Research Article | Vol. 10, No. 3 | 1 Mar 2019 | BIOMEDICAL OPTICS EXPRESS 1339
Biomedical Optics EXPRESS

Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy

Navid Borhani,[1,*] Andrew J. Bower,[2,3] Stephen A. Boppart,[2,3,4,5] and Demetri Psaltis[1]

[1] *Optics Laboratory, School of Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland*
[2] *Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[3] *Department of Electrical and Computer Engineering, Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[4] *Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[5] *Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[*] *navid.borhani@epfl.ch*

Abstract: Deep neural networks have been used to map multi-modal, multi-photon microscopy measurements of a label-free tissue sample to its corresponding histologically stained brightfield microscope colour image. It is shown that the extra structural and functional contrasts provided by using two source modes, namely two-photon excitation microscopy and fluorescence lifetime imaging, result in a more faithful reconstruction of the target haematoxylin and eosin stained mode. This modal mapping procedure can aid histopathologists, since it provides access to unobserved imaging modalities, and translates the high-dimensional numerical data generated by multi-modal, multi-photon microscopy into traditionally accepted visual forms. Furthermore, by combining the strengths of traditional chemical staining and modern multi-photon microscopy techniques, modal mapping enables label-free, non-invasive studies of *in vivo* tissue samples or intravital microscopic imaging inside living animals. The results show that modal co-registration and the inclusion of spatial variations increase the visual accuracy of the mapped results.

1. Introduction

Since its inception in the 19^{th} century, histological staining has become an established, standardised, and reliable diagnostic tool for cytologists [1]. In particular, end users are accustomed to diagnosing the nature of the direct visual results they provide. Staining protocols, such as haematoxylin and eosin (H&E), use reagents to chemically react with specific cellular structures and proteins to produce localised changes in colour and light absorption. Therefore, the resulting enhanced spatial intensity and colour profiles make these typically transparent and colourless structures visible when viewed by optical techniques such as brightfield microscopy. Depending on the nature of the investigation, different staining protocols are used to reveal the morphology and location of specific structures or bio-chemical activities.

Staining reagents can be classified as either vital and/or non-vital. Vital stains can enter and stain living cells for *in vivo* intravital observations. However, although non-toxic at low concentrations, vital stains can accumulate to degrade the tissue sample over time. Furthermore, these reagents are intrusive since they interfere with the biochemical processes within the living cells, thus introducing an unknown effect on the observed system. In contrast, non-vital stains can be absorbed by dead cells and are typically used on fixed *ex vivo* tissue samples.

Although tissue staining provides high throughput rates in automated histo-pathology laboratories, it has several short comings. These include standardisation, mechanical deformation of

#355205 https://doi.org/10.1364/BOE.10.001339
Journal © 2019 Received 12 Dec 2018; revised 4 Feb 2019; accepted 5 Feb 2019; published 19 Feb 2019

Biomedical Optics EXPRESS

the tissue samples, and the requirement for *ex vivo* tissue fixation for the application of certain staining protocols. Furthermore, it is not practical or ethical to apply histological staining for some applications, especially for intravital microscopy studies on living animals.

In contrast, the relatively recent development of multi-photon microscopy (MPM) techniques has provided a powerful set of tools for *in vivo* tissue imaging [2, 3]. These non-linear optical modalities can be divided into two classes: Firstly, intensity methods such as two-photon excitation fluorescence (TPEF) [4], second harmonic generation (SHG) [5], and coherent anti-Stokes Raman scattering (CARS) [6]. Secondly, time resolved methods such as fluorescence lifetime imaging (FLIM) [7].

Unlike one-photon excitation fluorescence (OPEF) microscopy techniques, where a fluorophore is excited by the absorption of a single high-energy photon typically from the ultraviolet spectrum, in MPM techniques, a fluorophore is excited by the simultaneous absorption of two or more lower energy photons typically from the near-infrared spectrum. As well as reducing the level of photodamage experienced by the tissue sample, the greater penetration depth of the near-infrared photons allows deeper tissue observations. Furthermore, rather than using invasive exogenous fluorophores, MPM techniques can also record the fluorescent response of endogenous biomarkers within the tissue sample. Such naturally occurring molecules include nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), keratin, and melanin within the cells; as well as extracellular structural proteins such as collagen and elastin across the tissue sample [8–10]. Therefore, MPM techniques can provide accurate quantitative structural, functional, and metabolic details of intra and extracellular activities to sub-micron resolutions without staining.

Since the different MPM modalities contain complementary information, they can be combined for multi-modal investigations of tissue samples [9–13]. This combination is practically simplified since the different modes are typically spatio-temporally co-registered as they are often measured on the same apparatus. On the other hand, although multi-modal MPM observations provide unprecedented details of *in vivo* cellular activities, they are relatively slow and produce high-dimensional numerical datasets that are not readily diagnosable by unspecialised histopathologists.

