-streptomycin. Human embryonic kidney
451 cells HEK293T were grown in DMEM supplemented with 100 µg/ml penicillin-streptomycin and 2mM
452 L-glutamine. All cells were kept at 37°C/5% CO_2 in a humidified atmosphere.

453 ## Generation and Verification of Knock Down *PAXILLIN* MeT5A Cells

454 Knockdown of MeT5A cells was obtained by a shRNA mediated knockdown using pKLO.1-vectors
455 coding for shRNAs mediated targeting *PAXILLIN* (TRC N0000123137) or control. HEK293T cells were
456 used for virus generation, and the virus was harvested, filtered and added to polybrene treated MeT5A
457 cells. The retroviral transduction of MeT5A cells was followed by puromycin selection. Stable
458 knockdown cell lines were verified by quantitative PCR (qPCR) (mRNA abundance) and Western blots
459 (protein abundance). For western blotting, cell lysates were prepared in reducing lysis buffer, boiled
460 for ten minutes and separated on conventional homemade SDS gels and developed onto Medical Blue
461 Sensitive X-Ray Film (Scientific Laboratory Supplies). The PAXILLIN specific antibody Y113 was used
462 (rabbit; ab32084, Abcam). Ponceau staining of the PVDF membrane was carried out to establish equal
463 loading. For quantitative PCR (RT-qPCR), mRNA was extracted from cells using the RNeasy plus mini
464 kit (QIAGEN). Complementary DNA (cDNA) was generated using the High-Capacity cDNA Reverse
465 Transcriptase kit (Applied Biosystems). qPCR was performed on 1 ng of cDNA using the Brilliant III
466 Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies) and the Applied Biosystems
467 QuantStudio 5 Real-Time PCR Systems. Expression of *PAXILLIN* was analysed and normalized to
468 Hypoxanthine-guanine phosphoribosyltransferase (*HPRT1*) levels. qPCR data originate from five
469 independent biological replicates, plotted in Prism (GraphPad) and analysed using Mann-Whitney.

470 ## Sample Preparation for Immunofluorescence Microscopy

471 If not stated otherwise, cells were plated onto #1.5 coverslips a day prior to fixation, then fixed with
472 4% paraformaldehyde for 15 min at 37°C, followed by a permeabilisation step with 2.5% FBS and 0.3%

473 TX 100 in PBS for 30min at room temperature [34]. HeLa cells that were labelled for actin were fixed
474 6h after plating. For immunostaining the following antibodies or stains were used: for PAXILLIN
475 labelling in MeT5A, U2OS and HeLa cells the monoclonal antibodies Y113 (1:250; rabbit; ab32084,
476 Abcam) and 5H11 (1:500; mouse; MA5-13356, Life Technologies); for α-tubulin labelling in MeT5A
477 cells the two monoclonal antibodies DM1A (1:250; mouse; T6199, Sigma Aldrich) and YOL1/34
478 conjugated to Alexa Fluor 488 (1:500; rat; ab195883, Abcam); to visualise the actin cytoskeleton in
479 MeT5A cells the phalloidin-ATTO-565 stain (1:100; 94072, Sigma Aldrich) and the monoclonal anti-β-
480 actin antibody AC-15 conjugated to FITC (1:250; mouse; ab6277, Abcam). Secondary antibody labelling
481 was conducted with anti-rabbit Alexa Fluor 488 (1:200; donkey; A-21206, Life Technologies) and anti-
482 mouse Alexa Fluor 633 (1:200; goat; A-11001, Life Technologies). The samples were mounted in
483 ProLong Glass Antifade Mountant (Life Technologies) onto microscope slides.

484 Imaging

485 The image pairs were taken with a commercial confocal laser scanning microscope (Leica TCS SP8,
486 Leica Microsystems), using a HC PL Apo 60x/1.4 N.A. CS2 objective and with the image size set to 4096
487 pixels x 4096 pixels with no digital zoom. Both antibodies and stains, respectively, in the individual
488 sample were imaged simultaneously using the two in-built photomultiplier detectors, each equipped
489 with a prism-based tunable spectral filter. Each field of view was imaged twice to accommodate later
490 neural network training for N2N. The samples dual-labelled for PAXILLIN and tubulin were excited with
491 a 488nm and 638nm laser line and the two spectral detectors were set to detect light between 495-
492 620nm and 645-750nm. The samples dual-labelled for actin were excited with a 488nm and 552 nm
493 laser line and the two spectral detectors were set to detect light between 495-540 and 560-700 nm.

