Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks

Jianyu Lin [a,b,*], Neil T. Clancy [a,c,d,e,f,*], Ji Qi [a,f], Yang Hu [a,b], Taran Tatla [g], Danail Stoyanov [c,d,e], Lena Maier-Hein [h], Daniel S. Elson [a,f,*]

[a] The Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK [b] Department of Computing, Imperial College London, London, UK [c] Wellcome/EPSRC Centre for Interventional & Surgical Sciences (WEISS), University College London, London, UK [d] Centre for Medical Image Computing, University College London, London, UK [e] Department of Computer Science, University College London, London, UK [f] Department of Surgery and Cancer, Imperial College London, London, UK [g] Department of Otolaryngology, Northwick Park Hospital, Harrow, UK [h] Division of Medical and Biological Informatics, German Cancer Research Center, Heidelberg, Germany

articleinfo

Article history:
Received 8 February 2018
Revised 9 May 2018
Accepted 7 June 2018
Available online 15 June 2018

Keywords:
Intra-operative imaging
3D reconstruction
Structured light
Hyperspectral imaging
Deep learning
Super-spectral-resolution

abstract

Surgical guidance and decision making could be improved with accurate and real-time measurement of intra-operative data including shape and spectral information of the tissue surface. In this work, a dual- modality endoscopic system has been proposed to enable tissue surface shape reconstruction and hy- perspectral imaging (HSI). This system centers around a probe comprised of an incoherent fiber bundle, whose fiber arrangement is different at the two ends, and miniature imaging optics. For 3D reconstruc- tion with structured light (SL), a light pattern formed of randomly distributed spots with different colors is projected onto the tissue surface, creating artificial texture. Pattern decoding with a Convolutional Neu- ral Network (CNN) model and a customized feature descriptor enables real-time 3D surface reconstruc- tion at approximately 12 frames per second (FPS). In HSI mode, spatially sparse hyperspectral signals from the tissue surface can be captured with a slit hyperspectral imager in a single snapshot. A CNN based super-resolution model, namely "super-spectral-resolution" network (SSRNet), has also been devel- oped to estimate pixel-level dense hypercubes from the endoscope cameras standard RGB images and the sparse hyperspectral signals, at approximately 2 FPS. The probe, with a 2.1 mm diameter, enables the sys- tem to be used with endoscope working channels. Furthermore, since data acquisition in both modes can be accomplished in one snapshot, operation of this system in clinical applications is minimally affected by tissue surface movement and deformation. The whole apparatus has been validated on phantoms and tissue (*ex vivo* and *in vivo*), while initial measurements on patients during laryngeal surgery show its potential in real-world clinical applications.

1. Introduction

In the past several decades, the development of technology has boosted the emergence of new surgical approaches. Minimal ac- cess surgery (MAS), for example, has limited trauma to the pa- tient by using smaller incisions resulting in reduced blood loss, less pain, fewer infections, quicker recovery and improved quality of life (Velanovich, 20 0 0; Darzi and Mackay, 2002). However, MAS also brings challenges to surgeons, such as the lack of tactile feedback, reduced depth perception, limited dexterity of sur- gical instruments and difficult hand-eye coordination, all of which contribute to a steeper learning curve. Providing specific additional intra-operative information is expected to help surgeons with clini- cal decision-making and surgical navigation. In this work, we focus on combining optical approaches and algorithmic implementations to obtain image data on the shape and function of the tissue.

Accurate pre- and intra-operative imaging of patient anatomy is essential. Typically, pre-operative image data is volumetric and used to plan a procedure, whereas during MAS, imaging is via 2D visualization enabled by monocular endoscopes. Surgeons there- fore lose a sense of depth when surveying the surgical scene, which leads to instrument mis-location and unwanted tissue in- teraction, and further results in steeper learning curves for effec- tive instrument manipulation. In robotic surgery, the application of active constraints on the manipulator end-effector also depend on accurate tissue surface shape measurement (Davies et al., 2006). Tissue surface shape also carries important information and is in- dicative of certain pathologies, for example colonic polyps, which vary in size and shape from pedunculated, to flat (Obuch et al., 2015). Furthermore, tissue surface shape has potential to be used in MAS to facilitate navigation by connecting the endoscopic view with the pre-operative image data. For example, augmented reality (AR) can be used to visualize the extent of a tumor and plan its resection using the displayed endoscopic surgical scene and pre-operative MRI (Pratt et al., 2012). It may also be used to allevi- ate the drawback of lack of direct view and tactile feedback during MAS (Bernhardt et al., 2017).

At the same time there is increasing current interest in us- ing additional optical imaging modalities during MAS to assist in disease detection and diagnosis, for instance using near in- frared fluorescence (Nagengast et al., 2017), fluorescence life- time (Marcu et al., 2014) or optical coherence tomography (Boppart et al., 2004). One of the techniques we and others have investigated, called multi/hyperspectral imaging (MSI/HSI), also provides important information for image-guided surgery. MSI/HSI aims to acquire

* Corresponding author.
E-mail addresses: xjtuljy@gmail.com (J. Lin), n.clancy@ucl.ac.uk (N.T. Clancy), daniel.elson@imperial.ac.uk (D.S. Elson).

https://doi.org/10.1016/j.media.2018.06.004 1361-8415/

optical reflectance spectra from all locations in a scene, resulting in a 3D ($x-y-\lambda$) hypercube, with a much higher number of narrow bands (10s to 100s) in the spectral di- mension than RGB images (Wolfe, 1997). The shape of the measured reflectance spectrum is influenced by the optical properties of the tissue, including the concentration of absorbers, such as hemoglobin, and scatterers, such as cells or structural connective tissues. The morphology of cells, and hence their scattering prop- erties, can be altered during disease, for instance, nuclear enlarge- ment is a marker for cancer (Ferris et al., 2001). Due to the rich spectral information it carries, MSI/HSI has been investigated to aid medical applications such as clinical diagnostics (Ferris et al., 2001), surgical guidance (Clancy et al., 2015), and pathology detec- tion (Lu and Fei, 2014). MSI and HSI differ mainly in the number of wavelength bands, and they are both referred to as HSI in the following context for convenience.

We have investigated an approach that allows 3D surface shape measurement and HSI with an endoscope system. The main chal- lenges addressed in this work are: 1) the appending of devices to the endoscope system to enable fast 3D reconstruction and HSI; 2) handling moving and deforming tissue in surgical scenes; 3) the spatial density of acquired intra-operative data; 4) AR that jointly displays intra-operative information on the surgical scene. The pro- posed apparatus has the advantages of being compatible with standard endoscope working channels (due to its small diameter) and having the ability to implement multiple modalities on the same platform. The new system was aided by algorithmic development for dense hypercube estimation from sparse hyperspectral signals and RGB images to permit complementary clinical studies of the separate and combined SL and HSI techniques for different surgi- cal applications.

1.1. Related work

In order to address the challenges mentioned above, we provide a brief introduction to SL based 3D reconstruction, HSI in medical applications, and the state-of-the-art super-resolution techniques.

1.1.1. Intra-operative 3D reconstruction with SL

One of the main challenges in reconstructing tissue surfaces in surgical scenes is the deformation and movement of tissue which requires fast single frame based 3D reconstruction. These require- ments can be potentially met by stereoscopy, Time-of-Flight (ToF), and SL, although stereoscopy depends on detecting tissue surface texture information and ToF suffers from systematic errors that cannot always be compensated by calibration (Maier-Hein et al., 2013). We have adopted SL as the 3D reconstruction technique since it overcomes these limitations, can be adapted to monocu- lar endoscopes and uses single shot based reconstruction that can handle moving and deforming tissues.

Conceptually similar to a two-camera passive stereo system, a typical SL system replaces one of the cameras with a projector, which projects a specific pattern of light on the target object. The reflectance is captured by a camera and the object surface can then be reconstructed using triangulation according to correspondences between the known (calibrated) ray projection of the customized light pattern and their positions in the captured image.

SL techniques can be divided into two categories, depending on how the correspondence problem is solved: multi-shot (sequential) and single-shot (Salvi et al., 2010). In multi-shot a sequence of light patterns is projected onto the object, forming multiple different codewords for specific locations on the surface. However, single- shot approaches, which project an unchanged pattern and use intensity features, are more common for clinical applications where the soft tissue or endoscope are continuously moving or deform- ing. In an early example, Hasegawa et al. developed a flexible fiber- based projector, using a spatial coding strategy of seven patterns to reconstruct the target surface (Hasegawa et al., 2002). Chan et al. projected a dot matrix light pattern through one of the channels in a dual-channel rigid endoscope, which was then validated on both reflective model surfaces, and tissues from the forearm and oral cavity (Chan et al., 2003). An SL-enabled colposcope was de- veloped by Wu and Qu, incorporating a grid light pattern for track- ing and correcting the motion of the cervix surface *in vivo* (Wu and Qu, 2007). Other notable examples include the laparoscopic use of the M-array algorithm during open surgeries (Maurice et al., 2012), capsule-based reconstruction of tubular cavities such as the tra- chea *ex vivo* (Schmalz et al., 2012), and the Pico Lantern miniatur- ized “pick-up” laser projector that can be passed through a trocar and held by an endoscopic grasper (Edgcumbe et al., 2015).

The ICL SL system uses an incoherent fiber bundle to create and transmit a pattern, which allows flexibility as well as compatibility with endoscope working channels and trocars (Clancy et al., 2011). The random fiber arrangement within the bundle and the imag- ing of the pattern by the GRIN lens results in a multi-colored spot pattern projected on the target object surface. Each spot contains a unique narrow spectrum, which is insensitive to modulation by the object color.

By analyzing the projected SL images via CNN models for real- time spot detection and local rigid registration-based spot identi- fication, the 3D locations of the projected spots were calculated, enabling 3D surface shape reconstruction (Lin et al., 2016). Valida- tion with phantom and *ex vivo* experiments showed that the ICL SL system was capable of reconstructing tissue surface with ≈ 0.7 mm average error at a working distance of ≈ 100 mm (Lin et al., 2015a).

Improvements to the ICL SL approach are required to compen- sate for some drawbacks which hinder its real-world application, such as: 1) lack of a white light (WL) view for normal color intra- operative visualization; 2) changeable relative pose between the probe and the laparoscope that might introduce reconstruction errors. These are addressed in the new ICL SLHSI system as described in Section 2.1 .

1.1.2. HSI in medical applications

Medical HSI measures the spectral reflectance of target tissue at all locations within the field-of-view (FoV) of the camera, ac- quiring a three-dimensional dataset known as a “hypercube” con- taining one wavelength and two spatial dimensions. It has been used to detect disease-specific changes in tissue optical proper- ties arising from structural changes within the tissue in the colon (Kumashiro et al., 2016), while macroscopic changes due to burn damage have been detected in skin (King et al., 2015). Ferris et al. proposed a HSI imaging system to capture both the reflectance and fluorescence from the cervical epithelium on the ectocervix, on a diverse population of women, proving the feasibility of ap- plying HSI in discriminating high grade cervical lesions, less severe lesions, and normal cervical tissue (Ferris et al., 2001). Several HSI systems have exploited the differing optical absorption spectra of oxy and deoxyhemoglobin to calculate tissue oxygen saturation (StO_2), based on hardware (Zuzak et al., 2011; Clancy et al., 2015) or software approaches (Wirkert et al., 2016; 2017; Jones et al., 2017).

