Learning colon centreline from optical colonoscopy, a new way to generate a map of the internal colon surface

Mohammad Ali Armin[1] ✉, *Nick Barnes*[1,2], *Florian Grimpen*[3], *Olivier Salvado*[1]

[1]*CSIRO (Data61) 3D Computer Vision, Canberra, Australia*
[2]*College of Engineering and Computer Science (ANU), Canberra, Australia*
[3]*Department of Gastroenterology and Hepatology, Royal Brisbane and Women's Hospital, Brisbane, Australia*
✉ *E-mail: m.a.armin@gmail.com*

Published in Healthcare Technology Letters; Received on 16th September 2019; Accepted on 2nd October 2019

Optical colonoscopy is known as a gold standard screening method in detecting and removing cancerous polyps. During this procedure, some polyps may be undetected due to their positions, not being covered by the camera or missed by the surgeon. In this Letter, the authors introduce a novel convolutional neural network (ConvNet) algorithm to map the internal colon surface to a 2D map (visibility map), which can be used to increase the awareness of clinicians about areas they might miss. This was achieved by leveraging a colonoscopy simulator to generate a dataset consisting of colonoscopy video frames and their corresponding colon centreline (CCL) points in 3D camera coordinates. A pair of video frames were used as input to a ConvNet, whereas the output was a point on the CCL and its direction vector. By knowing CCL for each frame and roughly modelling the colon as a cylinder, frames could be unrolled to build a visibility map. They validated their results using both simulated and real colonoscopy frames. Their results showed that using consecutive simulated frames to learn the CCL can be generalised to real colonoscopy video frames to generate a visibility map.

1. Introduction: Colorectal cancer is the second cause of cancer mortality in Australia, and worldwide [1, 2]. The chance of survival can be increased to 90% if it is diagnosed at early stages. Colonoscopy is a common practice to detect and remove colonic polyps, yet the chance of missing polyps is relatively high [3]. This might be due to the polyp structure and position (e.g. behind a fold) or lack of coverage of the colon surface. Under optimal conditions, it is expected that around 90–95% of the colon to be inspected, while in practice only 81% of the colon mucosa is typically visualised [4]. While polyp detection from colonoscopy videos has been widely investigated [5], fewer studies have investigated how to assist clinicians, particularly junior clinicians in ensuring complete coverage during the procedure [6–8].

One previous approach to detecting missed areas focuses on detecting regions behind haustral fold [6]. For example, Mahmood and Durr [8] proposed a generative adversarial network method to estimate colon depth from real images by using simulated images. However, here, we aim to generate a map of the internal colon surface (visibility map). Such a map can provide useful information about uncovered areas during colonoscopy, map the position of any detected polyp, and be used as a reference to follow up with patients. Previous work taking a similar approach [7, 9] generates a visibility map using the following steps: (i) estimate camera pose and infer 3D structure, (ii) fit a cylinder into the 3D structure to estimate an average radius, (iii) compute the centre-of-dark region for each frame, (iv) using camera parameters and the centre-of-dark region, project the cylinder onto endoscopy images and unroll the images (into band images), (v) stitch the band images to generate a visibility map. Since this method is based on camera pose estimation and 3D reconstruction, it can be computationally expensive and complex and can be unstable and fail for sparsely textured frames with the complex structure of the colon wall. In particular, Armin *et al.* [9] used traditional feature point matching methods, which can perform poorly under difficult visual conditions, such as in colonoscopy.

One approach to improve robustness is to use convolutional neural network (ConvNet) methods such as [8] or [10] to directly generate the 3D structure of a colon, or ConvNet approaches to compute optical flow that can then be used to estimate structure [11]. However, these methods either need a dataset annotated with depth, which is hard to obtain for real optical colonoscopy, or they predict depth from a single frame which adds complexity. Generating complete 3D information is not necessary to infer visibility and is a significant source of error. We propose a novel approach, we use regression and train a network to learn directly simple low-dimensional geometrical parameters of a colon segment (here, centreline) in camera coordinates and use this to directly estimate a visibility map. We propose a ConvNet to learn the colon centreline (CCL) and its direction from simulated colonoscopy video frames. Our proposed method consists of two phases: (i) train a ConvNet with simulated colonoscopy video frames for which their CCLs in camera coordinate are known (a pair of consecutive frames is used as input to the network and the output is a point of CCL and its direction), (ii) test the ConvNet on real colonoscopy video frames by generating a visibility map. The summary of our method is presented in Fig. 1. Our contributions are as follows: (i) an algorithm that combines motion and appearance cues from training data from an endoscopic simulator to learn to predict the centreline of the colon from red–green–blue-only image sequences obtained during real and simulated endoscopic procedures; (ii) this is the basis of a new algorithm that learns to project optical colonoscopy frames to a map to enable accurate visualisation of visual information over an endoscopic procedure; (iii) we show that this leads to the generation of a more accurate map with a smoother mapping and reduced artefacts compared to previous methods, without requiring additional information (e.g. computed tomography or hardware end-effector localisation [12, 13]).