494 Samples of fixed MeT5A cells dual-labelled for tubulin were further imaged with a Leica SP8 TCS 3X
495 STED microscope with a Leica HC PL APO 100x/1.40 Oil STED WHITE objective (Leica 15506378).
496 Samples were excited using a Supercontinuum White Light Laser at 488nm and 633nm for confocal
497 and at 633nm for STED imaging. For confocal imaging, two Leica PMT detectors were set to detect
498 light between 498-600nm and 643-743nm. For STED imaging, a Leica HyD hybrid detector was set to
499 detect light between 643-743nm. STED depletion was performed with a Leica 775 nm depletion laser
500 set to 50% with time gating from 0.3-8 ns. Pairs of confocal and STED images were acquired with a 15
501 nm pixel size.

502 Image Drift Correction

503 Images for N2N training were taken sequentially and were therefore affected by stage drift, as well as
504 high-frame average images. For the drift correction, a Gaussian filter with sigma=2 was applied to all

images. Then 100 image pairs of size 60 pixels x 60 pixels were generated for each image in a temporal image stack, using the first detected image as benchmark. For that, the *create_patches* function from the CSBDeep framework was used to create random image patch pairs in non-background areas of the image. For each created image pair, the SSIM index (filter size=11, sigma=1.5) was calculated by shuffling the uncorrected image patch relative to the benchmark patch by 5 pixels into all directions, effectively cutting both into 50 px x 50 px patches. The shuffling vector that, on average, yielded the highest structural similarity between both images was selected as optimal position for the sequential image.

Data Augmentation

Prior to the training data generation, the images were augmented to create a higher number of image pairs. To achieve this, the images were interpolated, using the zooming factors 0.5, 0.75, 1.25 and 1.5, and were randomly rotated by 0°, 90°, 180° and 270°.

Training the Convolutional Neural Network

For N2N and L2L training, the CSBDeep framework was used, which is a CNN with U-Net architecture [2] that was developed for content-aware image restoration (CARE) [10]. The CSBDeep framework (version 0.6.0) was downloaded from github and used with its default settings unless stated otherwise (https://github.com/CSBDeep/CSBDeep).

In the CSBDeep framework, training patches are generated randomly from the raw images after a percentile normalisation. The images were normalised to their 1st and 99.9^{th} percentile. The patch size was set to 128 pixels x 128 pixels. Half of the training patches were generated from the raw images (1024 patches per image), the other half from augmented images (see previous section). Validation data was generated via a 90/10 train/test split, and the training and validation loss were monitored to rule out overfitting.

Four different loss functions were used for the trainings: the least absolute deviation loss function L_1, and a multi-scale SSIM loss function in single-scale (L_{SSIM}), with 3 scales ($L_{3S\text{-}SSIM}$) and 5 scales ($L_{5S\text{-}SSIM}$). If not stated otherwise, to calculate the (multi-scale) SSIM indices, the settings from the original publication were used (see Introduction: Loss functions for training a convolutional neural network). For $L_{3S\text{-}SSIM}$, the weights were set to (0.2096, 0.4659, 0.3245). For $L_{5S\text{-}SSIM}$, the filter size was set to 7.

The settings that were applied for the L2L trainings are shown in **Table 4**. The same settings were used to train the network for N2N, but the training data was generated from two sequential images of the respective antibody, and the epoch number was cut in half for the PAXILLIN dataset.

Dataset	Actin	PAXILLIN	Tubulin
Input	AC-15	5H11	DM1A
Benchmark	Phalloidin	Y113	YOL1/34
No. of raw image pairs (4096px x 4096px)	68+272†	77+308†	51+204†
No. of validation/ training patches (128px x 128px)	13,927/ 125,337	15,770/ 141,926	10,445/ 94,003
Learning rate	2E-4	2E-4	2E-4
Batch size/ steps per epoch/ epoch number	32/100/400	32/100/500	32/100/500

† augmented images

Table 4 - Training settings.

Overview over the training data and settings in the CSBDeep framework for L2L training.

Repeated 8- or 10-fold cross validations were conducted on the datasets using the settings as outlined in **Table 5**. The selection of the raw image pairs from the total dataset for each 8- or 10-fold cross validation and the fold allocation were conducted randomly in Python. The image pairs for each training were generated as described above, including data augmentation, and disabling the train/test split in the CSBDeep Framework. The evaluation was conducted with the images in the fold that were excluded from the respective training.

Dataset	**Actin (N=68) / PAXILLIN (N=77)** *8-fold cross validations*				**Tubulin (N=51)** *10-fold cross validations*		
Number of repetitions	10	5	3	2	8	3	2
Total number of image pairs	8	16	32	64	10	30	50
Image patch size	128 px x 128 px				128 px x 128 px		
Loss function	$L_{3S\text{-}SSIM}$ / L_1				$L_{3S\text{-}SSIM}$		
Learning rate/ batch size/ steps per epoch	5E-4/32/150				5E-4/32/150		
Epoch number	20	40	60	120	20	60	100

Table 5 - Settings for repeated cross validations.

Overview over the settings in the CSBDeep framework that were used for repeated 8- or 10-fold cross validations, training the network for L2L.