Therefore, an attractive solution for non-intrusive *in vivo* observations is to use label-free imaging techniques and then post-process the recorded data to produce accurate visual representations that are renderings of the corresponding stained images. This *modal mapping* of the label-free *source* modes to the stained *target* modes provides several benefits: Firstly, it translates the abstract high-dimensional numerical data produced by multiple label-free imaging techniques to traditional stained images that are commonly diagnosable by histopathologists. Secondly, the histological samples can be stained with any protocol using standardised reagents, both *in vivo* and *ex vivo*, and then observed in a controlled manner to provide high quality brightfield images. The use of these reference target modes for obtaining the modal transfer functions ensures high quality reconstructions. Thirdly, multiple modal mapping functions can be obtained, each transferring the same source tissue type and imaging conditions to a different target staining protocol. This allows the end user to easily switch through different visualisations of the same tissue sample. Finally, once a modal mapping function has been determined, it can then be applied offline to any source dataset that matches the training conditions, namely tissue type and imaging modalities, to obtain the target stained images without physically implementing the staining. Therefore, this technique can be used in label-free *in vivo* or intravital animal microscopy studies to reconstruct images of *ex vivo* staining protocols that could be unethical to implement.

Previously, modal mapping techniques, using physical models for the propagation of light through tissue sections, have been developed to translate fluorescence microscopy observations to pseudo H&E-stained colour images. The source fluorescence modalities include confocal fluorescence microscopy [14], confocal fluorescence and reflectance microscopy [15, 16], and TPEF plus SHG [17].

Biomedical Optics EXPRESS

However, due to the multichannel and thus high-dimensional numerical nature of MPM and digitised stained image datasets, they are well suited for post-processing by machine learning algorithms to reveal even deeper dynamics. For example, machine learning has been used to differentiate between different tissue structures and type of cell death [13]. More generally, the application of deep neural networks (DNNs) to medical computer vision tasks has exploded during the last few years due to the availability of high-performance graphical processor units [18, 19]. DNNs use numerical models to idealise the activity of biological neurons [20], so that when trained with labelled source-target datasets they can learn coherent features of the input that statistically contribute to the outputs in a nonlinear fashion. The trained DNN can then be applied to similar unseen source datasets to classify or map them to new representations. DNNs have been used for improving the resolution of optical microscopes [21], tissue segmentation [22], cancer detection [23], and cancer prognosis [24]. Of particular relevance to this study, DNNs have been previously used to map the single channel total one-photon excited fluorescence response of label-free histological tissue samples to colour images corresponding to different staining protocols [25].

Modal mapping, particularly using DNNs, has several practical difficulties including co-registration of the modes, lack of signal, and DNN training efficiency. These will be discussed below.

A major technical difficulty for modal mapping is the spatial co-registration of the different imaging modalities, such that the source and target pixels coincide with the same location in the tissue sample during the DNN training. As such, modal co-registration is an active area of research [26, 27]. Modal de-registration occurs on both a global and local level. Globally, the translation, rotation and magnification of a tissue sample relative to a fixed datum can change as it is moved and observed by a different imaging system. On the other hand, local de-registration is caused by structural changes in the tissue sample between different observations. This is an important problem for the present application due to the deformation of the tissue sample during staining. These localised deformations arise from changes in the tissue rigidity and distortion during fixation, dehydration, and mechanical strains introduced across the tissue sample during mounting and sectioning. Therefore, the co-registration method should account for both global and local deformations.

A key requirement of modal mapping is the reconstructed stained images should contain the same structural contents, colours, and contrast as those obtained by the chemical staining of the corresponding tissue samples. Furthermore, there should be no processing artefacts which can lead to misdiagnosis. Therefore, the second short coming of the modal mapping technique is the lack of signal. In other words, some structures revealed through staining are not evident when viewed by other imaging modalities and thus cannot be reconstructed from these modalities alone. For example, an organelle revealed through staining may not auto-fluoresce when excited at a given wavelength. Therefore, it may not appear on the reconstructed stained mode since a DNN cannot directly learn its existence when trained with that modality alone. This implies the DNN should be trained using multiple source imaging modalities of the same tissue sample, thus increasing the probability that the stained mode of all relevant structures can be reconstructed.

Finally, DNNs require a large amount of data for their training. Therefore, due to the limited amount of multi-modal data available, it is often necessary to develop DNN architectures that maximise the utility of available data to provide acceptable results. Ideally, a single large field-of-view (FOV) multi-modal megapixel image of a tissue sample should be enough to train the DNN, where the FOV contains hundreds of different representations of each type of histological structure of interest. Therefore, rather than an area-to-area approach [25], where areas of the source modes are mapped onto corresponding areas in the target modes, we have developed techniques based on pixel-to-pixel and area-to-pixel mappings.

In this manuscript we use this approach to develop and implement modal mappings using