2. Method: A CCL is a curve in space that represents the centre of a surface of revolution that approximates the structure of the colon (the green line in Fig. 2). Note that the colon is not generally a surface of revolution, and our method does not require it to be. Our aim is to learn to predict a CCL in 3D camera coordinates. This provides the viewed colon segment centre and direction. Using this information, we locally generate a cylinder and project it onto the image to unroll it to build a small portion of a visibility map. These portions can then be joined to form a map of the internal colon surface. Considering the importance of

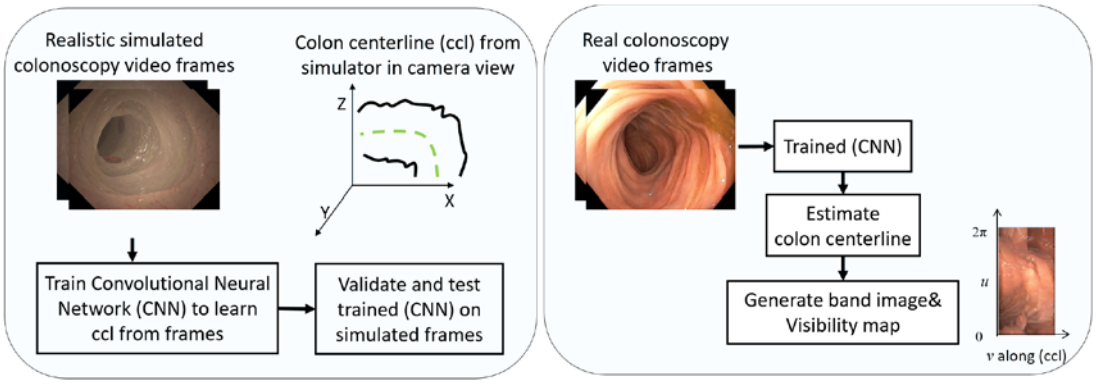

Fig. 1 *Schematic of the proposed processing pipeline*

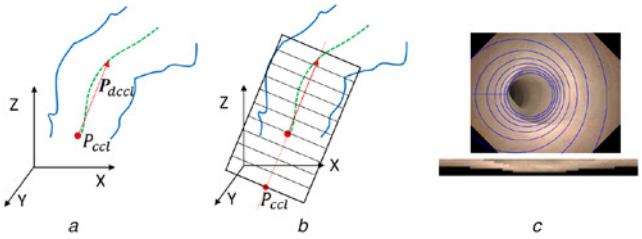

Fig. 2 *Colon centreline and band image*
a Colon with its centreline, P_{ccl} is a point on the CCL and $\boldsymbol{P}_{\text{dccl}}$ is the direction to a second point (red arrow)
b Cylinder generated in the direction of colon and
c Projection of the cylinder onto the image and the band image

understanding CCL and its application in our method, first, we explain how a visibility map can be generated by knowing the CCL, and then we introduce a ConvNet to learn the CCL.