Evaluation

Prior to the evaluation, the image pairs were normalised and rescaled to minimise the mean squared error (MSE) to account for the differing pixel intensity distribution between the input and the baseline image (as described in [10]).

The average peak-signal-to-noise-ratio (PSNR), normalised root mean squared error (NRMSE), SSIM index, 3-scale SSIM index and 5-scale SSIM index (see Introduction: Loss Functions for Training a Convolutional Neural Network) were calculated to compare image similarity. The PSNR and NRMSE are calculated using:

$$NRMSE = \frac{\sqrt{MSE}}{R}, \quad [9]$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{R^2}{\sqrt{MSE}}\right), \quad [10]$$

with *R* as the dynamic range of the image and the *MSE* between two images *X* and *Y* that follows:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(X_n - Y_n)^2, \quad [11]$$

where *N* is the number of pixels in the image [10].

References

[1] C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nat. Methods*, vol. 16, no. 12, pp. 1215–1225, Dec. 2019, doi: 10.1038/s41592-019-0458-z.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Med. Image Comput. Comput. Interv. – MICCAI 2015. MICCAI 2015. Lect. Notes Comput. Sci.*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[3] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Methods*, vol. 15, no. 11, pp. 917–920, Nov. 2018, doi: 10.1038/s41592-018-0111-2.

[4] E. M. Christiansen *et al.*, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," *Cell*, vol. 173, no. 3, pp. 792–803, Apr. 2018, doi: 10.1016/J.CELL.2018.03.040.

[5] T. Falk *et al.*, "U-Net: deep learning for cell counting, detection, and morphometry," *Nat. Methods*, vol. 16, no. 1, pp. 67–70, Jan. 2019, doi: 10.1038/s41592-018-0261-2.

[6] J. C. Caicedo *et al.*, "Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images," *Cytom. Part A*, vol. 95, no. 9, pp. 952–965, Sep. 2019, doi: 10.1002/cyto.a.23863.

[7] M. C. Daniel *et al.*, "Automated segmentation of the corneal endothelium in a large set of 'real-world' specular microscopy images using the U-Net architecture," *Sci. Rep.*, vol. 9, p. 4752, Dec. 2019, doi: 10.1038/s41598-019-41034-2.

[8] I. Jansen *et al.*, "Automated Detection and Grading of Non-Muscle-Invasive Urothelial Cell Carcinoma of the Bladder.," *Am. J. Pathol.*, vol. 190, no. 7, pp. 1483–1490, Jul. 2020, doi: 10.1016/j.ajpath.2020.03.013.

[9] T. Suzuki, T. Matsuzaki, H. Hagiwara, T. Aoki, and K. Takata, "Recent advances in fluorescent labeling techniques for fluorescence microscopy," *Acta Histochemica et Cytochemica*, vol. 40, no. 5. Japan Society of Histochemistry and Cytochemistry, pp. 131–139, 2007, doi: 10.1267/ahc.07023.

[10] M. Weigert *et al.*, "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods*, vol. 15, no. 12, pp. 1090–1097, Dec. 2018, doi: 10.1038/s41592-018-0216-7.

[11] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods*, vol. 16, no. 1, pp. 103–110, 2019, doi: 10.1038/s41592-018-0239-0.

590 [12] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2Void - Learning Denoising from Single Noisy Images,”
591 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2129–2137, 2019, doi:
592 10.1109/CVPR.2019.00223.

593 [13] M. Prakash, A. Krull, and F. Jug, “DivNoising: Diversity Denoising with Fully Convolutional Variational
594 Autoencoders,” *arXiv Prepr.*, 2020, Accessed: Jul. 24, 2020. [Online]. Available:
595 http://arxiv.org/abs/2006.06072.

596 [14] J. Lehtinen *et al.*, “Noise2Noise: Learning Image Restoration without Clean Data,” *Proc. 35th Int.*
597 *Conf. Mach. Learn.*, vol. 80, pp. 2965–2974, Jul. 2018.

598 [15] D. Tegunov and P. Cramer, “Real-time cryo-electron microscopy data preprocessing with Warp,”
599 *Nat. Methods*, vol. 16, no. 11, pp. 1146–1152, 2019, doi: 10.1038/s41592-019-0580-y.

600 [16] J. Lefebvre *et al.*, “Single-Molecule Localization Microscopy Reconstruction Using Noise2Noise for
601 Super-Resolution Imaging of Actin Filaments,” *2020 IEEE 17th Int. Symp. Biomed. Imaging*, pp.
602 1596–1599, 2020, doi: 10.1109/ISBI45749.2020.9098713.

603 [17] C. B. Saper, “A guide to the perplexed on the specificity of antibodies,” *J. Histochem. Cytochem.*,
604 vol. 57, no. 1, pp. 1–5, 2009, doi: 10.1369/jhc.2008.952770.

