International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-019-01940-2

ORIGINAL ARTICLE

Estimation of tissue oxygen saturation from RGB images and sparse hyperspectral signals based on conditional generative adversarial network

Qingbiao Li[1,2] · Jianyu Lin[1,3] · Neil T. Clancy[4,5] · Daniel S. Elson[1,2]

Received: 29 January 2019 / Accepted: 7 March 2019
© The Author(s) 2019

Abstract

Purpose Intra-operative measurement of tissue oxygen saturation (StO_2) is important in detection of ischaemia, monitoring perfusion and identifying disease. Hyperspectral imaging (HSI) measures the optical reflectance spectrum of the tissue and uses this information to quantify its composition, including StO_2. However, real-time monitoring is difficult due to capture rate and data processing time.

Methods An endoscopic system based on a multi-fibre probe was previously developed to sparsely capture HSI data (sHSI). These were combined with RGB images, via a deep neural network, to generate high-resolution hypercubes and calculate StO_2. To improve accuracy and processing speed, we propose a dual-input conditional generative adversarial network, Dual2StO2, to directly estimate StO_2 by fusing features from both RGB and sHSI.

Results Validation experiments were carried out on in vivo porcine bowel data, where the ground truth StO_2 was generated from the HSI camera. Performance was also compared to our previous super-spectral-resolution network, SSRNet in terms of mean StO_2 prediction accuracy and structural similarity metrics. Dual2StO2 was also tested using simulated probe data with varying fibre number.

Conclusions StO_2 estimation by Dual2StO2 is visually closer to ground truth in general structure and achieves higher prediction accuracy and faster processing speed than SSRNet. Simulations showed that results improved when a greater number of fibres are used in the probe. Future work will include refinement of the network architecture, hardware optimization based on simulation results, and evaluation of the technique in clinical applications beyond StO_2 estimation.

Keywords Intro-operative imaging · Optical imaging · Tissue oxygen saturation · Generative adversarial network

Introduction

Tissue perfusion and oxygenation are important clinical indicators of organ health during minimal access surgery (MAS). Endoscopic hyperspectral imaging (HSI) is a non-invasive optical technique to capture quantitative spectral information with a high spatial resolution based on narrow spectral bands over a virtually continuous spectral range for live tissue diagnostics and monitoring [1]. HSI can be used to estimate oxygen saturation (StO_2) and perfusion, which reflects tissue function and the health of an organ's blood supply. This, in turn, can be applied to various important clinical applications

✉ Qingbiao Li
ql716@ic.ac.uk

Jianyu Lin
xjtuljy@gmail.com

Neil T. Clancy
n.clancy@ucl.ac.uk

Daniel S. Elson
daniel.elson@imperial.ac.uk

1 The Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK

2 Department of Surgery and Cancer, Imperial College London, London, UK

3 Department of Computing, Imperial College London, London, UK

4 Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London, London, UK

5 Centre for Medical Image Computing (CMIC), Department of Computer Science, University College London, London, UK

Published online: 21 March 2019

[1], including monitoring of cortical haemodynamics during brain surgery [2], reperfusion during organ transplantation [3] and detection of intestinal ischaemia [4]. High-resolution spectral data can also be used to characterize tissue and detect subtle differences between normal and dysplastic areas [5]. HSI is a non-contact technique, compatible with conventional surgical light sources and endoscopes, and has some important advantages over competing optical techniques, such as photoacoustic tomography (PAT) [6], which requires ultrasound contact and a complex laser source.

HSI requires acquisition of a hypercube, which has one spectral and two spatial dimensions. Imaging hardware may use tunable filters or spatial scanning, but does not typically achieve real-time operation due to the data capture and processing times. Snapshot spectral imaging acquires the entire hypercube simultaneously, but the number of wavelengths or spatial resolution must be sacrificed to achieve high-speed acquisition. This trade-off between spectral information, spatial resolution and acquisition speed is a barrier for clinical use of HSI and other optical imaging techniques [1].

To overcome this, we previously developed a dual-mode structured light and hyperspectral imaging (SLHSI) system [7,8] to capture sparse hyperspectral images in real-time, as illustrated in Fig. 1a. The light (i.e. reflectance or fluorescence) from the tissue surface was imaged onto the 2D fibre array, and the bundle randomly re-ordered the fibres into a linear array at the other end. The spectrum carried by each fibre could then be captured by imaging the linear array onto the entrance slit of an imaging spectrograph. The data could then be rearranged computationally to generate sparse hyperspectral images (sHSI) in a snapshot. The 2.8 mm fibre bundle can be inserted through an endoscope biopsy port or attached to the endoscope or another surgical instrument [7]. The system could also be used to record spectrally encoded structured lighting (SL) images [7,8], although this capability is not explored further in this paper.

To process the acquisition, a super-spectral-resolution network, called SSRNet, was proposed to integrate dense RGB images and sHSI for pixel-level hypercube estimation [8]. The hypercube could be used to estimate StO_2 based on the modified Beer-Lambert law as illustrated in processing Route 1 (Blue line) in Fig. 1b. Previous work also explored the feasibility of estimating StO_2 directly from RGB images (Route 2, Fig. 1b) [9], and showed that hyperspectral information improves the accuracy of the result [10]. However, as the aim of SSRNet was to predict dense HSI hypercubes, it was not explicitly optimized for StO_2 estimation, and the value of combining RGB images with sparse HSI hypercubes has not been evaluated for estimating dense StO_2 maps.