According to the image acquisition mode, most HSI systems can be divided into three types: spatial scanning (Aiazzi et al., 2006), spectral scanning (Luo et al., 2014), and snapshot (Weitzel et al., 1996). HSI with spatial scanning refers to systems which generate the hypercube by sensing the whole spectrum at a single point or line, and scanning across the desired field of view. Spectral scan- ning captures an image of the whole scene at a certain wavelength, then scans this wavelength using a series of optical filters or tun- able light source. Snapshot methods, such as the one used in this paper, record the entire hypercube in a single acquisition. There is no “gold standard” HSI architecture as trade-offs between spa- tial/spectral resolution and acquisition speed depend on the appli- cation. Our system combines high resolution spectral data in a sin- gle snapshot, using computational techniques to increase the spa- tial resolution.

1.1.3. The super-resolution algorithms

In this work, we aim to recover pixel-level dense hypercubes from RGB images and spatially sparse hyperspectral signals. This proposed approach was enlightened by state-of-the-art super- resolution (SR) algorithms, which recover high-resolution (HR) data from low-resolution (LR) data either in the spatial or spectral di- mension.

The first milestone end-to-end CNN model used for SR prob- lems was the SRCNN model (Dong et al., 2016), a three-layer CNN model that simulates a sparse coding procedure to refine the de- tails in the upscaled image with bicubic interpolation. A deeper recursive neural network with skip connection (DRCNN) increased the receptive field to [41 × 41] (Kim et al., 2016b). In both SR- CNN and DRCNN, the input LR image needed to be upscaled to the size of the desired output HR image via bicubic interpola- tion before being fed into the network, hindering the process- ing speed. An efficient sub-pixel convolutional neural network (ES- PCN) was proposed for real-time SR (Shi et al., 2016), process- ing the input LR image with consecutive convolutional layers in LR scale, and upscaling the feature map only in the last layer. In the last upscaling step, a sub-pixel convolution was applied, which is in fact a reshaping procedure rather than real "convolu- tion". The problems of prone-gradient vanishing/explosion in deep neural networks (DNNs) was addressed by Kim et al. (2016a) by adopting residual blocks with skip-connections (He et al., 2015). This model has also been validated on SR problems with differ- ent scales, demonstrating its robustness. The same strategy has been adopted by Oktay et al. to solve MRI cardiac image SR prob- lems (Oktay et al., 2016). Generative adversarial networks (GANs) (Goodfellow et al., 2014) have also been adopted for the purpose of image generation to replace those pixel-wise loss functions, for instance the super-resolution generative adversarial network (SR- GAN) is a model proposed to solve the SR problems with GANs (Ledig et al., 2017). Instead of achieving high peak-to-signal-ratio (PSNR) and structural similarity (SSIM) index, this model aimed to produce more "photo-realistic" results.

Some recent studies have focused on estimating hypercubes from normal RGB images algorithmically, which can be considered as SR in the spectral domain. Arad and Ben-Shahar proposed a sparse coding based method to predict the hypercubes (Arad and Ben-Shahar, 2016), where the linear combination of hyperspectral basis can be found by searching for the counterpart for RGB ba- sis with orthogonal match pursuit (OMP) (Pati et al., 1993). Similar studies on recovering HR hypercubes from LR hypercubes and HR panchromatic (PAN) images were also popular in remote sensing (Loncan et al., 2015). More recently, CNN models have also been adopted for pan-sharpening (Zhong et al., 2016; Masi et al., 2016).

Inspired by state-of-the-art for SR approaches, we built our own model, the super-spectral-resolution network (SSRNet), to recover dense hypercubes from RGB images and spatially sparse hyper- spectral signals, as demonstrated in Section 2.3 .

1.2. Scope of work

In this paper, in order to accomplish practical intra-operative imaging during surgery based on the ICL SL system (Clancy et al., 2011; Lin et al., 2015a; 2015b; 2016; 2017), we propose a hybrid endoscope apparatus (the ICL SLHSI system), as shown in Fig. 1 , which

- integrates surface reconstruction and HSI into the endoscopic system;
- has a miniaturized dimension compatible with normal endo- scopic biopsy channels;
- is less affected by tissue deformation and moving due to the snapshot based data acquisition;
- enables real-time 3D reconstruction (12 FPS) with deep learn- ing and customized feature descriptors;
- uses SSRNet to measure pixel-level dense multispectral hyper- cubes (24 wavelength bands) from RGB images and capture spatially sparse hyperspectral signals;
- combines surface shape and hyperspectral information for AR that shows narrow-band images (NBI) or tissue oxygen satura- tion (StO_2) on reconstructed tissue surface.

The whole system has been validated in phantom experiments, on *ex vivo* / *in vivo* tissue samples and in preclinical animal proce- dures. The system has also been trialed *in vivo* during a human head-and-neck surgical case, demonstrating its potential clinical utility and compatibility with standard clinical workflows. Com- pared with the previous conference papers (Lin et al., 2016; 2017), this paper provides more details on the ICL SLHSI system, as well as further improvement to the SSRNet and significant extensions in the experimental validation.

2. Materials and methods

This section begins with a detailed description of the ICL SLHSI system. Then the algorithmic implementations in the SL and HSI modes are illustrated, which mainly focus on the 3D reconstruction and dense hypercube estimation.

2.1. The ICL SLHSI system

The proposed optical hardware in this work, namely the ICL SLHSI system, was extended from the ICL SL system but now op- erating in two different modes (SL and HSI), as shown in Fig. 2 . The assembly is based around a bundle of 171 fibers arranged in a

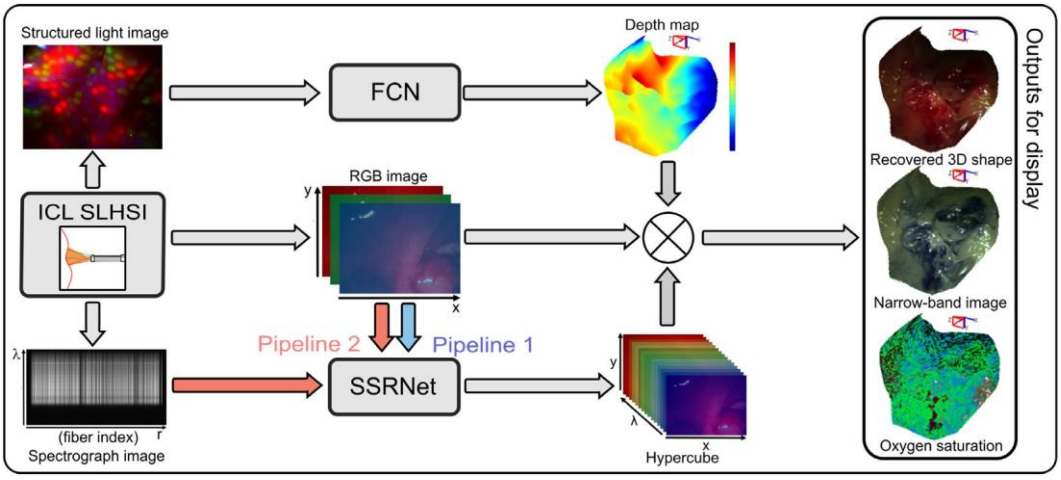

Fig. 1. The graphical abstract of this work. The ICL SLHSI optical system captures SL images, RGB images, and spectrograph images from the target tissue surface. The SL images are used to recover the depth maps of the tissue surfaces with a fully convolutional network (FCN); the spectrographs and the RGB images were jointly processed by the super-spectral-resolution network (SSRNet) to generate pixel-level dense hypercubes. For the purpose of AR, this system is able to combine the depth maps and hypercubes to provide surgeons with visualization of the recovered 3D surfaces, NB images, and the StO_2 maps. Sample images shown were acquired from the *in vivo* human larynx.

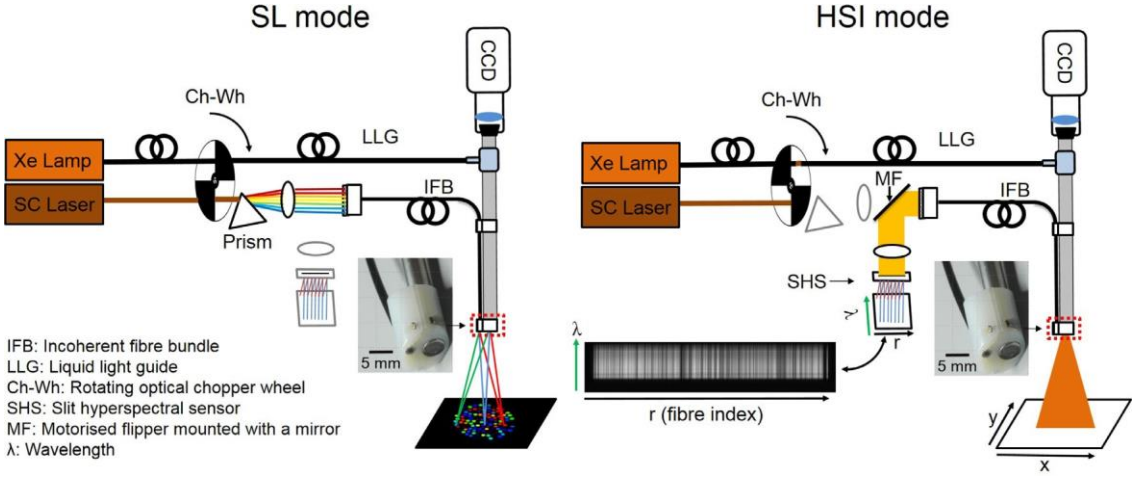

Fig. 2. Left: setup in SL mode. Supercontinuum laser is used to generate the structured light, and the white light is blocked. Right: setup in HSI mode and the captured spectrograph image (lower right). Supercontinuum laser is blocked and white light is shed on the tissue surface. The mirror mounted on a flipper is used to redirect the reflectance to the slit hyperspectral imager. Figure reproduced from Lin et al. (2017) .

linear array at one end, and a circular array at the other with the relative fiber positions randomized. A GRIN imaging lens was at- tached to the circular end and housed in a ferrule (maximum outer diameter of 2.1 mm), with the whole assembly – referred to here as the “probe” – packaged in a protective outer jacket. A 3D printed adapter (11 mm diameter) allowed the SL probe to be mounted at a fixed position relative to the tip of the rigid endoscope (5 mm di- ameter Hopkins II Optik 30 °, Karl Storz GmbH, Germany). Accord- ing to the FoV of the probe (≈ 65 °) and the endoscope (≈ 70 °), the angle and axial displacement (≈ 12 ° angle, 3 mm) of these chan- nels for the endoscope and probe were designed to maximize the overlapping imaging area. This enabled the SL system to recon- struct tissues surfaces within approximately 1 . 5 − 4 cm working distances, with a baseline ≈ 6 mm. The small size of the adaptor enables insertion of the probe and endoscope tip through a typical surgical access port.

A mirror mounted on a motorized flipper (MFF101/M, Thor- labs Ltd., UK) was positioned between the lens and the linear ar- ray end of the fiber bundle and was used to switch between the two modes. In SL mode (Fig. 2 (left)), the flipper mirror was removed from the light propagation path, allowing the visible light (420 − 750 nm) from a 4 W supercontinuum laser (SC400-4, Fia- nium Ltd., Southampton, UK) dispersed by an SF-11 prism to be focused onto the probe s linear array end such that each fiber car- ries light of a unique narrow spectral band. Rapid stroboscopic WL and SL acquisition was realized using a chopper wheel (3501 Op- tical Chopper; New Focus, Inc., USA) placed concurrently in the laser beam path and in a 2 mm gap between two light cables car- rying the white illumination light such that supercontinuum and white light were alternately blocked and transmitted out of phase. A computer-controlled signal generator (NI USB-6211; National In- struments Corporation, USA) was used to trigger the CCD camera (DCU 223C, Thorlabs Ltd., UK; resolution: 1024 × 768) and rotation of the chopper wheel so that WL and SL images were both cap- tured at a frequency of 8 FPS. The WL images were provided to the surgeons primarily for normal visualization but could be used for other purposes like tissue tracking or Structure-from-Motion (SfM) in future work.