Biomedical Optics EXPRESS

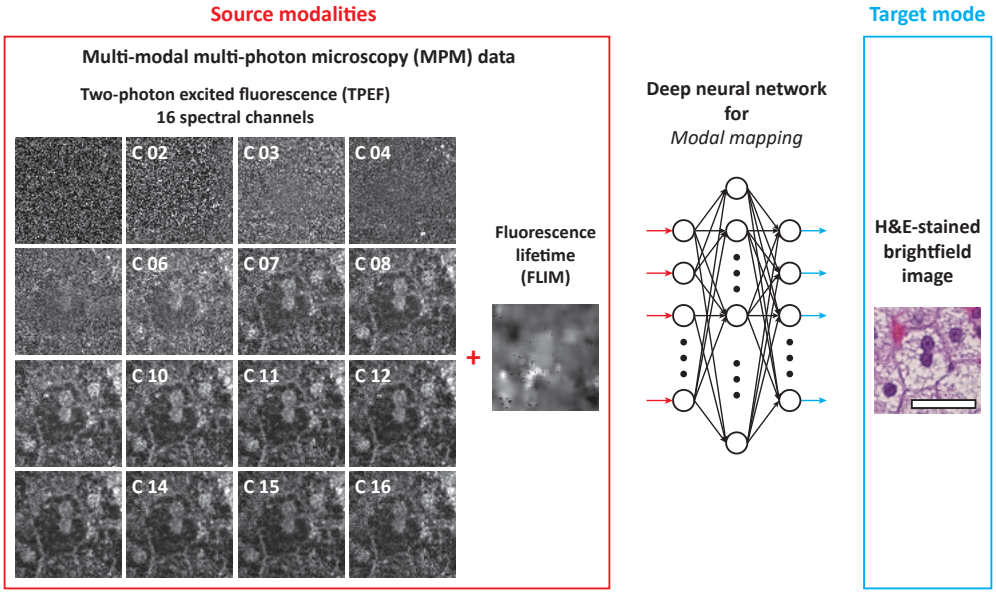

Fig. 1. Schematic of the modal mapping process to transfer the combined TPEF and FLIM observations, of a label-free rat liver tissue sample, to their corresponding H&E-stained brightfield microscope image. To provide contrast in this figure, the 17 MPM channels have been min-max scaled. The length-scale bar in the stained image corresponds to 30 μm.

DNNs. These DNNs were used to produce qualitatively accurate visual reconstructions of the stained images from label-free observations using two different MPM techniques. A combination of TPEF and FLIM was used as the source dataset to train the DNNs. An H&E-stained brightfield microscope image of the same tissue sample was used as the target dataset. An automated tile-wise scheme was developed for the global and local co-registration of the different imaging modalities. Different types of DNN architectures, for area-to-pixel and pixel-to-pixel modal mapping, were developed to maximise the utility of the available multi-modal training data. The relative performance of these DNNs are discussed and suggestions for their practical use are given.

2. Experimental and numerical methods

2.1. Multi-modal training dataset details

The multi-modal dataset used for this study comprised a 16 spectral channel TPEF mode, a single channel FLIM mode, and a 3 channel stained brightfield microscope image of the same rat liver tissue sample [13]. All the channels were recorded as 8-bits per pixel TIFF images. Details of these are given below and shown in Fig. 1.

The tissue section was a 10 μm thick slice of *ex vivo* label-free fixed rat liver tissue mounted on a glass microscope slide. It comprised hepatic cells to which capillaries deliver blood. It was first observed with an integrated multi-modal microscope capable of recording spatially co-registered TPEF, FLIM, SHG, and optical coherence tomography (OCT) modalities [28]. This instrument focuses the 730 nm centered beam from a titanium:sapphire laser (MaiTai HP, Spectra Physics) onto the tissue section with a 0.95 numerical aperture water-immersion objective lens (XLUMP20X, Olympus), to provide less than 7 mW of optical power at the focus. Two computer-controlled galvanometer mirrors (Micromax 671, Cambridge Technology) were

Biomedical Optics EXPRESS

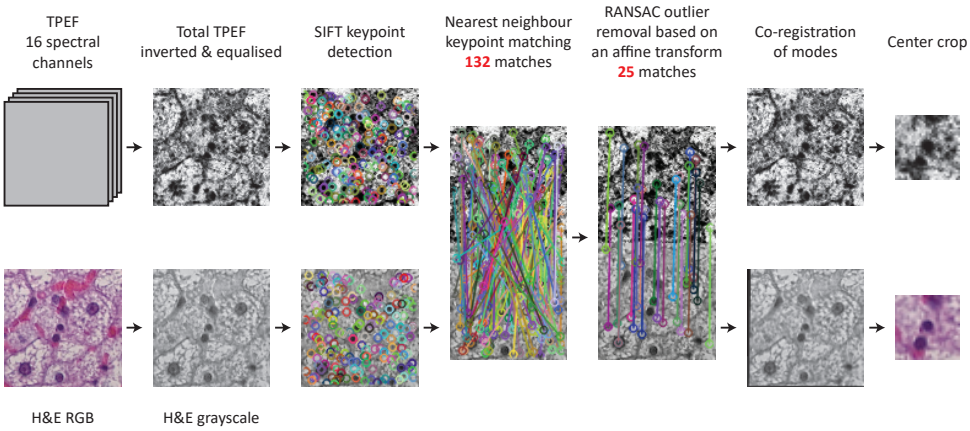

Fig. 2. Schematic of the multi-modal image co-registration scheme used to spatially warp the H&E-stained brightfield images onto the MPM modalities.

then used to raster scan the focused spot across the tissue section to construct the different 2-dimensional MPM images.

The spectro-temporal two-photon fluorescence response of the excited tissue section, at each interrogated pixel location, was measured by recording the output of a 16 spectral channel photomultiplier tube spectrometer (PML-16-C, Becker & Hickl) with a time-correlated single photon counting acquisition card (SPC 150, Becker & Hickl). The spectrally resolved TPEF and the FLIM images were then evaluated using software (SPCimage, Becker & Hickl); where the fluorescence lifetime was based on the wavelength integrated two-photon fluorescence (TPF) signal at each pixel. After measuring its TPF characteristics, the tissue was stained using H&E and observed under a brightfield microscope (Zeiss Axiovert 200).