2.1. Cylindrical model: A cylinder model can be determined in camera coordinates by estimating P_{ccl} as the point of intersection between the CCL and the image plane (or the closest point on the CCL to the centre of projection if they do not intersect (Figs. 2*a* and *b*)). A vector with the direction of the CCL in camera coordinates $\boldsymbol{P}_{\text{dccl}}$ from P_{ccl} is defined by taking a point on the CCL at a geodesic distance a from P_{ccl}, where $s \leq a$ is the length of the vector. Then a line segment in the direction of the CCL can be defined as

$$P(i) = \left\{ i\boldsymbol{P}_{\text{dccl}} + P_{\text{ccl}} | i \in [-0.01, d] \right\} \tag{1}$$

where d is the length of the cylinder. Then the surface of the cylinder $P_{\text{c}}(i)$ can be estimated by

$$\begin{aligned} P_{\text{c}}(i) = \{ & i\boldsymbol{P}_{\text{dccl}} + P_{\text{ccl}} + r\boldsymbol{u}\cos(\theta) + r\boldsymbol{v}\sin(\theta) | \\ & i \in [-0.01, d],\ \theta \in (0, 2\pi) \} \end{aligned} \tag{2}$$

where r is the radius of the cylinder and $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors orthogonal to the CCL. Note that defining $\boldsymbol{P}_{\text{dccl}}$ is based on a point at a distance so that smoothness of direction is maintained by being well ahead of the camera viewing location. This means that a local cylinder will remain in the overall direction of the colon and be less subject to rapid change at corners. The definition here is aimed to produce a smoothly changing mapping.

2.2. Band image and visibility map
2.2.1 Band image: By knowing intrinsic camera parameters, the above cylinder can be projected onto the colonoscopy video frame to unroll it and generate a radial strip called a band image (Fig. 2*c*) [7]. The radius of the cylinder is set to be constant throughout. The radius of 2 cm was empirically chosen for our experiments.

2.2.2 Visibility map: Band images were stitched by computing average motion flow in the x and y directions, estimated by FlowNet2 [14] from consecutive band images. A median filter was applied to the average motion flows to ensure a consistent motion. Some examples of a visibility map generated by our method and method explained in [7] are presented in Fig. 3 (first panel in each group).

2.3. Loss function: The loss function took the L2 norm between predicted and ground truth centreline information. Specifically, there were two separate terms, being the error for the CCL point P_{ccl_p}, and the direction $\boldsymbol{P}_{\text{dccl}_p}$ with λ as scale coefficient. The final loss is the sum of these terms

$$L = |P_{\text{ccl}} - P_{\text{ccl}_p}|^2 + \lambda |\boldsymbol{P}_{\text{dccl}} - \boldsymbol{P}_{\text{dccl}_p}|^2 \tag{3}$$

2.4. ConvNet and implementation details
2.4.1 Network architecture: In our experiments, we used VGG due to its high performance on our dataset. VGG is a ConvNet which consists of 16 convolutional layers, with a uniform architecture. We modified this architecture to take two consecutive frames as input, in a similar manner to FlowNet2 [11, 14]. The final fully connected output layer was reduced to predict six parameters, representing a CCL P_{ccl_p} and its direction $\boldsymbol{P}_{\text{dccl}_p}$.

2.4.2 Pre-image processing and implementation: Since the simulated frames have different colour distribution in comparison to real frames. We equalised both real and simulated video frames and resized them to comply with the input size of the ConvNet. A pre-trained VGG was used to learn CCL parameters. The learning rate was set to 1×10^{-4} and the network was trained for 100 epochs using the TensorFlow interface [15].

3. Experiments and results

3.1. Dataset
3.1.1 Simulated video frames: Simulated colonoscopy video frames with realistic appearance including specular reflection, texture, and blood vessels, were generated using a high fidelity simulator employing a complex parametric mathematical model. Using OpenGL, information such as CCL points and camera pose could be extracted for each frame. We used the CCL as ground truth in our experiments. Our simulated dataset consists of 20,827 video frames from segments of 12 different simulated colons which were generated by a variety of possible camera motions. Frames were generated at a simulated rate of 30 frames/s with a size of 676×540 pixels. This was split into 80% training and 20%