605 [18] D. M. Miller and D. C. Shakes, “Immunofluorescence Microscopy,” *Methods Cell Biol.*, vol. 48, pp.
606 365–394, Jan. 1995, doi: 10.1016/S0091-679X(08)61396-5.

607 [19] C. Wagner *et al.*, “Dynamic force spectroscopy on the binding of monoclonal antibodies and tau
608 peptides,” *Soft Matter*, vol. 7, no. 9, pp. 4370–4378, May 2011, doi: 10.1039/c0sm01414a.

609 [20] S. W. Hell *et al.*, “The 2015 super-resolution microscopy roadmap,” *J. Phys. D. Appl. Phys.*, vol. 48,
610 no. 443001, Oct. 2015, doi: 10.1088/0022-3727/48/44/443001.

611 [21] R. Dominguez and K. C. Holmes, “Actin Structure and Function,” *Annu. Rev. Biophys.*, vol. 40, no. 1,
612 pp. 169–186, Jun. 2011, doi: 10.1146/annurev-biophys-042910-155359.

613 [22] P. Vedula *et al.*, “Diverse functions of homologous actin isoforms are defined by their nucleotide,
614 rather than their amino acid sequence,” *eLife*, vol. 6, Dec. 2017, doi: 10.7554/eLife.31661.

615 [23] C. Suarez and D. R. Kovar, “Internetwork competition for monomers governs actin cytoskeleton
616 organization,” *Nature Reviews Molecular Cell Biology*, vol. 17, no. 12. 2016, doi:
617 10.1038/nrm.2016.106.

618 [24] V. Dugina, I. Zwaenepoel, G. Gabbiani, S. Clément, and C. Chaponnier, “Beta and gamma-
619 cytoplasmic actins display distinct distribution and functional diversity.,” *J. Cell Sci.*, vol. 122, pp.
620 2980–2988, Aug. 2009, doi: 10.1242/jcs.041970.

621 [25] F. Martino, A. R. Perestrelo, V. Vinarský, S. Pagliari, and G. Forte, "Cellular mechanotransduction:
622 From tension to function," *Frontiers in Physiology*, vol. 9. Frontiers Media S.A., p. 824, Jul. 05, 2018,
623 doi: 10.3389/fphys.2018.00824.

624 [26] V. Rausch and C. G. Hansen, "The Hippo Pathway, YAP/TAZ, and the Plasma Membrane," *Trends in*
625 *Cell Biology*, vol. 30, no. 1. Elsevier Ltd, pp. 32–48, Jan. 01, 2020, doi: 10.1016/j.tcb.2019.10.005.

626 [27] E. Nogales, "Structural insights into microtubule function," *Annual Review of Biochemistry*, vol. 69.
627 pp. 277–302, Nov. 2000, doi: 10.1146/annurev.biochem.69.1.277.

628 [28] F. Pellegrini and D. R. Budman, "Review: Tubulin function, action of antitubulin drugs, and new drug
629 development," *Cancer Investigation*, vol. 23, no. 3. pp. 264–273, 2005, doi: 10.1081/CNV-
630 200055970.

631 [29] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration with Neural
632 Networks," *IEEE Trans. Comput. IMAGING*, vol. 3, no. 1, 2017.

633 [30] Z. Wang, A. C. Bovik, H. Rahim Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error
634 Visibility to Structural Similarity," *IEEE Trans. IMAGE Process.*, vol. 13, no. 4, 2004, doi:
635 10.1109/TIP.2003.819861.

636 [31] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality
637 assessment," *Conf. Rec. Asilomar Conf. Signals, Syst. Comput.*, vol. 2, pp. 1398–1402, 2003, doi:
638 10.1109/acssc.2003.1292216.

639 [32] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 2032–2040, 1990, doi:
640 10.1364/josaa.7.002032.

641 [33] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, "Multicolor super-resolution imaging with
642 photo-switchable fluorescent probes," *Science*, vol. 317, no. 5845, pp. 1749–1753, Sep. 2007, doi:
643 10.1126/science.1146598.

644 [34] V. Rausch and C. G. Hansen, "Immunofluorescence Study of Endogenous YAP in Mammalian Cells,"
645 Humana Press, New York, NY, 2019, pp. 97–106.

646

Contributions

L.S.K. designed the method, prepared the samples for imaging, generated the imaging data and wrote/implemented Python scripts for pre-processing, trainings and evaluations. C.G.H. and O.S. developed the *PXN* knockdown cells. O.S. validated the knockdown. J.V. and L.S.K. took the confocal-STED image pairs. L.S.K., C.G.H. and G.M. wrote the manuscript, with input from all authors.