In this paper, we extend the previous published results, proposing a dual-input network using cGAN, Dual2StO2, to achieve dense StO_2 estimation using end-to-end learning, without the need for the intermediate spectral estimation step. The proposed network was inspired by the performance of GANs in super-resolution [11,12], to achieve super-resolution estimation in spatial (for sHSI) and spectral (for RGB) domains. A minimax two-player game was utilized between the generator and discriminator to further improve the accuracy of per-pixel regression for StO_2 estimation. By adding conditional input, the generator in cGAN would estimate StO_2 imitating the structure of the condition, instead of random image generation in GAN. The results from Mirza and Osindero [13] and Isola et al. [14] also supported that cGAN could achieve higher pixel-level accuracy than other GANs with the same settings. The relationship between two input modalities (RGB, sHSI) and output (StO_2) was known a priori, which enabled the network to be trained by supervised learning, achieving faster convergence and prediction accuracy. Additionally, a customized mask was added to filter saturated pixels and unreliable estimates at the pixel level. Furthermore, one of the key parameters in designing the MSI data acquisition system is the number of fibres in the bundle, and we have therefore additionally simulated the per-

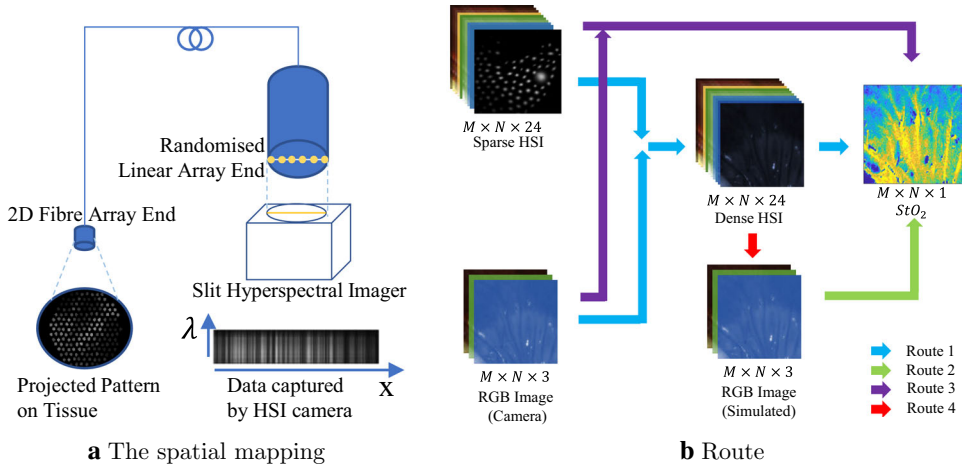
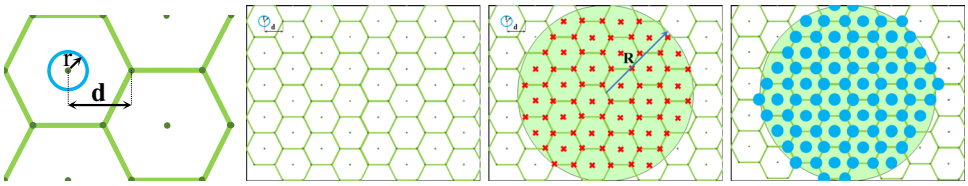


Fig. 1 **a** The spatial mapping between randomized end (2D) and the linear end (1D); **b** processing methods for StO_2 estimation



a Step1. **b** Step2. **c** Step3. **d** Step4.

Fig. 2 Illustration of the generation of the mask, where fibre core sensing areas (blue circle) are defined with respect to the hexagonal grid in Step 1 and 2. The radius of the fibre bundle (shaded green area) and individual fibre positions (red crosses) are defined in Step 3. Fibre cores within the bundle (blue dots) are masked off in Step 4

formance of this estimation for different fibre bundles. This approach is represented by Route 3 (Purple line) in Fig. 1b. This will allow optimization of future hardware designs to increase robustness.

In this paper, the "Data acquisition and preprocessing" section will describe data acquisition and HSI data synthesis, while Dual2StO2 is presented in "Dual-input network for StO2 estimation" section. The evaluation metrics and validation setup for this method are described in the "Experiments" section, followed by a validation of the network via an animal study on porcine bowel in vivo. The previous two-stage StO_2 estimation approach (Route 1, Blue line) developed by Lin et al. [10] in Fig. 1b was adopted as the baseline against which the performance of the proposed network was evaluated.

Materials and methods

Data acquisition and preprocessing

The porcine bowel in vivo data was captured by a liquid crystal (LCTF)-based HSI system in the wavelength range 460–700 nm with 10 nm interval, as described in a previous work [15]. Here, a subset of the spectral data from 460 to 690 nm was considered as a ground truth 24-channel hypercube with spatial size 256 × 192 pixels. A total of 50 acquisitions were selected from 15 separate animals.

Simulated RGB images The RGB image (Input-x) was simulated from the hypercube (Route 3 in Fig. 1b) using the known spectral response of a colour camera [15,16].