The animals used in this study were handled and cared for under a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Illinois at Urbana-Champaign.

2.2. Data pre-processing

2.2.1. Modal co-registration

An automated local method was developed to co-register the different imaging modalities. This involved identifying corresponding structures existing across the different modes, then using them to define spatial transformations to warp the target RGB H&E-stained images onto the MPM source modes. It should be noted that the TPEF and FLIM modes were already co-registered since they were recorded simultaneously on the same imaging instrument. Details of the implemented process are described below and shown in Fig. 2.

In the first step, the 16 channel spectrally-resolved TPEF image and the 3 channel RGB H&E-stained image are converted into single channel grayscale images, such that the intensity profiles and thus structural contents of the two images are similar. This allows corresponding features present in both modalities to be accurately identified and located. In this study, the first grayscale image was the intensity component of the RGB H&E-stained image. For the second grayscale image, the total TPEF response was determined by wavelength integration of the 16 TPEF channels. This was then min-max scaled between 0 and 1, inverted, and histogram equalised.

Once the above images were created, characteristic structural *keypoints* were located in each by applying the scale-invariant feature transform (SIFT) [29]. Each keypoint was localised to

Biomedical Optics EXPRESS

sub-pixel accuracy and described with a feature vector that characterised its surrounding intensity profile.

After feature detection, the images and their corresponding datasets were cropped into congruent tiled areas. The detected keypoints, between each corresponding source and target greyscale tiles, were then matched by carrying out a nearest neighbour search between their 128-dimensional keypoint descriptor vectors [29]. However, this typically produces many erroneous matches which prevent the direct evaluation of a continuous input-output warping function.

As a solution to this problem, the random sample consensus (RANSAC) algorithm was used to remove any outlying matches [30]. This approach assumes the in-plane deformations between the two modalities can be accurately described by a specific type of spatial warping function. In this study, it was assumed that the sliced tissue undergoes an affine spatial deformation over the area of each tile. As can be seen from Fig. 2, the outliers are effectively removed from the dataset, allowing the affine transform to be determined and applied to the H&E-stained image, which is then cropped to remove any edge distortions. The combined use of SIFT keypoints and RANSAC function fitting is a classic solution for medical image co-registration [31].

The implemented co-registration process comprised two stages at different length-scales: Firstly, the input and output imaging modes were globally aligned by using tile sizes equal to the full modal image sizes. This removed the mean affine distortion between the imaging modes, thus accounting for different translations and rotations of the tissue sample relative to the imaging systems, and ensuring the modes had the same magnification. In the second stage, matching small-scale structures between the different modes were aligned through a tile-wise local co-registration process. For the present study, the globally aligned modes were cropped into 96x96 pixels tiles with neighbouring tiles shifted by 32 pixels on a Cartesian grid. Once co-registered, the central 32x32 pixel areas of the aligned tiles were cropped to remove any edge effects, and then used to construct the fully co-registered modes in a piecewise manner.

2.2.2. Colour reduction

The aim of this study is to obtain a visually accurate reconstruction of the colour H&E-stained modes, rather than a quantitatively accurate one. Therefore, the colour depth of the brightfield microscope images was decreased. The main motivation for this quantisation was a decrease in the complexity of the required modal transfer DNN architectures, and an increase in the fidelity of the reconstructions.

Therefore, instead of regressing the MPM vectors to real number RGB triplets or classifying them into one of the 2^{24} colours defining each pixel in the true colour RGB images, an optimal colour palette comprising a significantly reduced number of unique colours was generated for the whole image. Such an optimal palette provides a visual representation of the original image with the minimum amount of degradation for the number of reduced colours chosen. Colour reduction also denoises the true colour RGB image by removing background fluctuations, thus providing a cleaner target dataset for training the DNNs.

This was achieved by k-means clustering of the 3-dimensional RGB vectors of all pixels within the acquired H&E-stained mode, where the number of clusters used for the unsupervised procedure is the number of unique colours required to describe the image [32].

Tests showed that the normalised mean squared error (NMSE) and the structural similarity index (SSIM) [33], between the reduced colour image and the ground-truth original, degraded monotonically as the number of colours decreased; see Fig. 3(a). SSIM was chosen as the preferred fidelity criterion since it accounts for the intensity profiles and thus structural similarities of the images; for example, adding a constant value to each pixel of the ground-truth image would produce a large NMSE but a very low SSIM. The SSIM varies from 0 to 1, with 1 corresponding to a perfect match. The results indicate that the degradation is not significant until the original

Biomedical Optics EXPRESS

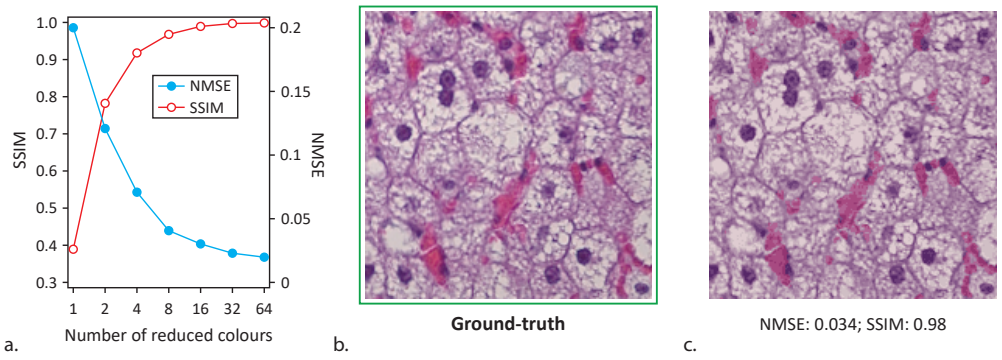

Fig. 3. Application of colour reduction on the H&E-stained images, showing the (a) variation of the NMSE and SSIM between the reduced and (b) original 24-bit colour ground-truth images with increasing number of colours; and (c) the resulting 16 colour reduced image.