Adding the flipper mirror enabled the snapshot HSI mode (Fig. 2 (right)) by allowing white light that had been reflected by the sample then collected by the SL probe to be reflected off the 45° mirror and directed towards an HSI detector. The linear fiber array end of the SL probe was then demagnified and focused, using a 250/50 mm focal length lens combination, onto the slit of a hyperspectral imager (Nano-Hyperspec; Headwall Photonics, Inc., USA). This slit hyperspectral imager was able to capture a spectrograph containing 640 spatial bands and 270 spectral bands (400 − 1000 nm). In the spectral dimension the HSI cameras dis- persive elements achieve 2.2 pixels per nm at a spectral resolution of 6 nm (full-

width at half-maximum). The imaging spectrograph was calibrated by manually finding the column numbers corre- sponding to each fiber in the probe. Therefore light reflected by different locations on the tissue surface was mapped to different positions on the input slit of the spectrometer, giving an $r-\lambda$ im- age (Fig. 2 (right)) that could, given the appropriate calibration, be remapped to an $x-y-\lambda$ cube in a single shot. This mapping pro- cedure used a one-off calibration to search for the correspondence between the fibers at the two ends of the bundle, which enabled a further mapping between the fiber indices in the spectrometer to individual spot locations in the SL mode, and hence the pre- requisite for AR that jointly displayed reconstructed surface with information extracted from the hyperspectral signals. During cali- bration, WL was coupled into the linear array end resulting in a spot pattern image on a flat screen at the distal end. A piece of card mounted on a translation stage was used to obscure all of the fibers in the linear array then sequentially uncover them. The emergence of the projected spots on the screen was recorded with a camera and used to identify their spatial locations.

The absorbance spectra (A) were calibrated and calculated to correct for wavelength-dependent system transmission by record- ing reference spectra from ambient light (the "dark spectrum" $I_{(d)}$) and a white reference target (the reflectance $I_{(0)}$, Spectralon; Lab- sphere, Inc., USA). As the absorbance A can be interpreted as a lin- ear combination of the absorption spectra from oxy and deoxyhe- moglobin (Sorg et al., 2005) with a constant offset accounting for losses due to scattering, the StO_2 could be estimated by linear fit- ting (Clancy et al., 2015).

2.2. 3D reconstruction pipeline in SL mode

The algorithmic pipeline of 3D reconstruction using the ICL SLHSI system can be divided into three steps: 1) system calibra- tion, which finds the intrinsic parameters of, and the relative pose between, the camera and the probe; 2) pattern decoding, which analyzes the pattern projected on the tissue to find the corre- spondences between the probe image plane and the camera image plane, referred to as "pattern decoding"; 3) triangulation, which measures the tissue surface shape based on the first two steps (Lin et al., 2015a; 2015b). In this paper we focus on illustrating the cur- rent pattern decoding algorithm, whose details have been updated compared with the previous work (Lin et al., 2016).

SL pattern decoding is the most important and challenging step for 3D reconstruction in terms of computational speed, accuracy, and robustness. The aim of pattern decoding in the ICL SLHSI sys- tem is to detect and recognize the projected spots from the cap- tured SL images, according to their color and neighborhood infor- mation so that they may be matched and triangulated using infor- mation from a reference SL image (one that shows all spots on a clean white surface). The proposed pattern decoding algorithm was divided into two steps: spot detection and spot identification. The former requires fast and accurate spot segmentation to precisely estimate spot centers, while the latter compares the reference SL image with the current SL image to find the matches.

2.2.1. Spot detection

In surgical environments spot detection is the bottleneck for accurate surface reconstruction due to the strong light-tissue interactions, appearance of mucus, and large curvature of the tissue surface in some surgical scenes. The first step for spot de- tection was image pre-processing for specular highlight removal and image smoothing, followed by a linear interpolation based im- age scaling operation to resize the images to [512 × 384] (width, height). In this work we consider spot detection as a binary im- age segmentation problem. While the spots in SL images pro- duced by our system are generally uniform in morphology, but with color differences, a FCN model proposed in Lin et al. (2016) , was adopted to detect the spot in this system. This model fol- lows the encoder-decoder structure (Long et al., 2015; Shelhamer et al., 2017; Badrinarayanan et al., 2017), with U-Net style short- cuts (Ronneberger et al., 2015) for feature map mergence between the encoder and decoder, as shown in Fig. 3 (a). The encoder con- sisted of five consecutive units with the same layer combination, and each halved the input image size while doubled the channel dimension. All the convolutional kernels had a size of [3 × 3], so that in the last layer of the encoder, each neuron had an effective receptive field of 46 × 46 pixels. The decoder also contained five consecutive units, where each upsampled the image and fused the upsampled feature maps and the counterpart from the encoder. The output feature map was cropped to the size of the input im- age, resulting in a 2-channel image, where each channel indicated the "probabilities" of being foreground (spots) or background. The number of layers was chosen empirically by considering the trade- off between computational speed and accuracy.

200 SL images, captured from phantom, *ex vivo* , and *in vivo* ex- periments, were manually annotated using the software "ilastic" (Sommer et al., 2011) to be used as the segmentation ground truth. The number of visible spots ranged from about 50–171, depending on the image quality and object surface curvature; the spot sizes differed as images were captured at different working distances; examples on tissue surfaces with blood and mucosa were also used to increase the variability of the training set. Data augmentation was applied including resizing in three scales, horizontal and ver- tical flipping, rotations, and affine transformations. Xavier initial- ization was applied to all the weights before training. For training a loss function consisting of cross-entropy loss and L2-regulariser was used, with Stochastic Gradient Descent (SGD) solver. The train- ing included a user-interactive coarse-to-fine tuning, with batch size 1, decay weight 0.0 0 05, and momentum 0.9. The learning rate was gradually changed from 0.01 to 0.0 0 05, in 50 0,0 0 0 iterations. The final loss converged at about 0.006.

In prediction, firstly the output mask of the "arguments of the maxima" (argmax) layer was used for background removal, then a "probability" map was obtained by the subtraction between the foreground and background channels. Next, the "probability" val- ues in individual isolated connected components of the mask were normalized to the range $[0-255]$. Finally the local maxima of the map was extracted via adaptive thresholding, and their centers of mass were treated as spot centers.

2.2.2. Fast feature matching with epipolar constraints

Based on the spot detection results, correspondences between the captured and reference SL images were found using a cus- tomized feature descriptor and epipolar constraints. An algorithm based on Delaunay triangulation with false connection pruning was applied to establish the neighborhood of detected spots, which were each allocated a customized $32 \times 32 \times 3$ feature descriptor to indicate the color distribution in 32 directions at 32 orientations (Fig. 3 (b)). The 1st dimension was the index of the starting direc- tion (orientation), the 2nd indicated the direction and the 3rd in- dicated the average color (normalized red, green, and blue values).

For every spot in the reference image, candidate matched spots in the captured image were firstly found using the epipolar con- straints, according to the distance from the spot centers to the epipolar line. The smallest distance between a spot descriptor on the captured image and all 32 descriptors of a spot on the refer- ence image was used to describe the distance between them. The match with closest distance smaller than a threshold was chosen, followed by a pruning procedure based on neighborhood informa- tion. An iterative method was then applied to propagate matches to other neighboring unmatched spots, followed by pruning, un- til the number of matches remained constant. After spot matching (Fig. 3 (c)), the target object surface shape was recovered by tri- angulation using calibrated parameters (Lin et al., 2015a). The 3D reconstruction pipeline was implemented in C ++ with external li- braries including OpenCV (Itseez, 2015), Caffe (Jia et al., 2014), and VTK (Schroeder et al., 2006). The computation time for reconstruc- tion from single SL image was ≈ 80 ms on a PC (OS: Ubuntu 14.04; processor: i7-3770; graphics card: NVIDIA GTX TITAN X).

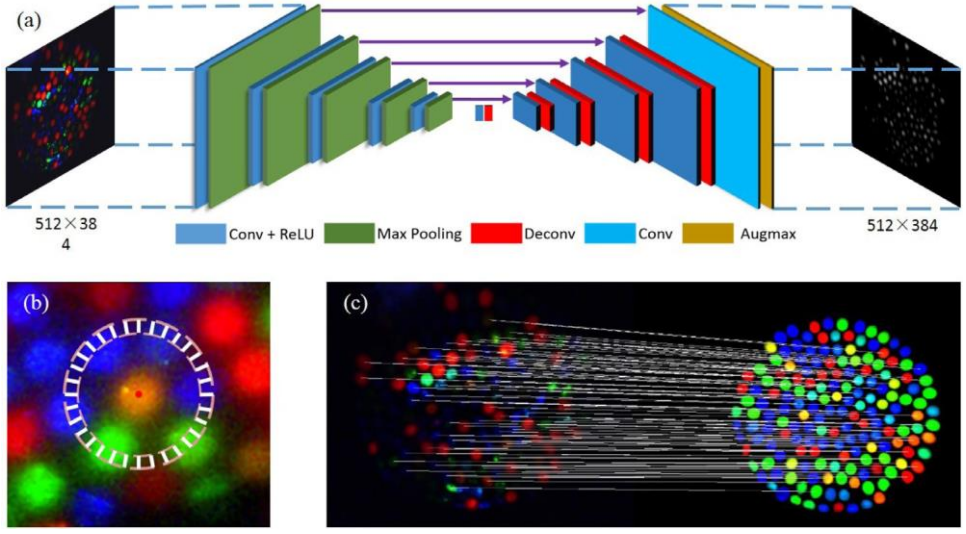

Fig. 3. (a) The deployed FCN model. (b) The working space of the feature descriptor marked by white segments. The normalized ratios between red, green, and blue in each segment were used as elements in the feature vector. (c) Spot (feature) matching between the captured (left) and the reference SL image (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. SSRNet for super-spectral-resolution

Due to the inherent limitations associated with acquiring a 3D dataset (the hypercube) on a 2D sensor (camera), the spatial reso- lution in the ICL SLHSI system has been sacrificed for snapshot ac- quisition, where only spatially sparse spectra could be directly col- lected in a snapshot. Although this still provided medically mean- ingful information that covered the FoV, the majority of the area was still left blank between the spot locations. In real-world ap- plications a hypercube with pixel-level dense spatial resolution is preferred. Therefore, we propose a CNN model for dense hypercube generation, which is referred to as SSRNet. The model described in this paper is named SSRNet_v2, which was extended and much improved over the first version SSRNet_v1 (Lin et al., 2017). There are two data sources from which the hypercube could be recov- ered: the RGB image captured by the CCD camera, and the spatially sparse spectral signals captured by the slit hyperspectral imager.

SR is an ill-posed problem, since it aims to recover HR im- ages from their LR counterparts, where multiple possible solutions exist. To deal with this we made three assumptions to constrain the search space: 1) HR images contain redundant HR information which could be partially extracted from the LR counterparts; 2) the mapping from LR to HR can be learnt from training sets contain- ing data similar to the unseen data; 3) the transmission spectra of the RGB camera is known. In general, the hypercube should be recovered mainly from the corresponding RGB image and partially from the sparse hyperspectral signals, since the former are spatially dense while the latter spatially sparse. Therefore, SSRNet_v2 can be divided into two steps, i.e. hypercube estimation using an RGB image, and hypercube refinement by integrating sparse hyper- spectral signals.