Analytical method to estimate StO_2 A well-established linear model based on the modified Beer–Lambert law was used in this paper to obtain ground truth StO_2. It uses linear regression to estimate the relative concentrations of oxygenated and deoxygenated haemoglobin (HbO_2 and Hb) and calculates StO_2 as the quantity of HbO_2 as a fraction of total haemoglobin (HbO_2 + Hb), subject to assumptions [15]. Experimental validation has also been carried out in our previous in vivo uterine transplantation and bowel surgery experiments [3,15] as well as by others [2,17]. The coefficient of determination (CoD) [18] was used to evaluate the accuracy of the linear regression estimation. CoD $\leq$ 0.85 is set as threshold for linear regression outliers, and related pixels were excluded for training and evaluation. Pixels located in non-tissue regions, insufficiently illuminated areas and specular reflections were also excluded.

Synthesized sparse hyperspectral images A number of different distal tip fibre arrangements may be chosen for the experimental hardware. To study how this may affect the performance of the Dual2StO2 network and thereby influence future decisions on the experimental setup, we have simulated data acquired with different fibre arrangements from the dense hyperspectral dataset in the "Data acquisition and preprocessing" section. The use of circular masks with high-resolution ground truth images to simulate and assess the performance of imaging fibre bundles has previously been demonstrated [19,20]. Masks were created in MATLAB (R2018a; The Math Works, Inc., USA) using a circular sensing area arranged on a hexagonal grid to represent the array of fibres, with the spatial information averaged within these areas, as illustrated in Fig. 2 and described in the following steps.

Step 1 Define a radius (r, representing the transmissive fibre cores), and horizontal and vertical spacing between the spot centres (d, a metric representing the core separations), where the ratio $\gamma = \frac{r}{d}$ is the fill factor that defines the relationship between the area of the projected spot and and the space. In reality, γ = stays unchanged when changing fibre numbers, as the dimensions of individual fibres and their cladding are consistent, which also described in Table 1;

Step 2 Generate a hexagonal grid across the whole image to simulate a hexagonal packed fibre probe ($[W_s, H_s, W_e, H_e] = [0, 0, W, H]$, where W and H are the image width and height, the subscripts s and e stands for start and end of the image range, respectively);

Step 3 Define the radius ($R = \frac{h}{2}$) of the fibre bundle (green circle);

Step 4 Generate a mask that includes all fibre cores within the bundle;

Step 5 Average the spatial information within each fibre sensing area to generate a single spectrum for each fibre.

Data augmentation The training data were augmented by horizontal, and vertical flipping, and image cropping using a sliding 96 × 96 window with a stride of 16. The cropped images were resized to the target size, 256 × 256 through bilinear interpolation, which augments 231 times of 38 original images and results in 8778 images for training. In order to maintain information consistency, the 96 × 96 central regions cropped from the original 12 images were resized to 256 × 256 through bilinear interpolation and used as test set.

Dual-input network for StO2 estimation

Dual2StO2 is a cGAN-based image-to-image translation network for StO_2 estimation utilizing dual-input modalities (RGB, sHSI), which was implemented in Pytorch 4.0. In analogy with automatic language translation, image-to-image translation defined by Isola et al. [14] is a task that translates the representation of one scene into another, which is implemented as a general framework called pix2pix for per-pixel classification and regression. Its fundamental network was based on cGAN, where additional conditions were added for both the generator and discriminator [13].

Generator (***G***) Inspired by the network architecture of pix2pix [14], it was adopted as the base model in the generator of Dual2StO2, because the relationship between two input modalities (RGB, sHSI) and output (StO_2) was known a priori (suitable for supervised learning). The network architecture of the generator was modified based on a multi-input unsupervised learning image-to-image translation framework, called In2I [21], as illustrated in Fig. 3.

- The encoder (light orange box) was designed to first extract features from the RGB image (256 × 256 × 3) and sparse HSI (256 × 256 × 24) (pink region), fuse the feature map from these two modalities by concatenation (grey box), and extract further features from the fused feature map;
- The decoder (light green box) was introduced to decode the feature map and output the StO_2 estimation;
- Residual block (brown arrow, with process illustrated in the brown box) proposed by He et al. [22] was adopted in both the encoder and decoder;
- Instance normalization was adopted based on comparison work [23,24], where the results indicated that instance normalization has better performance in image generation tasks than batch normalization;
- One mask was created to filter the position of pixels with saturated pixel values due to specular reflections (NaN), and those with a coefficient of determination (CoD) ≤ 0.85;

In the training stage of the simulation experiment, the two input modalities (simulated RGB, and the synthesized sparse hyperspectral image called synthesized sHSI) defined as $S = \{S_{\mathrm{RGB}}, S_{\mathrm{sHSI}}\}$ were fed into the generator (Fig. 3) and in Fig. 3 and trained to learn a forward transformation $f_{S \to T(s)}$ to output a single set of images (StO_2) from the "Data acquisition and preprocessing" section, under the condition of source domain S. Here, the source and target domain were defined by S and T, with the data distributions of domain S and T as $p_{\mathrm{data}(s)}$ and $p_{\mathrm{data}(t)}$. The similar notations in In2I [21] are used here.