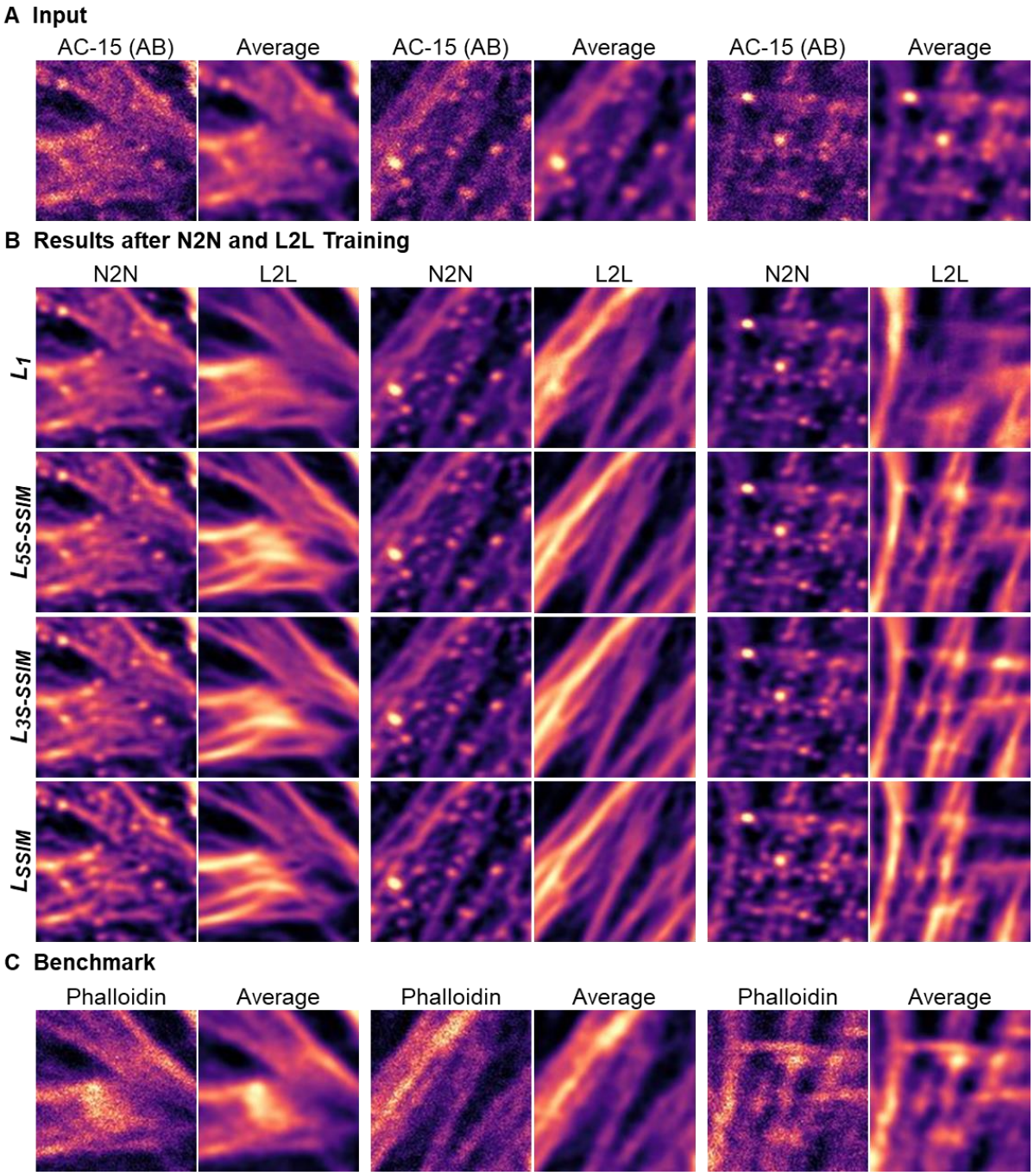
Competing Interests

The authors have no competing interests to declare.