2.3.1. Recover hypercubes from RGB images (Pipeline 1)

An RGB image can be seen as a special hypercube with three spectral bands. In Pipeline 1, the proposed model, SSR- Net_v2, searched for a mapping from an $M \times N \times 3$ hypercube to an $M \times N \times 24$ one, where M and N indicated the width and height of the hypercube. The proposed Pipeline 1 consisted of two stages:

1. RGB image upscaling along the spectral dimension. This was re- alized by three consecutive units (Fig. 4 (c)), each consisting of a deconvolution layer (transposed convolutional layers) followed by a residual block (He et al., 2015). The residual block provided a shortcut connection between lower and higher layers, helping to avoid gradient vanishing problems for deep neural networks. Each of these units doubled the number of the spectral bands, hence resulting in a mapping of the hypercube from a size of $M \times N \times 3$ to $M \times N \times 24$ with three such units.
2. High frequency signal extraction. This extracted the high fre- quency signal and then added this to the LR data, realized by a residual-like block. In contrast to a normal residual block the “shortcut” in our model contained one instead of zero convo- lutional layer. With this structure, we expected a mergence of the low-frequency and high-frequency data to improve the pre- dicted hypercube.

All the convolutional kernels in Pipeline 1 were set to $[1 \times 1 \times 3]$ to ensure that the convolution only occurred along the spectral dimension. This facilitated the training, since individual spectra could be used as the training set, and the trained network could be applied to input RGB images with arbitrary spatial dimen- sions. Pipeline 1 is demonstrated in Fig. 4 (a).

2.3.2. Refine hypercubes with sparse hyperspectral signals (Pipeline 2)

Pipeline 1 was found to provide generally good spectral pre- diction (see Section 3.2), however the problem itself is highly ill- posed leading to inevitable errors. Pipeline 2 was developed on top of Pipeline 1 to predict a hypercube from two data sources: an RGB image ($M \times N \times 3$), and a sparse hypercube ($M \times N \times 24$) containing the HSI signals at the relevant spatial positions with all back-

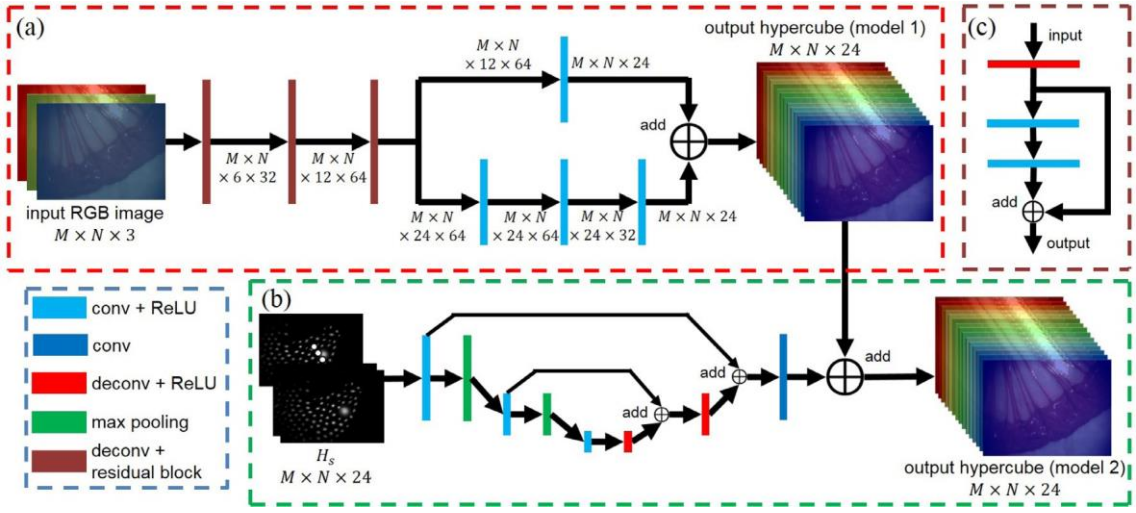

Fig. 4. The schematic of Pipeline 1 (a) and Pipeline 2 (a + b) in SSRNet_v2. (a) The schematic of Pipeline 1 that predicts a hypercube from an input RGB image. This pipeline contains two steps: 1) spectrally upscaling the RGB image; 2) Extraction of high-frequency signals and mergence with the low frequency ones. (b) The "mergence" part for Pipeline 2 on top of Pipeline 1, which merges the sparse hyperspectral signals for hypercube refinement. (c) Deconvolution layer followed by a residual block.

ground pixels as 0. An additional feature map was extracted from the sparse hypercube and added to the estimated hypercube from Pipeline 1 (Fig. 4 (b)). It is expected that combing these two data sources can allow correction of the RGB data to deal with inter- and intra-patient variability. Due to its spatial sparsity, informa- tion from local areas was propagated to the neighborhood of the sparse hypercube realized by a shallow FCN with all convolutional kernels sizes set to [3 × 3 × 3], to guarantee a sufficiently large re- ceptive field. This last convolutional layer produced a feature map that could be added to the output of Pipeline 1 to "fine-tune" the estimated hypercube.

Compared to SSRNet_v1, significant network architecture mod- ification has been carried out to enhance the model performance. In Pipeline 1, a kernel size of (1 × 1 × 3) was adopted instead of (1 × 1 × 2) for the deconvolutional layer for a larger receptive field in the spectral dimension; additional residual blocks were ap- pended to the deconvolutional layer for better feature extraction. A shallow U-Net style FCN was applied in the "mergence" part of Pipeline 2 to enlarge the receptive field of pixels in the output to guarantee a more effective propagation of sparse hyperspectral signal to the surrounding area. These improvements are shown in Section 3.2 .

2.3.3. *Training and prediction*

For training and testing, hypercubes collected using a liquid crystal tunable filter (LCTF) endoscopic imager during *in vivo* an- imal trials were used, including procedures on porcine bowel, rab- bit uterus, and sheep uterus. The LCTF is an electronically tunable optical filter that transmits light of selected wavelengths across the visible range. The wavelength interval in the animal trial was set to 10 nm with a range of 460 – 690 nm. A subset of images within this range (500 – 620 nm) was used for StO_2 estimation in pre- vious studies (Clancy et al., 2015; 2016). The misalignment of in- dividual hypercube slices at adjacent wavelengths due to tissue movement was compensated with the registration method pro- posed by Du et al. (Du et al., 2015). The transmission spectrum (h) of the CCD camera was used to generate the synthetic RGB images R from hypercubes, with $R = h * H$. The density map D_{hsi} of the sparse hyperspectral signals was produced using the spot segmentation results in the SL mode, where Gaussian distributions (max = 1) was simulated at the detected spot locations (where the sparse hyperspectral signals came from). The sparse HSI stack (H_s) was obtained by the element-wisely multiply the density map and the hypercube ($H_s = D_{hsi} . * H$).

Data augmentation including horizontal/vertical flipping and ro- tations has been applied to increase the size of the training set. Be- fore training all the weights were initialized with Gaussian distri- bution. According to a previous study Arad and Ben-Shahar (2016) , there exists a strong assumption that the influence from adjacent pixels is limited for spectra prediction. Therefore, in Pipeline 1 we trained our network with individual pixel spectra, in order to guar- antee sufficient training samples. In Pipeline 1 all the convolution occurred only along the spectral dimension, so the trained model can be fed with inputs with arbitrary spatial dimensions, since no spatial convolution was applied. When training Pipeline 2, the network was initialized with trained Pipeline 1. Data augmenta- tion including flipping and rotation was carried out to increase the training set and the generalization. A two-stage training strategy was adopted: the parameters in the shared layers with Pipeline 1 were frozen while the rest were updated; then all the parame- ters were updated until convergence. Adam optimizer and L2-norm loss function were empirically used for training. In prediction, RGB images were captured by the CCD camera used for synthetic RGB image generation, while the sparse hyperspectral signals and the density map came from the slit hyperspectral imager. Training and prediction were implemented using Tensorflow (Abadi et al., 2016) with a ≈ 2 FPS processing speed for hypercube generation.

3. Experimental results

In this section, we present the experimental results extended from Lin et al. (2016) and Lin et al. (2017) , on phantoms, *ex vivo* and *in vivo* to validate and on clinical cases to demonstrate the proposed pattern decoding and super-spectral-resolution algo- rithms.

3.1. *Pattern decoding and 3D reconstruction*

For stereo based reconstruction methods including the ICL SL system, the reconstruction accuracy is mainly limited by two fac- tors: calibration and feature matching. In this section, we mainly demonstrate the accuracy and robustness of our pattern decoding algorithm. Statistical evaluation has also been provided on 3D re- construction accuracy assessment, which shows how the pattern decoding affects the reconstruction results.

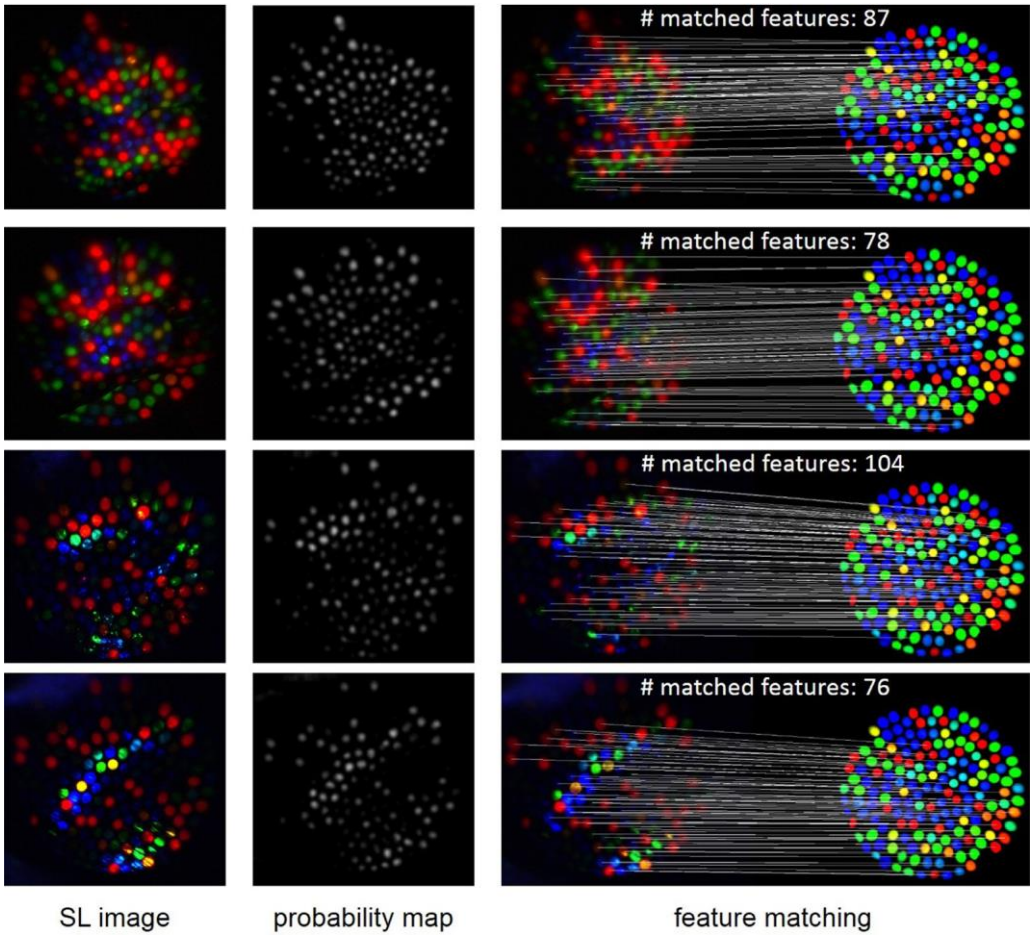

Fig. 5. Examples of pattern decoding using the proposed method. The estimated probability maps and the matching between the captured and reference SL images are shown for four example images. The upper two are for the ovine heart, and the lower two for the ovine liver.