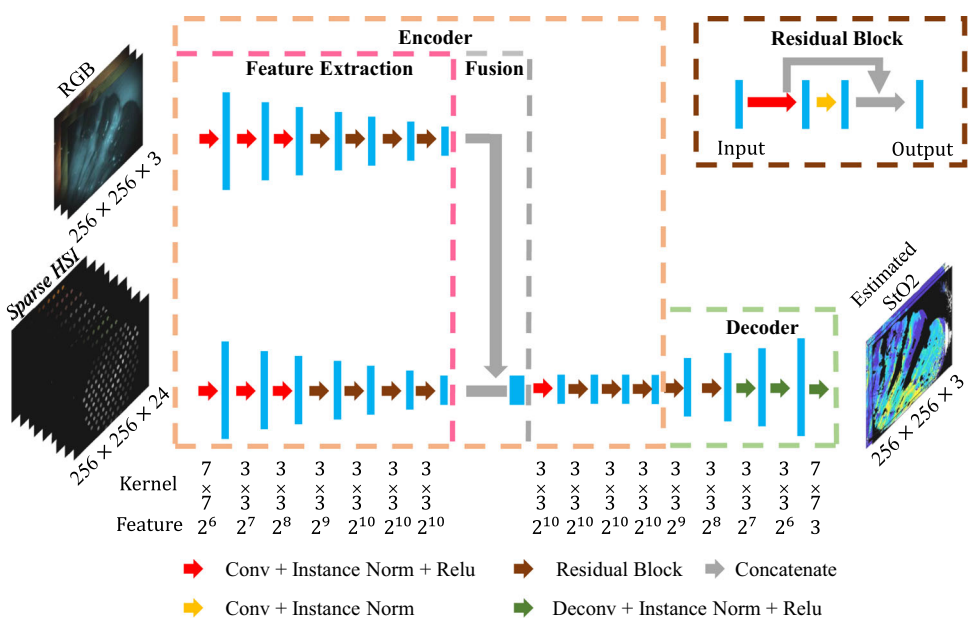


Fig. 3 Network architecture of the generator **G** in proposed Dual2StO2

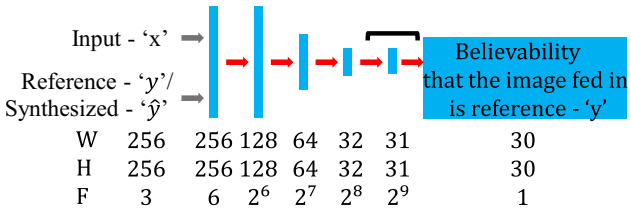


Fig. 4 Network architecture of the discriminator **D** to output believability between 0 to 1 that the image is synthesized $\hat{\mathbf{y}}$ to reference image y, where W, H and F are the width, height and channel size of feature map

Discriminator (***D***) Under the condition of observed image ($\mathbf{x}$) from input domain S, the discriminator **D** will estimate the probability of whether the image is the ground truth image ($\mathbf{y}$) from target domain T, or the synthesized image ($f_{S\to T(s)}, \hat{\mathbf{y}}$) generated by generator **G**. A convolutional network, called PatchGAN, was first introduced by Li et al. [25] to classify real or fake images based on individual image patches. A comparison on different patch size was carried by Isola et al. [14] and showed that the performance of image-to-image translation was best with 70×70 patch size. This size was adopted into the implementation of the discriminator. Concat ($\mathbf{x}, \mathbf{y}$) and Concat ($\mathbf{x}, \hat{\mathbf{y}}$) are put into a discriminator separately, which outputs the probability of the input to be $\mathbf{y}$. Here Concat() is concatenate, the probability map is a $30 \times 30 \times 1$ map which is useful for pixel-level rather than image-level translation. The discriminator network architecture is shown in Fig. 4.

Adversarial learning During the training stage, the generator tries to generate a synthesized image ($f_{S\to T(s)}$) as real as possible to cheat the discriminator to consider it as real. On the other hand, the discriminator will also improve its ability to make correct judgement on whether the images are the ground truth image from target domain T, or the synthesized image ($f_{S\to T(s)}$) generated by generator. Hence, this forward transformation ($f_{S\to T(s)}$) is trained by the adversarial loss function Eq. 1:

$$\begin{aligned}\mathcal{L}_{total} &= \mathcal{L}_{\text{cGAN},S\to T} + \beta\mathcal{L}_1(S \to T)\\ &= E_{t\sim p_{\text{data}(t)}}[\log D_T(t)]\\ &\quad + E_{s\sim p_{\text{data}(s)}}[1 - \log D_T(f_{S\to T(s)})]\\ &\quad + \beta E_{s\sim p_{\text{data}(s)}}[||(f_{S\to T(s)}) - t||_1]\end{aligned} \tag{1}$$

where D is the discriminator and β is the weight of the L1 norm, set as 400 based on previous work [9].

A minimax two-player game is introduced in this network to train the generator and discriminator through an adversarial process. Hence, the generator in Fig. 3 is trained to maximize the probability of discriminator to be a false positive. Our final objective is defined by Eq. 2:

$$\mathcal{G} = \arg\min_G \max_D [\mathcal{L}_{\text{cGAN}}(G, D) + \beta\mathcal{L}_1(G)] \tag{2}$$

where G is the generator.

Experiments

In this section, the evaluation metrics are firstly defined in the "Evaluation metrics" section to quantitatively analyse the performance of StO_2 estimation, followed by the experiment setup in the "Experimental Setup" section.