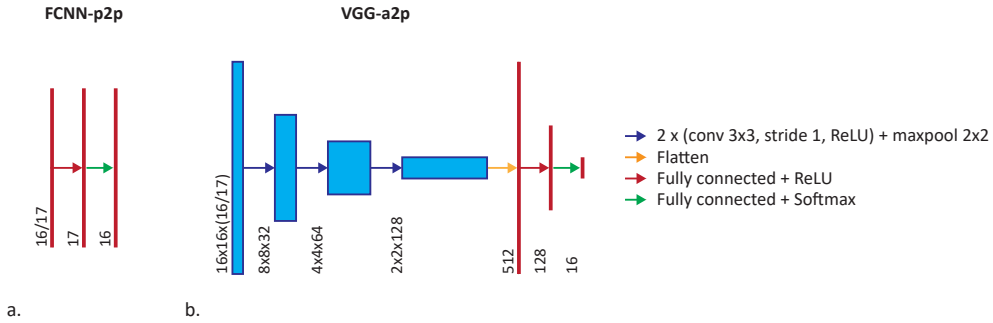

Fig. 4. Schematics of the DNN architectures used to map the multiple source MPM modalities onto the target reduced colour H&E-stained image: (a) FCNN-p2p comprising a fully connected classifier, and (b) VGG-a2p with a CNN spatial feature detector frontend and a fully connected classifier backend.

16.7 million colours have been drastically reduced. Therefore, as a compromise, a palette of 16 optimal colours was chosen for this study, providing an NMSE of 0.034 and an SSIM of 0.98. Comparison of the ground-truth original and the resulting 16 colour reduced image, shown in Figs. 3(b) and 3(c) respectively, highlight the fidelity of this approximation.

2.3. *Deep neural network details*

In order to enhance the efficiency of the learning process, pixel-to-pixel and area-to-pixel mapping solutions were developed; these are described below.

The pixel-to-pixel approach maps the multi-dimensional MPM vector of a given pixel to one of 16 reduced colours defining its H&E-stained value. It makes no account of the local spatial distribution of the input data and thus of any spatially coherent structures within the observations. The DNN used to achieve this was a 2-layer fully connected classifier. We refer to this network as FCNN-p2p and it is shown in Fig. 4(a).

In contrast, area-to-pixel mapping uses a convolutional neural network (CNN) to learn local spatial structures within the input data area which statistically relate to the measured outputs. In this study, a Visual Geometry Group (VGG) type network [34], with a CNN spatial feature detector frontend and a 2-layer fully connected classifier backend, was used to a map a tiled

Biomedical Optics EXPRESS

section of the MPM dataset to the reduced colour of its central pixel location. We refer to this network as VGG-a2p and it is shown in Fig. 4(b).

In order to train the DNNs, a 300x300 pixel region of the 950x950 pixel co-registered multi-modal dataset was first cropped out to act as the test set. For the FCNN-p2p case, the remaining pixels were then randomly split into 610k training and 200k validation samples; the latter were used to monitor the convergence of the learning process and to prevent over-fitting. For the VGG-a2p case, the multi-modal MPM dataset was cropped to form 16x16 pixel tiles with a 1 pixel shift between neighbouring tiles. These were then randomly split into around 500k training and 200k validation samples. The inputs into both networks were either 16 or 17 dimensional, corresponding to TPEF only and the multi-modal TPEF and FLIM cases, respectively. The amplitude of each measured MPM channel was between 0 and 1, and they were not scaled before the application of the DNN. It should be noted that the use of 3-layer fully connected classifiers in both DNN architectures did not improve their reconstruction performance.

For both DNNs, an Adam optimiser [35] with an adaptive learning rate was used to minimise a mean square error cost function using a back-propagation algorithm. The networks were trained with batch sizes of 10k samples until the cost function no longer improved with increasing training epochs; this typically occurred before 50 epochs. The DNNs were implemented using the TensorFlow 1.5 library on an NVIDIA 1080ti graphical processor unit.

The performance of the DNNs was characterised by the NMSE and SSIM between the reconstructions and the known ground-truth. The ground-truth for all comparisons was the original 24-bit RGB H&E-stained test image, and the NMSE was normalised with the mean value of this ground-truth.

3. Results and discussion

The relative performance of the different modal mapping schemes with respect to different DNN architectures, modal co-registration, and number of source MPM modalities is shown in Fig. 5. This shows the reconstructed H&E-stained images generated by the respective trained DNNs when applied to the unseen cropped test source dataset. The SSIM and NMSE values shown in Fig. 5 represent the mean and standard deviation of 9 different training instances. For each instance, a different independent test crop image was taken from the dataset to act as the ground-truth for evaluating the modal mapping accuracies. The remaining data were then randomised before being split into training and validation subsets.