Experiments have been carried out for validation on a silicone heart phantom, *ex vivo* ovine heart and liver, and porcine liver and bowel, with the probe adaptor used to fix the probe next to the endoscope at an angle ≈ 12 °, and baseline ≈ 6 mm. Videos were recorded for all experiments, in which the working distance was always kept at 1 . 5 − 4 cm, with the endoscope optical axis roughly perpendicular to the tissue surface. The endoscope move- ments were planned to simulate the use in the operating theater by including sudden fast and axial movements.

Fig. 5 provides several examples of the *ex vivo* "probability map" estimation and the feature matching results. It is clear that the proposed pattern decoding algorithm can process poor quality images even under heterogeneous illumination or specular high- lights. The corresponding "probability" for brighter and darker spots differ in size, as well as the maximal value, however, a nor- malization step forces the maximum value in individual binary masks to be 255 and ensures the detection of spots with different sizes. This algorithm may fail in over-exposed areas or where mas- sive specular reflections occur, but it generally guarantees robust performance for typical tissue surfaces. Also, it is noticeable that despite being successfully detected, some isolated spots induced by large surface curvatures could not be matched.

Statistical results are provided in Table 1 , for 10 frames per recorded video on different objects. Automatic feature matching results were compared with those from manual annotation. The true positive (correctly detected and matched spots), annotated matches (manually annotated spot matches), together with the matching sensitivity and precision are used as indicators. Compared to (Lin et al., 2016), extra experimental results on porcine liver and bowel were provided for validation.

Table 1 shows that the pattern decoding algorithm functions ro- bustly in general. The average sensitivity on the phantom experi- ments was higher than 0.9 due to higher image quality, reducing to around 0.85 in the *ex vivo* experiments, where the image qual- ity was affected by the appearance of mucosa, and specular reflec- tions. This indicates the capability of the algorithm for spot de- tection. It is also noticeable that for all data the feature matching precision was higher than 0.99, showing the increased reliability of the spot detection over previous work (Lin et al., 2015a). Mean- while, the real-time performance (12 FPS) guarantees real-world practicality. These improvements are owing to the FCN based de- tection algorithm, and the efficient feature matching enabled by customized feature descriptors and epipolar constraints.

In this work, we compared the reconstruction accuracy achieved using the proposed pattern decoding algorithm, with that using the manual pattern decoding, on the same eight SL images from the previous work Lin et al. (2015a) for reconstruction. These SL images were captured on the silicone heart phantom from dif- ferent angles, at a working distance around 100 mm. The ground truth surface shapes were measured by an MCAx25 handheld laser scanner. The average/maximum distances, from the reconstructed surface to the ground truth after iterative closest point (ICP) based rigid registration, were used as indicators for reconstruction accuracy evaluation (Table 2).

From Table 2 , it can be found that the proposed pattern decoding algorithm sacrificed little reconstruction accuracy, compared

Table 1
Validation of feature matching. Manual annotation is used as the reference.

Object	Annotated matches	True positive	Sensitivity	Precision
Silicone heart phantom	170 ± 1	154 ± 17	90.7% ± 10.1%	99.7% ± 0.4%
Ovine heart	134 ± 8	113 ± 10 110 ±	84.4% ± 3.8%	99.7% ± 0.6%
Ovine liver				99.4% ± 0.5%
Porcine liver	128 ± 11	13	86.1% ± 5.2%	99.7% ± 0.5%
Porcine bowel	136 ± 7	117 ± 9	85.8% ± 3.2%	99.1% ± 0.6%
	119 ± 14	91 ± 17	76.6% ± 7.1%	

Table 2
Validation of reconstruction accuracy (manual vs. proposed pattern decoding).

	Sensitivity	Precision	Average error (mm)	Max error (mm)
Manual	100%	100%	0.64 ± 0.14	2.78 ± 1.39
Proposed	92.5% ± 4.3%	99.4% ± 0.5%	0.68 ± 0.13	3.19 ± 1.45

with the manual method for the silicone heart phantom, which is due to the high spot identification precision (99.4% on average). From this experiment we show that given adequate pattern decod- ing results with the proposed method, accurate reconstruction re- sults can be achieved.

3.2. *Super-spectral-resolution*

We evaluated the proposed SSRNets (SSRNet_v1 proposed in Lin et al. (2017) and SSRNet_v2 proposed in this paper) for Pipeline 1 and 2 with intuitive and statistical analysis using including 50 hypercubes from porcine bowel, 21 from rabbit uterus, and 10 from sheep uterus *in vivo* . In order to compare and evaluate the cross- domain performance of the proposed algorithms, a dataset con- taining 243 hypercubes was generated by data augmentation and mixing, which was then divided 5 fold such that each experiment contained 200 hypercubes for training and the remaining 43 for testing using leave-one-out cross-validation (LOOCV) (Fig. 9). The peak signal-to-noise ratio ($PSNR = 20 \log_{10}(255/MSE)$, where MSE stands for mean square error) was deployed as the main indicator for validation.

In order to intuitively show the difference between Pipeline 1 and 2, the estimated hyperspectral signals from 8 points, chosen from representative areas in one pig bowel image, were compared (Fig. 6 (a)). It can be observed that both Pipeline 1 (blue line) and 2 (red line) enable hyperspectral signal prediction which is close to the ground truth (GT) (black line) measured by LCTF. In some cases Pipeline 1 (blue line) leads to inaccurate estimation for the wave- length range (580 − 600 nm) which may further lead to unwanted errors when using the estimated hypercube for StO_2 estimation which uses the wavelength range 580 − 600 nm (Clancy et al., 2015). Pipeline 2 outperforms Pipeline 1, especially when a sharp change is observed in the GT, and Fig. 6 (b)–(e) further compares PSNR and relative error for the pipelines. Both indicators in every pixel location are represented by their average values across the spectral domain, with Pipeline 2 slightly outperforming Pipeline 1. Excluding the specular reflection areas, for this example Pipeline 1 and 2 achieve a PSNR at 39.51 and 41.44 on average, with a rel- ative error of 0.80% and 0.63%, respectively. These intuitive exam- ples suggest that the SSRNet leads to promising results, and the integrated sparse hyperspectral signals are able to aid in refining the hypercube estimation in surrounding areas.

The models and pipelines are compared using a recovered hy- percube at different wavelengths, as shown in Fig. 7 , and according to the error maps SSRNet_v2 considerably outperforms SSRNet_v1. For SSRNet_v2, Pipeline 2 also managed to “propagate” the correct information from the captured sparse hyperspectral signals to the surrounding areas not only the local areas where the signals were collected. However, in the previous work, the propagation of hyperspectral reconstruction away from the sparse hyperspectral seed points was not always successful, resulting in “dot patterns”, which are “blotchy” artefacts with size and structure corresponding to the SL illumination (and sparse HSI detection) region (area indicated by red arrows in Fig. 7) that can be observed in the error map of Pipeline 2 from SSRNet_v1 at 580 nm.

Fig. 8 (a) shows an improvement from SSRNet_v1 (black curve) to SSRNet_v2 for Pipeline 2 (red curve) using the change in PSNR with wavelength, which is mainly caused by the deeper network and larger receptive field in the latter. By comparing Pipeline 1 and Pipeline 2 from SSRNet_v2, Fig. 8 (b) shows that integrat- ing the sparse hyperspectral signal improves the performance in most of the wavelengths, although not always significantly. This is because RGB images were the main contributors to hyper- cube prediction, while hyperspectral signals, being far sparser, could only be used to improve some “details” of the prediction achieved by the former. It is also noticeable that relatively poor PSNRs were obtained at the wavelength of 580, 590, and 600 nm, from all models. One of the main reasons is that the GT hyper- spectral signals in this range suddenly fluctuate, making accurate spectrum estimation more difficult, which can be observed from Fig. 6 (a). As end-to-end CNN models tend to produce smoother outcomes, this might lead to inaccurate estimation of signals us- ing the SSRNet at these wavelengths. These errors, however, could be compensated by hardware-based HSI with improved spatial res- olution, rather than by software based super-resolution methods only.

Evaluation of the cross-domain performance is of great im- portance on validating machine learning algorithms, especially for clinical applications, where high generalization is expected for proposed models. Therefore, we trained SSRNet on different data sources and tested on each other. Fig. 9 compares the PSNR of hy- percubes recovered by Pipeline 1 and 2 from SSRNet_v2.

As demonstrated in Fig. 9 , Pipeline 2 outperforms Pipeline 1 by integrating captured sparse hyperspectral signals. However, SS- RNet_v2 does not provide consistently excellent results: for exam- ple, the prediction on porcine bowel is relatively poor when the models are trained on other data. There are several reasons that might hinder the cross-domain application of SSRNet: 1) The train- ing set is not sufficiently large, for instance, the model trained on sheep uterus, which contains fewest examples, led to relatively poor results. 2) The training set was collected on *in vivo* ani- mal trials, where tissue movement introduced misalignment be- tween adjacent frames in hypercubes measured by the LCTF, which could not be completely compensated by automatic registration algorithms. However, despite this limited training set, promising results were still achieved with SSRNet according to the valida- tion above, demonstrating its potential to be used in real clinical cross-domain applications, given sufficient and accurate training sets.