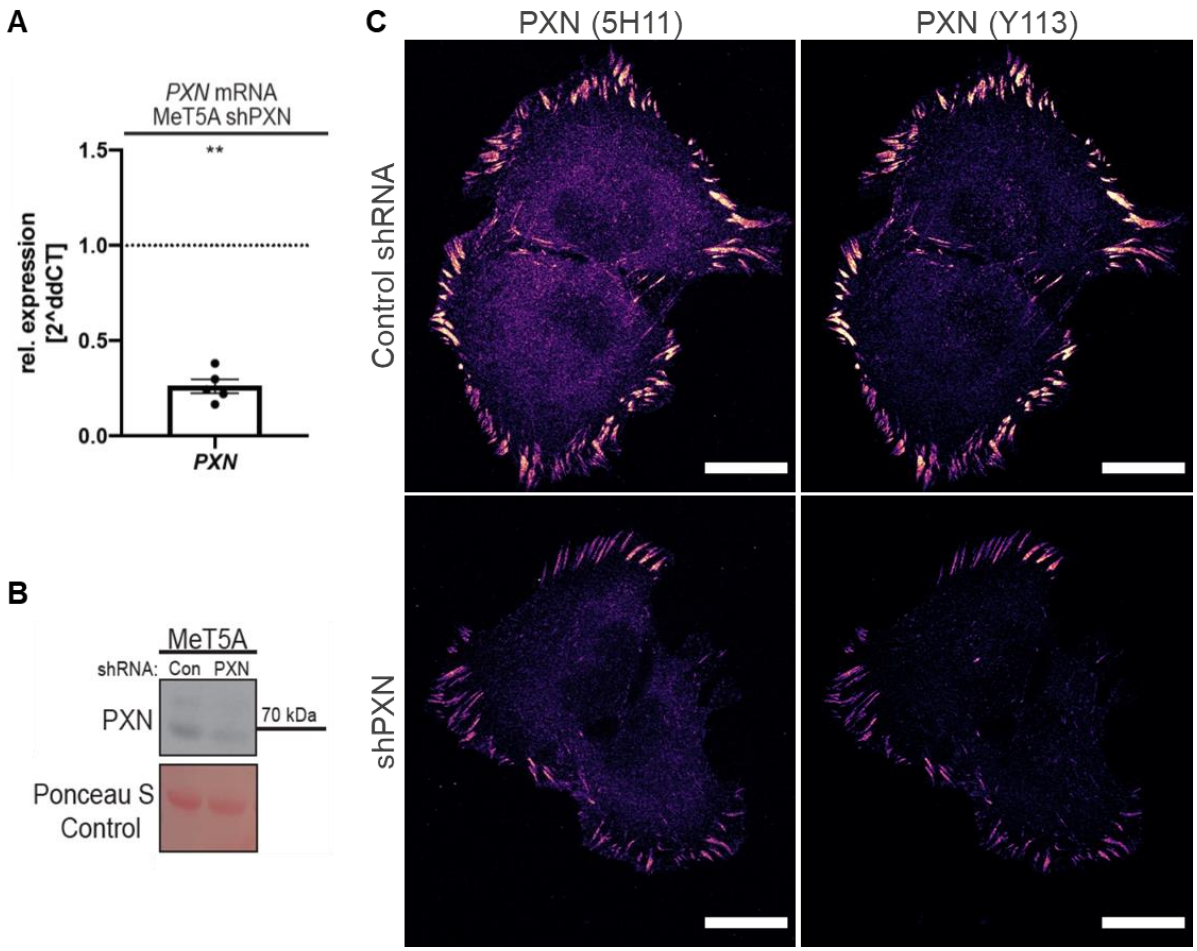
Acknowledgements

L.S.K. was supported by the Medical Research Council and Engineering and Physical Sciences Research Council Centre for Doctoral Training in Optical Medical Imaging (Optima), grant number EP/L016559/1. J.V. was supported by the Wellcome Trust, grant number 208345/Z/17/Z. STED imaging was performed at the Edinburgh Super Resolution Imaging Consortium (ESRIC), which is supported by the MRC and the Wellcome Trust. Work on-going in the Gram Hansen lab is supported by a University of Edinburgh Chancellor's Fellowship and start-up fund, as well as by the Worldwide Cancer Research and the June Hancock Mesothelioma Research Fund (JHMRF). G.M. was supported by the grants MR/K015583/1 and BB/P02565X/1. The authors thank Martin Weigert (Swiss Federal Institute of Technology Lausanne, Switzerland) for his helpful feedback.