The results show the importance of accurate modal co-registration during training: For the pixel-to-pixel mapping of FCNN-p2p, co-registration increases the level of detail found in the stained images, resulting in an improvement of the performance metrics. This is because the fluorescence vectors are mapped onto the correct stained colours for each pixel during training, thus enhancing any correlations present. The necessity for co-registration is particularly evident for the area-to-pixel mapping of VGG-a2p. In this case, the lack of spatial congruence, between corresponding structures in the source and target modes, hinders feature extraction by the CNN frontend during training. This greatly reduces the fidelity of the mapping.

The results also indicate that the additional structural and functional contrasts, provided by using multiple source MPM modalities, result in more accurate H&E-stained image reconstructions. This is because different liver tissue structures display different TPF signatures when observed by different MPM modalities [13]. Namely, the hepatocyte nuclei have a greater response than the red blood cells when observed by TPEF, whilst the red blood cells have a slightly greater response when viewed by FLIM. The improvement in modal mapping accuracy is particularly evident for the FCNN-p2p case where the haemoglobin is almost absent when using only the TPEF mode, whilst it becomes evident when trained with both the TPEF and FLIM modes; albeit with the generation of many false artefacts.

Comparison between the two DNN architectures indicates VGG-a2p provides a much better

Biomedical Optics EXPRESS

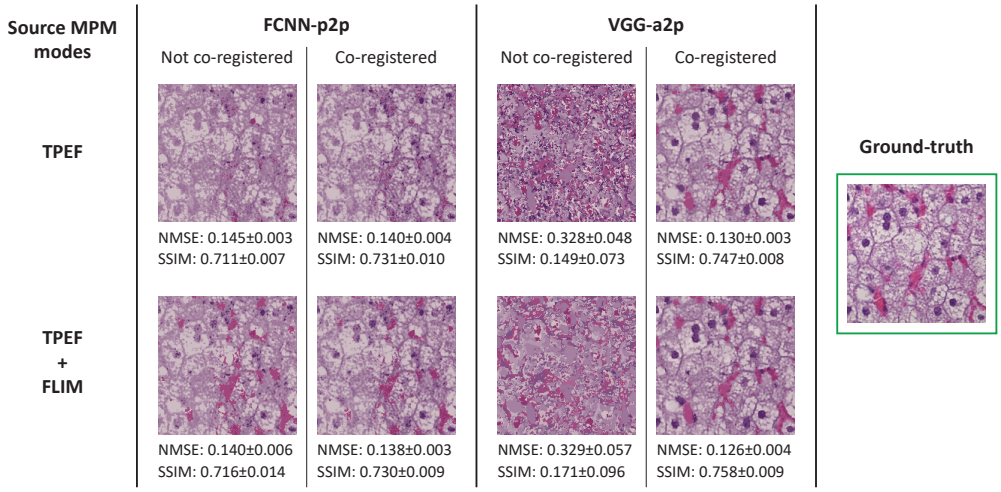

Fig. 5. Reconstructed H&E-stained images generated by different DNNs when applied to unseen test source datasets. The NMSE and SSIM accuracies are relative to the 24-bit colour ground-truth image. As can be seen, the reconstruction fidelity increases when multiple source modes (TPEF+FLIM) are used, and when spatial variations are learnt (VGG-a2p).

visual representation of the stained images than FCNN-p2p for the same source modes. This is evident from the presence and appearance of structures such as hepatocyte nuclei and red blood cells, as well as a reduction in the number of erroneous artefacts. This is because during training the area-to-pixel approach learns to associate the presence of specific structures in the source data with the pixel colour values of the target, thus allowing them to be identified in the test source data and mapped accurately onto the target. For example, as shown in the bottom right-hand corner of the reconstructed H&E-stained images, FCNN-p2p can generally infer the presence of haemoglobin from the TPF signature of a pixel, whilst VGG-a2p is better able to spatially differentiate the haemoglobin signal to either red blood cells or that free-floating in the blood serum.

Figure 6 shows the results when DNNs trained on the rat liver tissue section are used for modal mapping from different types of tissues. These samples were mouse ovary and rat mammary tumour tissues [13]; shown in Figs. 6(a) and 6(b), respectively. The poor performance obtained in these cases emphasises that trained DNNs can only be applied to source datasets which closely match their training conditions. These include focus, magnification, and tissue type. The main reason for this is that different tissue types have different structural and functional components with different TPF signatures. Therefore, a DNN trained to learn a specific set of these, in order to generalise a particular modal mapping scheme, can not be applied to unlearnt sets unless the training set is enlarged to include these variable conditions.

For example, the sensitivity of the DNNs to magnification is due to two reasons: Firstly, for both the FCNN-p2p and VGG-a2p cases, the fluorescence signatures of two different adjacent structures may become mixed as the magnification is reduced. Therefore, a DNN trained to recognised the resolved case will not recognise the new mixed signature at the lower magnification, thus leading to reconstruction errors. This problem can be reduced by training the DNNs with multi-modal data taken at different magnifications. Secondly, since VGG-a2p is not scale invariant it cannot recognise a learnt structural feature if its size is different in the unseen dataset. A solution to this scale dependency is the use of scale invariant DNNs.