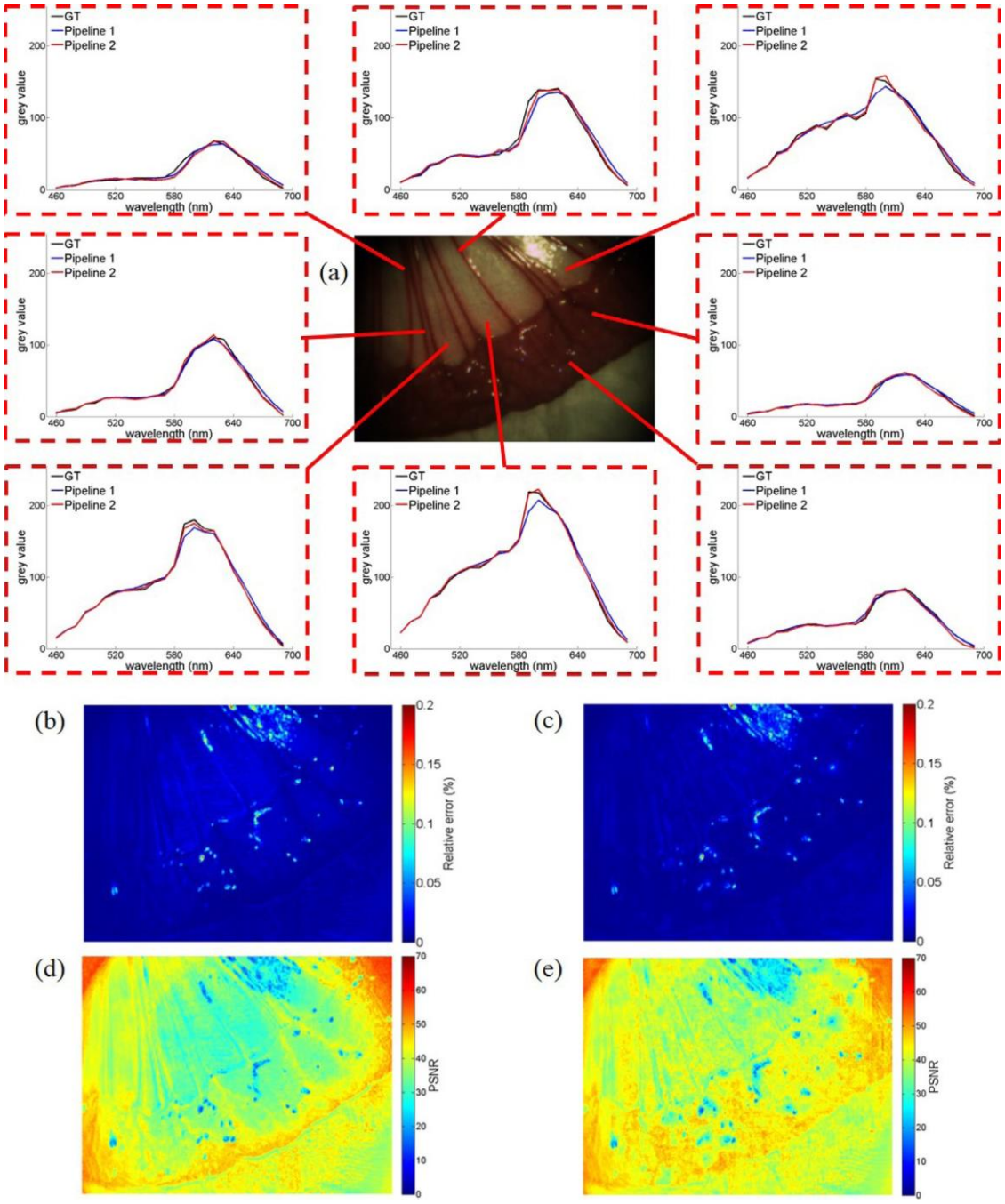

Fig. 6. (a) The RGB image and the estimated spectra (Pipeline 1: blue; Pipeline 2: red) vs. ground truth (black) from 8 locations. Relative error maps for Pipeline 1 (b) and 2 (c); PSNR map for Pipeline 1 (d) and Pipeline 2 (e) regarding the same sample (a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. *Clinical experiments*

Hypercubes provide rich information that can be utilized for tissue classification and cancer detection, or can function as the prerequisite for narrow-band imaging (NBI) or StO_2 measurement, aiding diagnosis and surgical navigation. Some intra-operative imaging examples on patient larynx tissue in laryngeal surgeries are provided in Fig. 10 to show feasibility for clinical application of the ICL SLHSI system with the proposed 3D reconstruction and SS- RNet based super-spectral-resolution algorithms. Our system was placed on a trolley and moved beside the operating table during the procedure, prior to diseased tissue resection. The endoscope with the SL probe was held by the surgeon and inserted through a laryngoscope of 1.5 cm diameter, imaging the larynx tissue. Data acquisition took approximately 5 minutes, after which the clinical procedure proceeded as normal. The captured SL/WL and hyper- spectral signal were analyzed offline post-operatively due the lim- ited computational power of the laptop used for camera control during imaging.

In order to jointly display reconstructed surface and estimated information, thin-plate splines (TPS) were applied to the sparsely reconstructed tissue surface to generate a smooth and dense surface. Next, the NB images and StO_2 maps, estimated from the predicted hypercubes, were overlaid onto these dense reconstructed surfaces (Fig. 10). NBI image generation (Fig. 10)(b), (e), (h)) was carried out by projecting the estimated hypercube onto narrow-

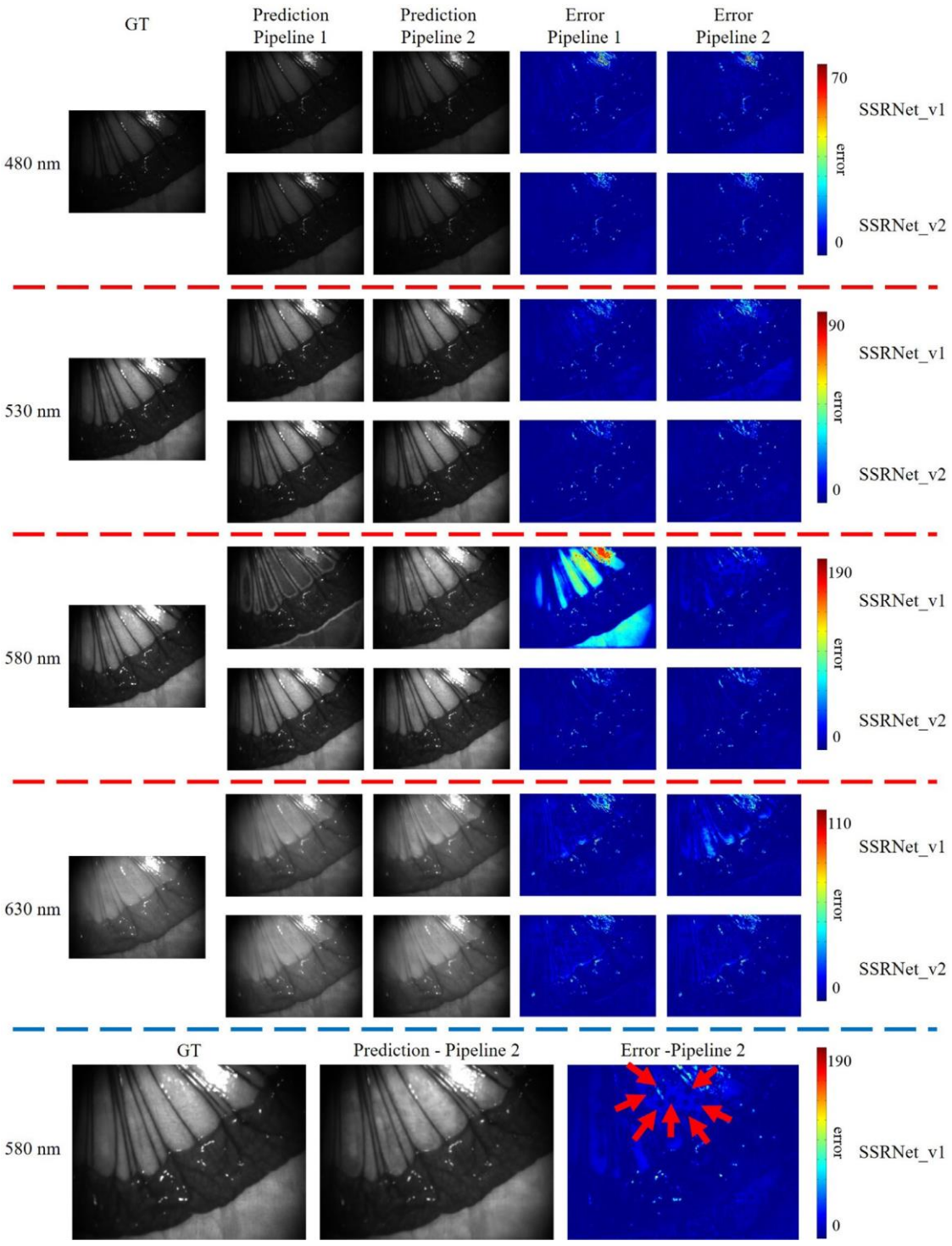

Fig. 7. Comparison of the recovered hypercube at different wavelengths for the two different pipelines and models. At each wavelength, the recovered hyperspectral images and the error maps obtained by different models are displayed. The bottom row provides "zoomed-in" versions of the predicated image and error map using Pipeline 2 from SSRNet_v1 at 580 nm, where the red arrows in the error map indicate the "dot patterns" mentioned in the discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

banded blue and green channels, where absorption by hemoglobin is highest, resulting in enhanced visualization of blood vessels. This emulates clinical systems currently used in endoscopy to increase image contrast and reveal areas of dysplasia (Matsuda et al., 2017). In StO_2 measurement examples (Fig. 10 (c), (f), (i)), areas with poor linear fitting were left blank. This situation normally occurs when no vessels exist or the predicted hypercube is erroneous due to e.g. camera over-exposure.

4. Discussion and conclusions

In this paper, we proposed the ICL SLHSI system, a dual- modality instrument that is capable of tissue surface shape 3D sensing and HSI.

In 3D reconstruction, more SL images were prepared to train our pattern decoding FCN model compared to the previous work (Lin et al., 2016), enhancing the robustness. Meanwhile, more *in*

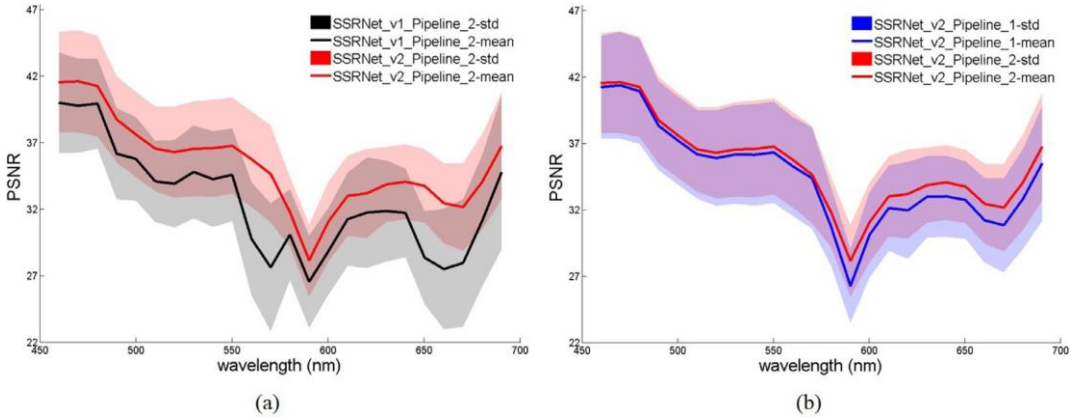

Fig. 8. (a) Average PSNR value along the wavelength for SSRNet_v1_Pipeline2 (black line), SSRNet_v2_Pipeline2 (red line), with their standard deviations (grey and red shadow). (b) Average PSNR value along the wavelength for SSRNet_v2_Pipeline1 (blue line), SSRNet_v2_Pipeline2 (red line), with their standard deviations (blue and red shadow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