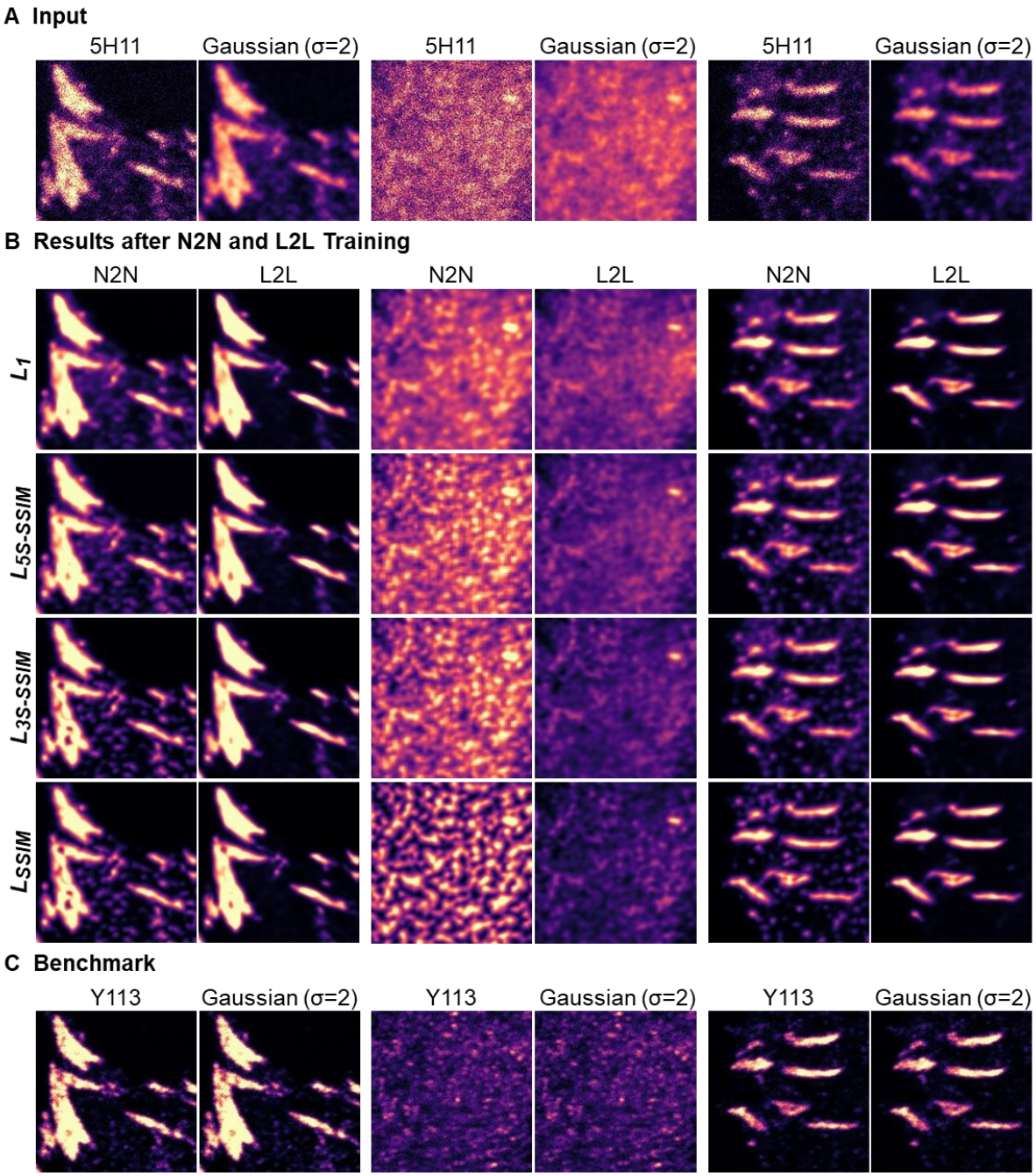
Supplementary Information



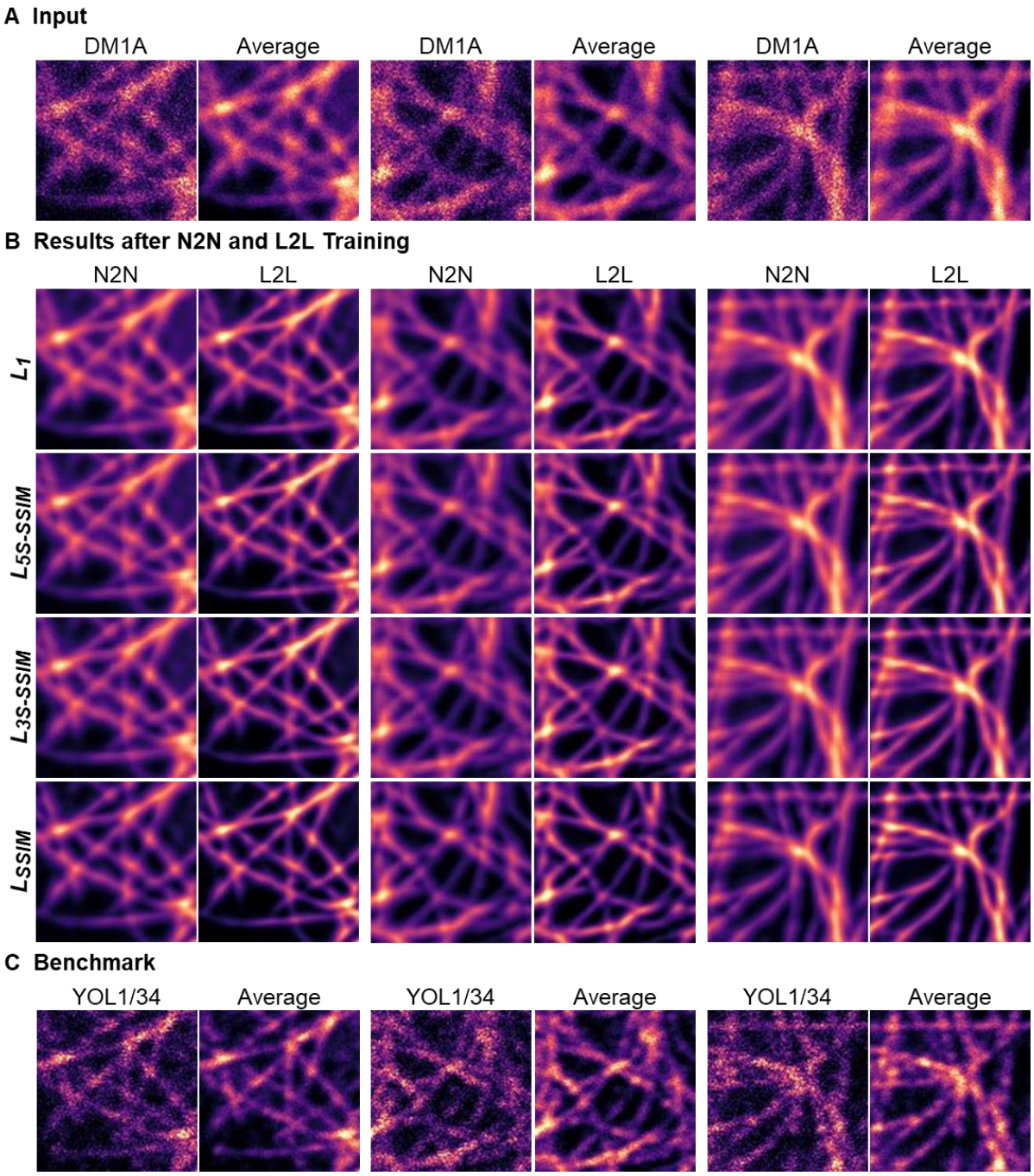
SI Figure 1 – Loss function-dependent label2label (L2L) and noise2noise (N2N) results for images of actin. Three representative examples of image pairs of HeLa cells that were dual-labelled with the anti-β-actin antibody (AB) AC-15 and a phalloidin-ATTO-565 stain. (**A**) Images of AC-15 in a HeLa cell and corresponding 20-frame average images. (**B**) Restored images of AC-15 by a CNN after N2N and L2L training, dependent on the training loss function. For N2N training, two sequential images of AC-15 in HeLa cells were used as training input and benchmark. For L2L training, benchmark images were replaced by (**C**) cell images of a phalloidin stain. The shown images were excluded from the trainings and contrast adjusted to the 2nd and 99.8th percentile (image dimensions: 4.5 µm x 4.5 µm).