In addition, the pixel-to-pixel mapping provided by FCNN-p2p appears to produce relatively

Biomedical Optics EXPRESS

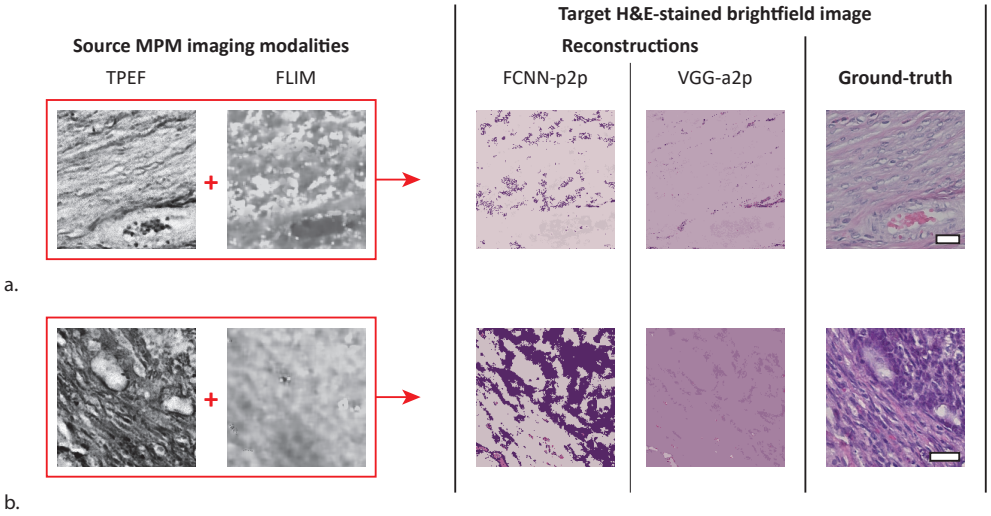

Fig. 6. Modal mapping results when DNNs, trained to reconstruct H&E-stained images from multi-modal MPM observations of a rat liver tissue sample, are applied to (a) mouse ovary, and (b) rat mammary tissues. The length-scale bars in the ground-truth stained images correspond to 30 μm.

better results. However, this apparent segmentation is a numerical artefact since pixels with similar MPM feature vectors in the new tissue images will tend to be classified into the same reduced colour, regardless of their similarity to the trained vectors. Furthermore, the lack of similar spatial features between the different tissues greatly degrades the H&E-stained reconstructed images produced by the VGG-a2p scheme.

Therefore, in summary, the results show that the best modal mapping performance is obtained by training a CNN-based area-to-pixel classification scheme with co-registered data comprising multiple source imaging modalities.

4. Conclusions

Deep neural networks have been successfully trained to generate a visually accurate reconstruction of the brightfield H&E-stained image of a label-free tissue sample observed by multiple MPM modalities.

The applied DNN were simple and maximised the utility of the available multi-modal training data. Colour reduction was used to simplify the modal mapping process to a colour classification problem. The use of the complementary information provided by two source modes, namely TPEF and FLIM, provided a more faithful reconstruction of the target H&E-stained mode; thus highlighting the need for high-dimensional multi-modal MPM datasets for such applications. The visual accuracy of the reconstructions also improved with modal co-registration and the inclusion of spatial contrast variations.

Modal mapping using trained DNNs can provide several advantages for histology: Firstly, they can be applied to any similar source data sets to generate the unobserved stained images. Secondly, the abstract MPM modalities are mapped onto more accessible stained images which are readily diagnosable; thus also promoting the wider use of MPM techniques for pathology. Thirdly, due to the label-free nature of MPM techniques using endogenous fluorophores, modal mapping can provide many different corresponding vital and non-vital staining protocol visualisations of *in vivo* or intravital studies. This can be achieved either real-time or offline. Finally, only a small

Biomedical Optics EXPRESS

histological biopsy sample is enough to simultaneously provide many different staining protocol visualisations.

Foreseen applications of the above DNN modal mapping include label-free deep tissue imaging, *in vivo* drug screening, real-time intravital surgical monitoring, and post-therapy *in vivo* clinical evaluations.

Funding

U.S. National Institutes of Health (R01 EB023232 and R01 EB013723); the U.S. National Science Foundation (CBET 18-41539).