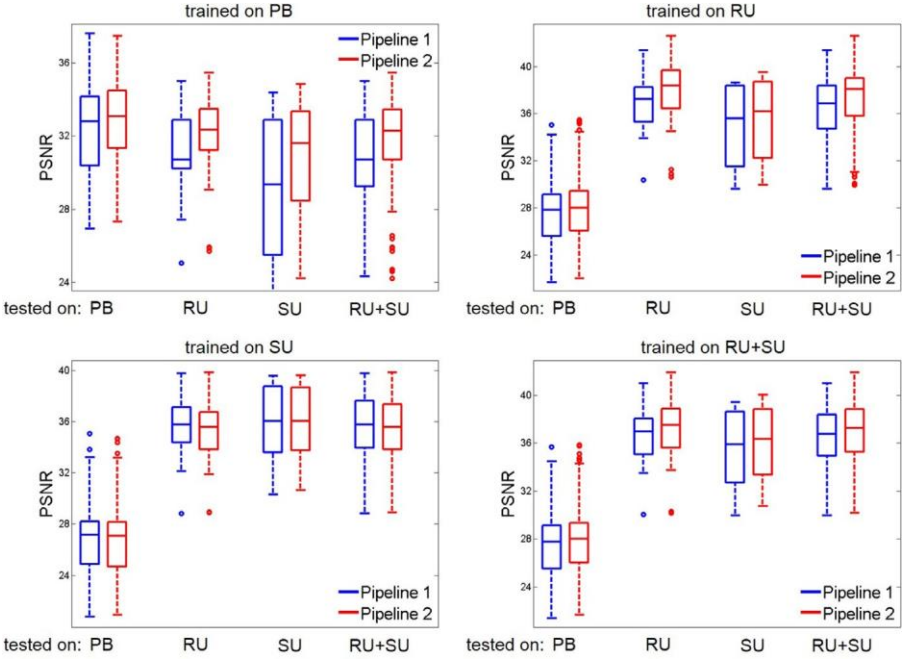

Fig. 9. The cross-domain performances (indicated by PSNR) with Pipeline 1 (blue) and Pipeline 2 (red) from SSRNet_v2. Both models were trained and tested on 4 datasets: pig bowel (PB), rabbit uterus (RU), sheep uterus (SU), and mixed rabbit/sheep uterus (RU + SU). 5-fold cross validation was applied to calculate the PSNR when training and testing on data from the same source. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vivo animal trials have been carried out to validate the perfor- mance of our algorithm. Through phantom experiments, we also show that the proposed method leads to accurate pattern decod- ing and consequently accurate 3D reconstruction. However, the al- gorithm could still be affected by image quality, mainly in terms of density, where generally fewer spots can be efficiently detected in images due to severe specular reflections, over-exposed regions, and diffusion of the spots within the mucosa. This could be com- pensated by 1) further improvements to the FCN model used for spot detection; 2) generation of a larger training set with higher generalization; 3) additional pre-processing like image deblurring and specular highlight suppression before pattern decoding; 4) combination with other 3D reconstruction techniques (e.g. monoc- ular SfM) using the WL images. As acquiring ground truth tissue surface shape information *in vivo* is challenging, more experiments on phantoms that simulate the *in vivo* tissue behavior will be car- ried out for further evaluation.

Another focus of this work is the super-spectral-resolution (SSR) problem. We conducted significant hardware improvements as well as algorithmic development to upgrade the ICL SL system to the SLHSI system, which is able to predict pixel-level dense hyper- cubes, containing 24 spectral bands at 10 nm intervals, from RGB images and spatially sparse hyperspectral signals. The spectral res- olution of the estimated hypercube has been demonstrated to be sufficient for StO_2 estimation as well as NBI. However, estimated hypercubes with different bandwidth and higher spectral resolu- tion would be preferred if we aim to use the them for other ap- plications, such as tumor detection. The optimization of the band selection in the predicted hypercube will be one of the research foci in the near future. During experimental evaluation on *in vivo* animal data, obvious improvements have been found from SSRNet_V1, a previous model, to SSRNet_v2, the latest model. This is mainly owning to the larger convolutional kernel and added resid- ual blocks that effectively deepened the network depth. The difference between Pipeline 1 and Pipeline 2 in SSRNet_v2, however,

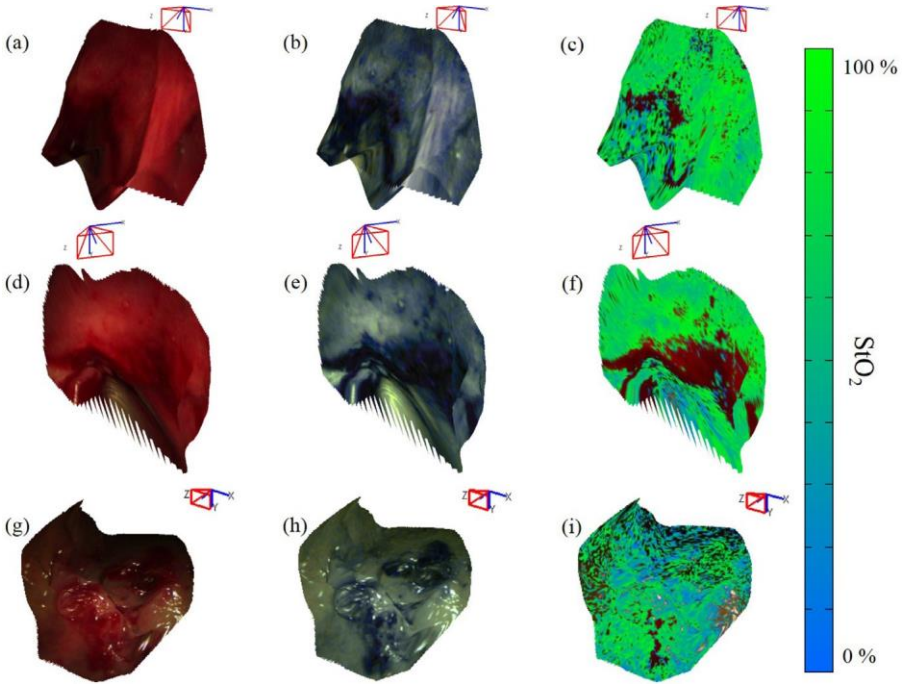

Fig. 10. Example *in vivo* (a), (d) and *ex vivo* (g) reconstructed patient larynx tissue surface. (b), (e), (h) Synthetic narrow band images and (c), (f), (i) synthetic oxygen saturation maps overlaid on these surfaces. In (c), (f), (i), brighter green means high StO_2, while deep blue means low StO_2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is small, as the latter only marginally outperformed the former at some wavelengths. This is mainly caused by the highly imbal- anced information distribution between two data sources: RGB im- ages are spatially dense and functioned as the main contributors for SSR, while the hyperspectral signals captured by our system are much sparser and could only be used to refine some details of the former prediction. It was also found that unexpected spec- trum fluctuation in the 580 − 600 nm wavelength range led to rel- atively inaccurate prediction in these bands. This could be poten- tially alleviated by increasing the spectral resolution of the ground truth/predicted hypercubes, increasing the number of training sets, or assigning higher weights for these bands during training. How- ever, it is hoped that the sparse hyperspectral data will eventu- ally allow correction of the RGB data to deal with inter- and intra- patient variability as well as acquisition artefacts, which are com- monly reported issues in hyperspectral imaging.

Usability of the system during clinical experiments, which aimed to generate NBI and StO_2 maps, has also been demonstrated. Currently, the data processing is still applied offline, due to the limited computational power of the hardware that is used in the operating theater. However, the fast processing speed of our pro- posed methods make the integration of the system into real-time clinical workflow possible. These algorithms still have potential to be sped up by fine-tuning the network architecture. The current acquisition speed (8 FPS) of the WL/SL images is limited by the CCD camera, which could be upgraded to enable signal synchro- nization at a higher framerate. During the patient experiments, the ICL SLHSI system was used for imaging before the tissue dissec- tion, which added only approximately 5 min to the normal clin- ical workflow. The whole system was located on a trolley beside the operating table to provide this potentially valuable additional tissue information to the surgeons without significant interruption of the normal clinical workflow. More *in vivo* animal and exper- iments should be carried out to validate and improve the whole system. Patient experiments in other surgery are also desired to enable further study on tissue classification (e.g. tumor detection) with hyperspectral signal analysis.

To summarize, the ICL SLHSI system is proposed to enable both tissue surface shape 3D sensing and HSI, as well as to jointly dis- play information from different imaging modalities. The 3D recon- struction was based on a structured light technique, while HSI was realized via addition of a slit hyperspectral imager to the system. The probe's flexibility and dimension can enable its integration with standard surgical tools in MAS, for example to access the ab- domen, gastrointestinal tract, or larynx. This system can potentially be used for clinical research including tissue depth measurement, StO_2 estimation, NBI, and AR.

Acknowledgements

This work was funded by ERC 242991 and an Imperial College Confidence in Concept award. Jianyu Lin was supported by IGHI scholarship. The authors gratefully acknowledge infrastructure sup- port from the Cancer Research UK Imperial Centre, the Imperial Ex- perimental Cancer Medicine Centre and the National Institute for Health Research Imperial Biomedical Research Centre.

Ethics statement

The ethics approval for human study was covered by Central London Research Ethics Committee (reference No. 10/H0718/55), the animal study was conducted under UK Home Office license (reference No. 70/24843, 70/7508, 70/6927, 8012639).

Conflict of interest

None of the authors has a conflict of interest.

References

Abadi, M. , Barham, P. , Chen, J. , Chen, Z. , Davis, A. , Dean, J. , Devin, M. , Ghemawat, S. , Irving, G. , Isard, M. , Kudlur, M. , Levenberg, J. , Monga, R. , Moore, S. , Mur- ray, D.G. , Steiner, B. , Tucker, P.A. , Vasudevan, V. , Warden, P. , Wicke, M. , Yu, Y. , Zhang, X. , 2016. Tensorflow: a system for large-scale machine learning. CoRR abs/1605.08695 .

Aiazzi, B. , Alparone, L. , Barducci, A. , Baronti, S. , Marcoionni, P. , Pippi, I. , Selva, M. , 2006. Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers. Ann. Geophys. 49 (1) .

Arad, B. , Ben-Shahar, O. , 2016. Sparse Recovery of Hyperspectral Signal from Natural RGB Images. Springer International Publishing, Cham, pp. 19–34 .

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 24 81–24 95. doi: 10.1109/TPAMI.2016.2644615 .

Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C., 2017. The status of augmented re- ality in laparoscopic surgery as of 2016. Med. Image Anal. 37, 66–90. doi: 10.1016/j.media.2017.01.007 .

Boppart, S.A., Luo, W., Marks, D.L., Singletary, K.W., 2004. Optical coherence tomog- raphy: feasibility for basic research and image-guided surgery of breast cancer. Breast Cancer Res. Treat. 84 (2), 85–97. doi: 10.1023/B:BREA.0 0 0 0 018401.13609. 54 .

Chan, M., Lin, W., Zhou, C., Qu, J.Y., 2003. Miniaturized three-ddimensional endo- scopic imaging system based on active stereovision. Appl. Opt. 42 (10), 1888– 1898. doi: 10.1364/AO.42.001888 .

Clancy, N.T., Arya, S., Stoyanov, D., Singh, M., Hanna, G.B., Elson, D.S., 2015. Intraop- erative measurement of bowel oxygen saturation using a multispectral imaging laparoscope. Biomed. Opt. Express 6 (10), 4179–4190. doi: 10.1364/BOE.6.004179 .

Clancy, N.T., Saso, S., Stoyanov, D., Sauvage, V., Corless, D.J., Boyd, M., Noakes, D.E., Thum, M.-Y., Ghaem-Maghami, S., Smith, J.R., Elson, D.S., 2016. Multispectral imaging of organ viability during uterine transplantation surgery in rabbits and sheep. J. Biomed. Opt. 21. doi: 10.1117/1.JBO.21.10.106006 . 21–21–7.

Clancy, N.T., Stoyanov, D., Maier-Hein, L., Groch, A., Yang, G.-Z., Elson, D.S., 2011. Spectrally encoded fiber-based structured lighting probe for intraoperative 3d imaging. Biomed. Opt. Express 2 (11), 3119–3128. doi: 10.1364/BOE.2.003119 .

Darzi, A., Mackay, S., 2002. Recent advances in minimal access surgery. BMJ 324 (7328), 31–34. doi: 10.1136/bmj.324.7328.31 .

Davies, B., Jakopec, M., Harris, S.J., Baena, F.R.Y., Barrett, A., Evangelidis, A., Gomes, P., Henckel, J., Cobb, J., 2006. Active-constraint robotics for surgery. Proc. IEEE 94 (9), 1696–1704. doi: 10.1109/JPROC.2006.880680 .

Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image super-resolution using deep con- volutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307. doi: 10.1109/TPAMI.2015.2439281 .