Evaluation metrics

- *Structural similarity index (SSIM)* A perception-based method proposed by Wang et al. [26] comparing the local patterns of pixel intensities that have been normalized for luminance and contrast. The similarity between the ground truth and synthesized image could be measured and quantified between 0 and 1, where $SSIM = 1$ was considered as identical.
- *Mean prediction error* ($\bar{e}$) The difference in StO_2 value between the ground truth and synthesized image at each pixel measured by the L1 norm.

$$\begin{aligned}\bar{e}(i, j) &= \frac{\sum_{i=0}^{W}\sum_{j=0}^{H} e(i, j)}{n_{\text{effective}}}\\ &= \frac{\sum_{i=0}^{W}\sum_{j=0}^{H} ||I_{\text{syn}(i,j)} - I_{\text{gt}(i,j)}||_1}{n_{\text{effective}}}\end{aligned} \tag{3}$$

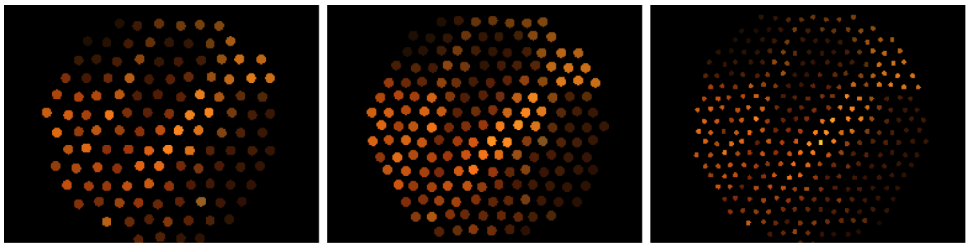
  where I_{syn} and I_{gt} are the absolute values for a pixel at column i, row j, in the synthesized and ground truth images with width W and height H, respectively, and $n_{\text{effective}}$ is the total number of pixels in the image excluding saturated and low CoD pixels
- *Fraction of pixels with high accuracy level* (p_{HAP}) Accuracy is defined above a certain level compared to the pixel data in the ground truth image.

$$p_{\text{HAP}} = \frac{n_{\text{HAP}}}{n_{\text{effective}}} \tag{4}$$

  where n_{HAP} is the number of pixels with high prediction accuracy (i.e. $1 - e_{(i,j)} \geq 95\%$).

Experimental setup

Animal studies were carried out to validate the performance of Dual2StO2 on the in vivo acquisitions by separating animals into training and test data sets. The training set consisted of 38 acquisitions captured for 10 animals (animal ID: 1–10), while 12 further test acquisitions were from the 5 remaining animals (animal ID: 11–15).



a $n_{spot} = 121$ **b** $n_{spot} = 171$ **c** $n_{spot} = 300$

Fig. 5 Synthesized sHSI displayed as three-channel images generated by taking intensities at selected wavelengths ($\lambda = 460, 520, 590$ nm)

The bundles with circular distal cross sections and different numbers of fibres ($n_{\text{spot}} = 0, 121, 171, 300$) were simulated. Two of these configurations (121 and 171 spots) were chosen as they match the existing hardware available, complemented by a fibre bundle with a high number of spots (300). To confirm any benefit of integrating sparse HSI, a fibre bundle with zero spots was used as the control group. Figure 5 illustrates sample synthesized sHSI images generated by the corresponding masks. These sHSI and simulated RGB images were fed into Dual2StO2. Route 1 in Fig. 1b based on the SSRNet developed by Lin et al. was adopted as the baseline to compare the performance of StO_2 based on the same simulated RGB and synthesized sHSI as input.