SI Figure 2 - *PAXILLIN* (*PXN*) knockdown in MeT5A cells. PXN is a focal adhesion molecule. (**A**) Results of a RT-qPCR analysis, including the standard error of the mean: the relative expression of the *PXN* knockdown cells is 26.14±0.04%. (**B**) Western analysis of cell lysates of MeT5A cells visualising the PXN protein knockdown, with Ponceau stain ensuring equal loading. The monoclonal anti-PXN antibody Y113 was used. The PXN expression is reduced by around a half in the knockdown cells compared to the control. (**C**) Immunofluorescence images of MeT5A cells that were dual-labelled with the anti-PXN antibodies (*left*) 5H11 and (*right*) Y113 after (*top*) mock-infection and (*bottom*) shRNA mediated *PXN*-knockdown. The low signal of cytosolic protein in the knockdown cells for both antibodies indicate that the difference in cytosolic signal between the two, as observed in the control cells, originates from clone-dependent antibody binding to cytosolic PXN rather than unspecific binding (scale bar = 20 µm).



SI Figure 3 – Loss function-dependent label2label (L2L) and noise2noise (N2N) results for images of focal adhesions. Three representative examples of image pairs of MeT5A cells that were dual-labelled with the anti-PAXILLIN (PXN) antibodies 5H11 and Y113. (**A**) Images of 5H11 in MeT5A cells and corresponding Gaussian filter images (σ=2). (**B**) Restored images of 5H11 in a MeT5A cell by a CNN after N2N and L2L training, dependent on the training loss function. For N2N training, two sequential images of 5H11 in MeT5A cells were used as training input and benchmark. For L2L training, benchmark images were replaced by (**C**) cell images of the anti-PXN antibody Y113. The shown images were excluded from the trainings. Images of 5H11 and Y113 were contrast adjusted to the 2nd and 99.8th percentile of the respective original image (image dimensions: 9.1 µm x 9.1 µm).



SI Figure 4 – Loss function-dependent label2label (L2L) and noise2noise (N2N) results for images of tubulin. Three representative examples of image pairs of MeT5A cells that were dual-labelled with the anti-α-tubulin antibodies DM1A and YOL1/34. (**A**) Images of DM1A in MeT5A cells and corresponding 20-frame average images. (**B**) Restored images of DM1A in MeT5A cells by a CNN after N2N and L2L training, dependent on the training loss function. For N2N training, two sequential images of the DM1A signal in MeT5A cells were used as training input and benchmark. For L2L training, benchmark images were replaced by (**C**) cell images of the anti-α-tubulin antibody YOL1/34. The shown images were excluded from the trainings and contrast adjusted to the 2nd and 99.8^{th} percentile (image dimensions: 4.5 µm x 4.5 µm).