Acknowledgments

The authors wish to thank Eric Chaney for animal handling and tissue processing, Ronit Barkalifa for managing the animal study protocol, and Darold Spillman for logistical and IT support.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. S. K. Suvarna, C. Layton, and J. D. Bancroft, *Bancroft's Theory and Practice of Histological Techniques, 7th Edition* (Churchill Livingstone Elsevier, 2013).
2. C. Lefort, "A review of biomedical multiphoton microscopy and its laser sources," J. Phys. D Appl. Phys. **50**, 423001 (2017).
3. E. E. Hoover and J. A. Squier, "Advances in multiphoton microscopy technology," Nat. Photonics **7**, 93–101 (2013).
4. P. T. C. So, C. Y. Dong, B. R. Masters, and K. M. Berland, "Two-photon excitation fluorescence microscopy," Annu. Rev. Biomed. Eng. **2**, 399–429 (2000).
5. P. J. Campagnola and L. M. Loew, "Second harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms," Nat. Biotechnol. **21**, 1356–1360 (2003).
6. L. G. Rodriguez , S. J. Lockett, and G. R. Holtom, "Coherent anti-Stokes Raman scattering microscopy: A biological review," Cytom. Part A **69A**, 779–791 (2006).
7. W. Becker, "Fluorescence lifetime imaging – techniques and applications," J. Microsc. **247**(2), 119–136 (2012).
8. H. G. Breunig, H. Studier, and K. König, "Multiphoton excitation characteristics of cellular fluorophores of human skin *in vivo*," Opt. Express **18**(8), 7857–7871 (2010).
9. E. Benati, V. Bellini, S. Borsari, C. Dunsby, C. Ferrari, P. French, M. Guanti, D. Guardoli, K. Koenig, G. Pellacani, G. Ponti, S. Schianchi, C. Talbot, and S. Seidenari, "Quantitative evaluation of healthy epidermis by means of multiphoton microscopy and fluorescence lifetime imaging microscopy," Skin Res. Technol. **17**(3), 295–303 (2011).
10. Y. Zhao, M. Marjanovic, E. J. Chaney, B. W. Graf, Z. Mahmassani, M. D. Boppart, and S. A. Boppart, "Longitudinal label-free tracking of cell death dynamics in living engineered human skin tissue with a multimodal microscope," Biomed. Opt. Express **5**(10), 3699–3716 (2014).
11. H. Tu, Y. Liu, D. Turchinovich, M. Marjanovic, J. K. Lyngsø, J. Lægsgaard, E. J. Chaney, Y. Zhao, S. You, W. L. Wilson, B. Xu, M. Dantus, and S. A. Boppart, "Stain-free histopathology by programmable supercontinuum pulses," Nat. Photonics **10**, 534–541 (2016).
12. A. J. Bower, M. Marjanovic, Y. Zhao, J. Li, E. J. Chaney, and S. A. Boppart, "Label-free *in vivo* cellular-level detection and imaging of apoptosis," J. Biophotonics **10**(1), 143–150 (2017).
13. A. J. Bower, B. Chidester, J. Li, Y. Zhao, M. Marjanovic, E. J. Chaney, M. N. Do, and S. A. Boppart, "A quantitative framework for the analysis of multimodal optical microscopy images," Quant. Imaging Med. Surg. **7**(1), 24–37 (2017).
14. J. Dobbs, S. Krishnamurthy, M. Kyrish, A. P. Benveniste, W. Yang, and R. Richards-Kortum, "Confocal fluorescence microscopy for rapid evaluation of invasive tumor cellularity of inflammatory breast carcinoma core needle biopsies," Breast Cancer Res. Treat. **149**(1), 303–310 (2015).
15. D.S. Gareau, "Feasibility of digitally stained multimodal confocal mosaics to simulate histopathology," J. Biomed. Opt. **14**(3), 034050 (2009).
16. J. Bini, J. Spain, K. Nehal, V. Hazelwood, C. DiMarzio, and M. Rajadhyaksha, "Confocal mosaicing microscopy of human skin *ex vivo*: spectral analysis for digital staining to simulate histology-like appearance," J. Biomed. Opt. **16**(7), 076008 (2011).
17. M. G. Giacomelli, L. Husvogt, H. Vardeh, B. E. Faulkner-Jones, J. Hornegger, J. L. Connolly, and J. G. Fujimoto, "Virtual hematoxylin and eosin transillumination microscopy using epi-fluorescence imaging," PLoS ONE **11**(8), e0159337 (2016).

Biomedical Optics EXPRESS

18. H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," IEEE T. Med. Imaging **35**(5), 1153–1159 (2016).
19. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," Med. Image Anal. **42**, 60–88 (2017).
20. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
21. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**(11), 1437–1443 (2017).
22. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumour segmentation using convolutional neural networks in MRI images," IEEE T. Med. Imaging **35**(5), 1240–1251 (2016).
23. D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," arXiv:1606.05718 (2016).
24. P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. V. Vega, D. J. Brat, and L. A. D. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," PNAS **115**(13), E2970–E2979 (2018).
25. Y. Rivenson, H. Wang, Z. Wei, Y. Zhang, H. Günayfin, and A. Ozcan, "Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue," arXiv:1803.11293 (2018).
26. J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," Med. Image Anal. **2**(1), 1–36 (1998).
27. M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration - under review," Med. Image Anal. **33**, 140–144 (2016).
28. Y. Zhao, B. W. Graf, E. J. Chaney, Z. Mahmassani, E. Antoniadou, R. DeVolder, H. Kong, M. D. Boppart, and S. A. Boppart, "Integrated multimodal optical microscopy for structural and functional imaging of engineered and natural skin," J. Biophotonics **5**(5-6), 437–448 (2012).
29. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis. **60**(2), 91–110 (2004).
30. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," Commun. ACM **24**(6), 381–395 (1981).
31. G. Lippolis, A. Edsjö, L. Helczynski, A. Bjartell, and N. C. Overgaard, "Automatic registration of multi-modal microscopy images for integrative analysis of prostrate tissue sections," BMC Cancer **13**, 408 (2013).
32. M. E. Celebi, "Improving the performance of k-means for colour quantization," Image Vision Comput. **29**(4), 260–271 (2011).
33. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE T. Image Process. **13**(4), 600–612 (2004).
34. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
35. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 (2014).