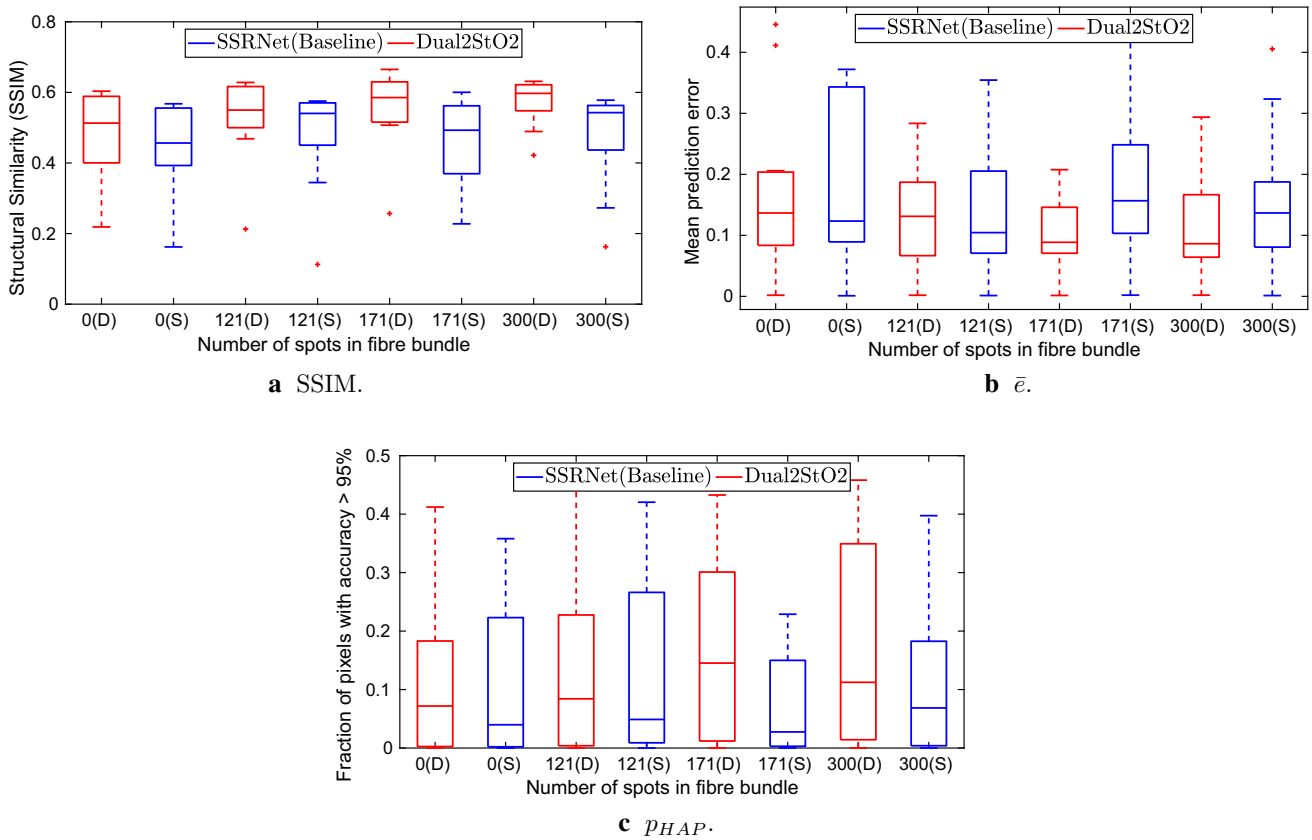
Results

Table 1 summarizes the performance of Dual2StO2 and SSRNet compared to the ground truth. The proposed network is superior to SSRNet in terms of SSIM and pixel-level accuracy across all fibre bundle configurations. When the fill factor (γ) is unchanged, the predicted images by Dual2StO2 are structurally closer to the ground truth (16.5% higher average SSIM), and have 3.6% lower average $\bar{e}$ than SSRNet for $n_{\text{spot}} = 300$. Figure 6a shows that even when the number of fibres increased in the bundle the Dual2StO2 predicted images are still structurally closer to the ground truth, with higher SSIM and less variance across different animals indicated by smaller interquartile range (IQR) at high SSIM value than SSRNet. Figure 6b, c also presents lower $\bar{e}$ and larger p_{HAP} by Dual2StO2 than that by SSRNet. Faster StO_2 estimation ($\approx$ 35 ms) can be achieved by Dual2StO2 due to its end-to-end estimation without the intermediate spectral estimation step and light-weight architecture, while the estimation required over 500 ms by SSRNet [8]. This was validated on a PC (OS: Ubuntu 16.04; processor: i7-3770; graphics card: NVIDIA GTX TITAN X).

A better StO_2 estimation was achieved with a higher number of fibres in the bundle with $n_{\text{spot}} = 300$ achieving the best result for both Dual2StO2 and SSRNet. The overall performance of StO_2 estimation was better with additional sHSI information than when compared to that from RGB images only. When sHSI was added, comparing $n_{\text{spot}} = 121$ to $n_{\text{spot}} = 0$, the structural similarity increased by 10% and the average mean error reduced by 2.3%. An experiment has also been carried out to estimate StO_2 using only sHSI. The results from Table 1 indicate that the single input network can estimate StO_2 and achieve pixel-level accuracy, evaluated by averaged mean prediction error, to some extent. However, the general structure similarity, evaluated by SSIM, is lower than that the dual-input network combined with RGB images.

Table 1 The average SSIM and average mean prediction error of StO_2 with standard deviation estimation by Dual2StO2, SSRNet, and single input network with different sHSI parameters, where the case that achieves best performance are hightlighted in bold

Fibre bundle setting					Network	Results	
n_{spot}	γ	r	d	Demo figure		Average SSIM	Average $\bar{e}$
300	0.25	2.6	10	Figure 5c	Dual2StO2	**0.63 ± 0.17**	**0.11 ± 0.09**
					SSRNet	0.54 ± 0.26	0.15 ± 0.12
					Only sHSI	0.54 ± 0.22	0.13 ± 0.11
171	0.25	3.5	14	Figure 5b	Dual2StO2	0.61 ± 0.19	0.12 ± 0.1
					SSRNet	0.52 ± 0.24	0.17 ± 0.12
					Only sHSI	0.54 ± 0.22	0.15 ± 0.10
121	0.25	4	16	Figure 5a	Dual2StO2	0.59 ± 0.21	0.14 ± 0.13
					SSRNet	0.54 ± 0.24	0.14 ± 0.10
					Only sHSI	0.53 ± 0.23	0.16 ± 0.10
0	0	0	–	–	Dual2StO2	0.53 ± 0.23	0.17 ± 0.14
					SSRNet	0.51 ± 0.24	0.18 ± 0.14



a SSIM.

b $\bar{e}$.

c p_{HAP}.