Du, X., Clancy, N., Arya, S., Hanna, G.B., Kelly, J., Elson, D.S., Stoyanov, D., 2015. Robust surface tracking combining features, intensity and illumination com- pensation. Int. J. Comput. Assist. Radiol. Surg. 10 (12), 1915–1926. doi: 10.1007/ s11548-015-1243-9 .

Edgcumbe, P., Pratt, P., Yang, G.-Z., Nguan, C., Rohling, R., 2015. Pico lantern: surface reconstruction and augmented reality in laparoscopic surgery using a pick-up laser projector. Med. Image Anal. 25 (1), 95–102. doi: 10.1016/j.media.2015.04. 008 .

Ferris, D.G. , Lawhead, R.A. , Dickman, E.D. , Holtzapple, N. , Miller, J.A. , Grogan, S. , Bambot, S. , Agrawal, A. , Faupel, M.L. , 2001. Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia. J. Low. Genit. Tract Dis. 5 (2), 65–72 .

Goodfellow, I. , Pouget-Abadie, J. , Mirza, M. , Xu, B. , Warde-Farley, D. , Ozair, S. , Courville, A. , Bengio, Y. , 2014. Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neu- ral Information Processing Systems 27. Curran Associates, Inc., pp. 2672–2680 .

Hasegawa, K., Noda, K., Sato, Y., 2002. Electronic endoscope system for shape mea- surement. In: Pattern Recognition, 2002. Proceedings. 16th International Confer- ence on, 1, pp. 792–795. doi: 10.1109/ICPR.2002.1044878 .

He, K. , Zhang, X. , Ren, S. , Sun, J. , 2015. Deep residual learning for image recognition. CoRR abs/1512.03385 .

Itseez, 2015. Open source computer vision library. https://github.com/itseez/opencv .

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: convolutional architecture for fast feature embedding. arXiv:1408.5093 .

Jones, G., Clancy, N.T., Helo, Y., Arridge, S., Elson, D.S., Stoyanov, D., 2017. Bayesian estimation of intrinsic tissue oxygenation and perfusion from rgb images. IEEE Trans. Med. Imaging 36 (7), 1491–1501. doi: 10.1109/TMI.2017.2665627 .

Kim, J. , Lee, J.K. , Lee, K.M. , 2016. Accurate image super-resolution using very deep convolutional networks. In: 2016 IEEE Conference on Computer Vision and Pat- tern Recognition (CVPR), pp. 1646–1654 .

Kim, J. , Lee, J.K. , Lee, K.M. , 2016. Deeply-recursive convolutional network for im- age super-resolution. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1637–1645 .

King, D.R., Li, W., Squiers, J.J., Mohan, R., Sellke, E., Mo, W., Zhang, X., Fan, W., Di- Maio, J.M., Thatcher, J.E., 2015. Surgical wound debridement sequentially char- acterized in a porcine burn model with multispectral imaging. Burns 41 (7), 1478–1487. doi: 10.1016/j.burns.2015.05.009 .

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 analy- sis for colorectal cancer detection. Anticancer Res. 36 (8), 3925–3932 . http: //ar.iiarjournals.org/content/36/8/3925.full.pdf+html .

Ledig, C., Theis, L., Huszr, 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), pp. 105–114. doi: 10.1109/ CVPR.2017.19 .

Lin, J., Clancy, N.T., Elson, D.S., 2015. An endoscopic structured light system using multispectral detection. Int. J. Comput. Assist. Radiol. Surg. 10 (12), 1941–1950. doi: 10.1007/s11548-015-1264-4 .

Lin, J. , Clancy, N.T. , Hu, Y. , Qi, J. , Tatla, T. , Stoyanov, D. , Maier-Hein, L. , Elson, D.S. , 2017. Endoscopic depth measurement and super-spectral-resolution imaging. Springer International Publishing, Cham, pp. 39–47 .

Lin, J. , Clancy, N.T. , Stoyanov, D. , Elson, D.S. , 2015. Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system. Springer International Publishing, Cham, pp. 405–412 .

Lin, J. , Clancy, N.T. , Sun, X. , Qi, J. , Janatka, M. , Stoyanov, D. , Elson, D.S. , 2016. Probe-Based Rapid Hybrid Hyperspectral and Tissue Surface Imaging Aided by Fully Convolutional Networks. Springer International Publishing, Cham, pp. 414–422 .

Loncan, L., de Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobi- geon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., Tourneret, J.Y., Vegan- zones, M.A., Vivone, G., Wei, Q., Yokoya, N., 2015. Hyperspectral pansharpening: a review. IEEE Geosci. Remote Sens. Mag. 3 (3), 27–46. doi: 10.1109/MGRS.2015. 2440094 .

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recog- nition (CVPR), pp. 3431–3440. doi: 10.1109/CVPR.2015.7298965 .

Lu, G., Fei, B., 2014. Medical hyperspectral imaging: a review. J. Biomed. Opt. 19 (1), 010901. doi: 10.1117/1.JBO.19.1.010901 .

Luo, H., Xu, J., Binh, N.H., Liu, S., Zhang, C., Chen, K., 2014. A simple calibration procedure for structured light system. Opt. Lasers Eng. 57 (0), 6–12. doi: 10.1016/ j.optlaseng.2014.01.010 .

Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., Stoyanov, D., 2013. Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. Med. Image Anal. 17 (8), 974–996. doi: 10.1016/j.media.2013.04.003 .

Marcu, L. , French, P.M.W. , Elson, D.S. , 2014. Fluorescence lifetime spectroscopy and imaging: principles and applications in biomedical diagnostics. CRC Press .

Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolu- tional neural networks. Remote Sens (Basel) 8 (7). doi: 10.3390/rs8070594 .

Matsuda, T. , Ono, A. , Sekiguchi, M. , Fujii, T. , Saito, Y. , 2017. Advances in image en- hancement in colonoscopy for detection of adenomas. Nat. Rev. Gastroenterol. Hepatol. 14 (5), 305(10) .

Maurice, X., Albitar, C., Doignon, C., de Mathelin, M., 2012. A structured light-based laparoscope with real-time organs' surface reconstruction for minimally inva- sive surgery. In: Engineering in Medicine and Biology Society (EMBC), 2012 An- nual International Conference of the IEEE, pp. 5769–5772. doi: 10.1109/EMBC. 2012.6347305 .

Nagengast, W.B., Hartmans, E., Garcia-Allende, P.B., Peters, F.T.M., Linssen, M.D., Koch, M., Koller, M., Tjalma, J.J.J., Karrenbeld, A., Jorritsma-Smit, A., Kleibeuker, J.H., van Dam, G.M., Ntziachristos, V., 2017. Near-infrared fluo- rescence molecular endoscopy detects dysplastic oesophageal lesions using topical and systemic tracer of vascular endothelial growth factor a. Gut doi: 10.1136/gutjnl- 2017- 314953 .

Obuch, J.C., Pigott, C.M., Ahnen, D.J., 2015. Sessile serrated polyps: detection, erad- ication, and prevention of the evil twin. Curr. Treat. Options Gastroenterol. 13 (1), 156–170. doi: 10.1007/s11938- 015- 0046- y .

Oktay, O. , Bai, W. , Lee, M. , Guerrero, R. , Kamnitsas, K. , Caballero, J. , de Marvao, A. , Cook, S. , O'Regan, D. , Rueckert, D. , 2016. Multi-input Cardiac Image Super-Reso- lution Using Convolutional Neural Networks. Springer International Publishing, Cham, pp. 246–254 .

Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S., 1993. Orthogonal matching pursuit: re- cursive function approximation with applications to wavelet decomposition. In: Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, 1, pp. 40–44. doi: 10.1109/ACSSC.1993.342465 .

Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.-Z., 2012. An effective visualisation and registration system for image-guided robotic partial nephrectomy. J Robot Surg 6 (1), 23–31. doi: 10.1007/s11701-011- 0334- z .

Ronneberger, O. , Fischer, P. , Brox, T. , 2015. U-Net: convolutional Networks for Biomedical Image Segmentation. Springer International Publishing, Cham, pp. 234–241 .

Salvi, J., Fernandez, S., Pribanic, T., Llado, X., 2010. A state of the art in structured light patterns for surface profilometry. Pattern Recognit. 43 (8), 2666–2680. doi: 10.1016/j.patcog.2010.03.004 .

Schmalz, C., Forster, Schick, A., Angelopoulou, E., 2012. An endoscopic 3d scanner based on structured light. Med Image Anal 16 (5), 1063–1072. doi: 10.1016/j. media.2012.04.001 .

Schroeder, W. , Martin, K.M. , Lorensen, W.E. , 2006. The Visualization Toolkit (4th ed.). Kitware .

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651. doi: 10. 1109/TPAMI.2016.2572683 .

Shi, W., Caballero, J., Huszr, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub- pixel convolutional neural network. In: 2016 IEEE Conference on Computer Vi- sion and Pattern Recognition (CVPR), pp. 1874–1883. doi: 10.1109/CVPR.2016.207 .

Sommer, C., Straehle, C., Kthe, U., Hamprecht, F.A., 2011. Ilastik: Interactive learning and segmentation toolkit. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 230–233. doi: 10.1109/ISBI.2011.5872394 .

Sorg, B.S., Moeller, B.J., Donovan, O., Cao, Y., Dewhirst, M.W., 2005. Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hy- poxia development. J. Biomed. Opt. 10 (4), 044004–044004–11. doi: 10.1117/1. 2003369 .

Velanovich, V., 20 0 0. Laparoscopic vs open surgery. Surg. Endosc. 14 (1), 16–21. doi: 10.10 07/s0 0464990 0 0 03 .

Weitzel, L. , Krabbe, A. , Kroker, H. , Thatte, N. , Tacconi-Garman, L.E. , Cameron, M. , Genzel, R. , 1996. 3D: the next generation near-infrared imaging spectrometer. Astron. Astrophys. Suppl. Ser. 119 (3), 531–546 .

Wirkert, S.J., Kenngott, H., Mayer, B., Mietkowski, P., Wagner, M., Sauer, P., Clancy, N.T., Elson, D.S., Maier-Hein, L., 2016. Robust near real-time estimation of physiological parameters from megapixel multispectral images with inverse monte carlo and random forest regression. Int. J. Comput. Assist. Radiol. Surg. 11 (6), 909–917. doi: 10.1007/s11548- 016- 1376- 5 .

Wirkert, S.J. , Vemuri, A.S. , Kenngott, H.G. , Moccia, S. , Götz, M. , Mayer, B.F.B. , Maier-Hein, K.H. , Elson, D.S. , Maier-Hein, L. , 2017. Physiological parameter esti- mation from multispectral images unleashed. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer-Assisted Intervention MICCAI 2017. Springer International Pub- lishing, Cham, pp. 134–141 .

Wolfe, W.L. , 1997. Introduction to imaging spectrometers / William L. Wolfe. SPIE Bellingham, Wash .

Wu, T.T., Qu, J.Y., 2007. Optical imaging for medical diagnosis based on active stereo vision and motion tracking. Opt. Express 15 (16), 10421–10426. doi: 10.1364/OE. 15.010421 .

Zhong, J., Yang, B., Huang, G., Zhong, F., Chen, Z., 2016. Remote sensing image fu- sion with convolutional neural network. Sens. Imaging 17 (1), 10. doi: 10.1007/ s11220- 016- 0135- 6 .

Zuzak, K.J., Francis, R.P., Wehner, E.F., Litorja, M., Cadeddu, J.A., Livingston, E.H., 2011. Active dlp hyperspectral illumination: a noninvasive, in vivo, system character- ization

visualizing tissue oxygenation at near video rates. Anal. Chem. 83 (19), 7424–7430. doi: 10.1021/ac201467v . PMID: 21842837.