Fig. 6 Boxplot of averaged SSIM, averaged mean prediction error ($\bar{e}$) and p_{HAP} in different simulation experiments

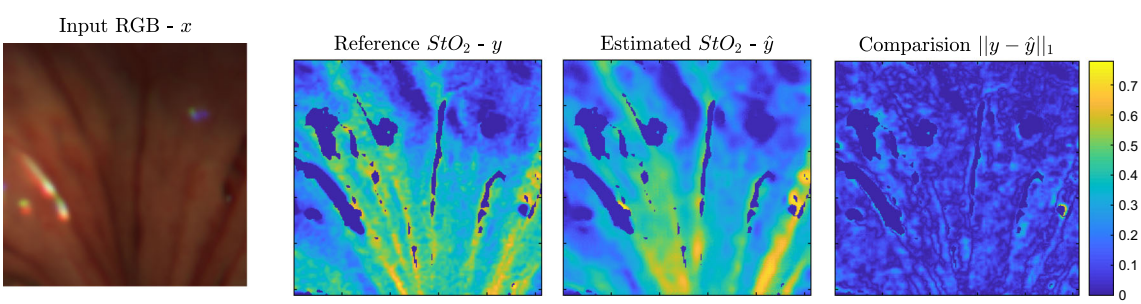


Fig. 7 Estimation of StO_2 based on Dual2StO2 with $n_{\text{spot}} = 300$, which achieved 0.61 in SSIM and 0.07 in $\bar{e}$ compared with the reference StO_2

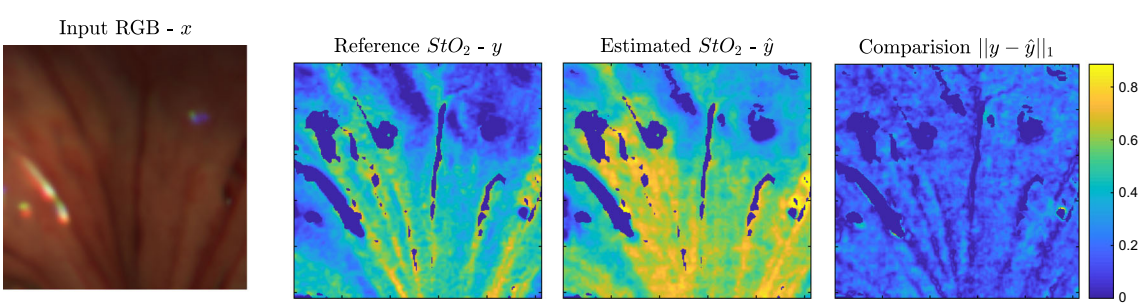


Fig. 8 Estimation of StO_2 based on SSRNet with $n_{\text{spot}} = 300$, which achieved 0.58 in SSIM and 0.15 in $\bar{e}$ compared with the reference StO_2

As the number of spots increased the performance of StO_2 estimation by single input network also improved.

Figures 7 and 8 illustrate the typical performance of StO_2 estimation by Dual2StO2 and SSRNet with the number of fibres in the bundle ($n_{spot} = 300$) on the second acquisition in animal ID 13 (porcine bowel). The input RGB image, reference StO_2 and estimated StO_2 by Dual2StO2 and SSRNet are displayed, while the StO_2 value difference between them is also presented. These demonstrate that, with an end-to-end learning training/testing architecture, Dual2StO2 outperforms the two-stage method, i.e. estimating StO_2 from hypercubes generated by SSRNet.

Discussion and conclusions

A dual-input network, called Dual2StO2, was designed to estimate StO_2 based on sHSI and RGB images. Simulations of three fibre bundles ($n_{spot} = 121, 171, 300$) and a control group ($n_{spot} = 0$) were carried out to investigate the impact of integrating sHSI, and to examine the relationship between the number of fibres and prediction accuracy. The results showed that with same fibre bundle, Dual2StO2 has better performance in StO_2 estimation (higher SSIM and lower $\bar{e}$ with smaller IQR, larger p_{HAP} and faster prediction) than SSRNet. Compared with the control group ($n_{spot} = 0$, using RGB data alone), the simulation results showed that the overall performance of StO_2 estimation with both Dual2StO2 and SSRNet was improved by adding sHSI. Performance was also improved as the number of fibres increased from 121 to 300, in terms of prediction accuracy and structural similarity. The result of the control group also indicated that StO_2 can be estimated directly from RGB although with consistently lower accuracy. This is in agreement with our previous works [9,10]. It was also observed that although RGB data could produce realistic spectral estimation, large errors at individual wavelengths were common. While StO2 estimation may be relatively insensitive to these underlying errors, spectral fidelity will be crucial to solving more subtle diagnostic problems such as the detection of cancer. This will be explored further in our future clinical work.

For real fibre bundles, the transmission characteristics of each fibre differ and cross-talk between fibres may result in measurement noise. This does not affect the result of the Dual2StO2 versus SSRNet comparison, but will affect the spatial accuracy of the sHSI-only results in Table 1 although it is unlikely to be significant. Furthermore, the sHSI presented here is simulated from an LCTF-based hyperspectral camera, which has lower spectral resolution (10–20 nm) than the spectrograph used in the real SLHSI system ($\approx$ 5 nm). Therefore, it is likely that the overall StO2 accuracy would be improved when trained with data from the real SLHSI bundle. Nevertheless, the simulations presented here serve as a useful testbed to allow comparative testing of network performance and to guide future design of an optimized fibre bundle. The network architecture of Dual2StO2 will be further customized for better performance, including exploration of custom-designed networks to extract features from RGB and sHSI images separately. The proposed dual-input network could potentially be modified to achieve dual output and generate, for example, narrow band images (NBI). The pyramid architecture of multi-generator and discriminator proposed by Wang et al. [27] could also be adopted to enhance the quality of generation. Our network can be further extended to real-time StO_2 imaging based on video-to-video synthesis [28].

Acknowledgements The authors appreciate the assistance of Xiao-Yun Zhou on the GPU configuration and academic advice, and help from Maria Leiloglou in the operation of SLHSI system. We would like to thank NVIDIA Corporation for the donation of the Titan X GPU. This work was carried out with support from the CRUK Imperial Centre, Imperial ECMC and the NIHR Imperial BRC.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants. All applicable international, national and/or institutional guidelines for the care and use of animals were followed.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Lu G, Fei B (2014) Medical hyperspectral imaging: a review. J Biomed Opt 19(1):010901
2. Mori M, Chiba T, Nakamizo A, Kumashiro R, Murata M, Akahoshi T, Tomikawa M, Kikkawa Y, Yoshimoto K, Mizoguchi M, Sasaki T, Hashizume M (2014) Intraoperative visualization of cerebral oxygenation using hyperspectral image data: a two-dimensional mapping method. Int J Comput Assist Radiol Surg 9(6):1059–1072
3. Clancy NT, Saso S, Stoyanov D, Sauvage V, Corless DJ, Boyd M, Noakes DE, Thum MY, Ghaem-Maghami S, Smith JR, Elson DS (2016) Multispectral imaging of organ viability during uterine transplantation surgery in rabbits and sheep. J Biomed Opt 21(10):106006. https://doi.org/10.1117/1.JBO.21.10.106006
4. Akbari H, Kosugi Y, Kojima K, Tanaka N (2010) Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging. IEEE Trans Biomed Eng 57(8):2011–2017
5. Kumashiro R, Konishi K, Chiba T, Akahoshi T, Nakamura S, Murata M, Tomikawa M, Matsumoto T, Maehara Y, Hashizume M (2016) Integrated endoscopic system based on optical imaging and hyperspectral data analysis for colorectal cancer detection. Anticancer Res 36(8):3925–3932

6. Wang LV, Hu S (2012) Photoacoustic tomography: in vivo imaging from organelles to organs. Science 335(6075):1458–1462
7. Clancy NT, Stoyanov D, Maier-Hein L, Groch A, Yang GZ, Elson DS (2011) Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging. Biomed Opt Express 2(11):3119–3128
8. Lin J, Clancy NT, Qi J, Hu Y, Tatla T, Stoyanov D, Hein LM, Elson DS (2018) Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. Med Image Anal 48:162–176
9. Li QB, Zhou XY, Lin J, Zheng JQ, Clancy NT, Elson DS (2018) Estimation of tissue oxygen saturation from RGB images based on pixel-level image translation. In: Hamlyn symposium on medical robotics
10. Lin J, Clancy NT, Hu Y, Qi J, Tatla T, Stoyanov D, Maier-Hein L, Elson DS (2017) Endoscopic depth measurement and super-spectral-resolution imaging. In: Medical image computing and computer-assisted intervention (MICCAI) 2011. Springer, Berlin, pp 39–47
11. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 105–114
12. Yi X, Walia E, Babyn P (2018) Generative adversarial network in medical imaging: a review. arXiv preprint arXiv:1809.07294
13. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784
14. Isola P, Zhu J, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 5967–5976. https://doi.org/10.1109/CVPR.2017.632
15. Clancy NT, Arya S, Stoyanov D, Singh M, Hanna GB, Elson DS (2015) Intraoperative measurement of bowel oxygen saturation using a multispectral imaging laparoscope. Biomed Opt Express 6(10):4179–4190
16. Clancy NT, Stoyanov D, James DR, Di Marco A, Sauvage V, Clark J, Yang GZ, Elson DS (2012) Multispectral image alignment using a three channel endoscope in vivo during minimally invasive surgery. Biomed Opt Express 3(10):2567–2578
17. Sorg BS, Moeller BJ, Donovan O, Cao Y, Dewhirst MW (2005) Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development. J Biomed Opt 10(4):044004–044004-11
18. Tjur T (2009) Coefficients of determination in logistic regression models a new proposal: the coefficient of discrimination. Am Stat 63(4):366–372
19. Kyrish M, Kester R, Richards-Kortum R, Tkaczyk T (2010) Improving spatial resolution of a fiber bundle optical biopsy system. In: Endoscopic microscopy V, SPIE, vol 7558, p 755807
20. Shao J, Liao WC, Liang R, Barnard K (2018) Resolution enhancement for fiber bundle imaging using maximum a posteriori estimation. Opt Lett 43(8):1906–1909
21. Perera P, Abavisani M, Patel VM (2018) I2I: unsupervised multi-image-to-image translation using generative adversarial networks. In: 24th international conference on pattern recognition, ICPR 2018
22. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/CVPR.2016.90
23. Ulyanov D, Vedaldi A, Lempitsky VS (2016) Instance normalization: the missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022
24. Zhou X, Yang G (2019) Normalization in training U-Net for 2-D biomedical semantic segmentation. IEEE Robot Autom Lett 4(2):1792–1799. https://doi.org/10.1109/LRA.2019.2896518
25. Li C, Wand M (2016) Precomputed real-time texture synthesis with Markovian generative adversarial networks. In: European conference on computer vision ECCV 2016. Springer, Cham, pp 702–716
26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
27. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B (2018) High-resolution image synthesis and semantic manipulation with conditional gans. In: 2018 IEEE conference on computer vision and pattern recognition (CVPR)
28. Wang TC, Liu MY, Zhu JY, Liu G, Tao A, Kautz J, Catanzaro B (2018) Video-to-video synthesis. In: Advances in neural information processing systems (NeurIPS)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.