doi: 10.1049/htl.2019.0073

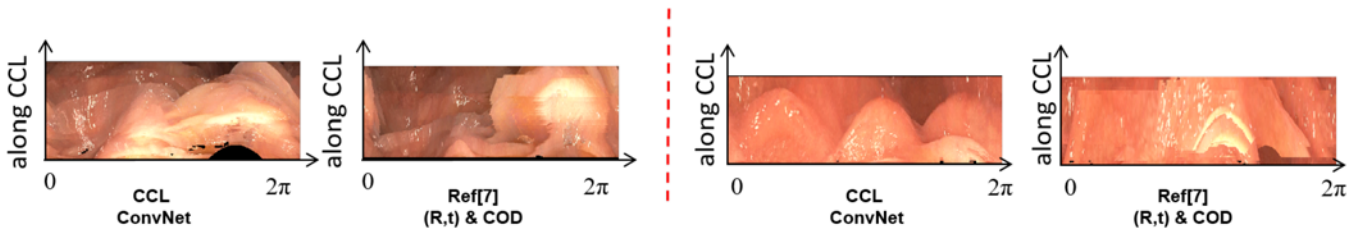

Fig. 3 *Two sample visibility maps generated from two different colon segments generated by our method and [7]. On the left our map shows an uncovered area, this has not appeared on the map generated by Armin et al. [7]. The difference between our method and [7] is also clear as the map generated by our method has fewer artefacts and shows colon folds due to its consistency in CCL detection*

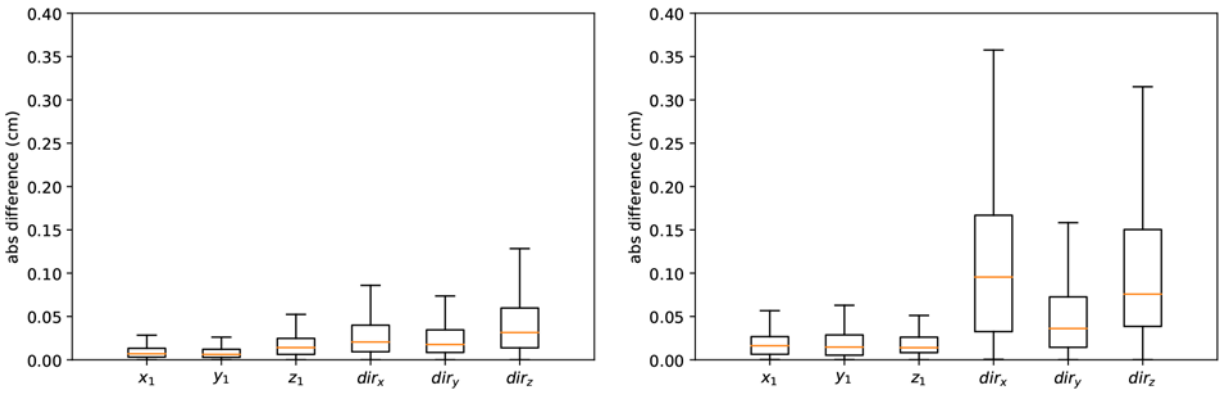

Fig. 4 *Validation of trained VGG on simulated training dataset (left panel), test network for generalisation by using a simulated colonoscopy video that has not been used in training or validation (right panel). x, y, z represent centreline points and dir is for the direction of colon*

validation frames. We also generated a separate video for test that was not used in training or validation.

3.1.2 Real colonoscopy video frames: We tested our trained network on ten different segments of real colonoscopy videos from five different patients. Uninformative frames (frames with no technical or medical information) were removed. The videos were captured by a 190HD Olympus endoscope, with a frame size 1352 × 1080 pixels, and a capture rate of 50 frames/s. In total 2515 real video frames were used to test our method.

3.2. Performance on simulated colonoscopy videos: The absolute difference error results for the validation set are presented by boxplot in Fig. 4 left panel and results showing the generalisation ability of the network to predict the CCL evaluated using the test video are shown in Fig. 4 right panels. In general, the absolute difference error for the centreline parameters for validation and test sets was <0.15 and 0.40 cm, respectively. We performed an ablation study to demonstrate the performance of our proposed method when only one frame was used for training versus two frames. The results are shown in Fig. 5.

3.3. Performance on real colonoscopy videos: As the CCL was not estimated in [7] and camera parameters along with the centre-of-dark region were used to project cylinder onto the image, we were unable to directly compare CCL results from our method with them. Instead, we implemented the method of Armin *et al.* [7] to generate a visibility map and compared it with a map generated by our proposed method. The projection of the cylinder using our CCL ConvNet method is shown as the first row of Fig. 6 and the second row shows those generated by Armin *et al.* [7] using the centre-of-dark region and camera R,t. In these sequences, camera translates to the left side of colon occluding part of the darkest region, though the centreline does not shift dramatically. For our method, the projected cylinder is correctly stretched to the side

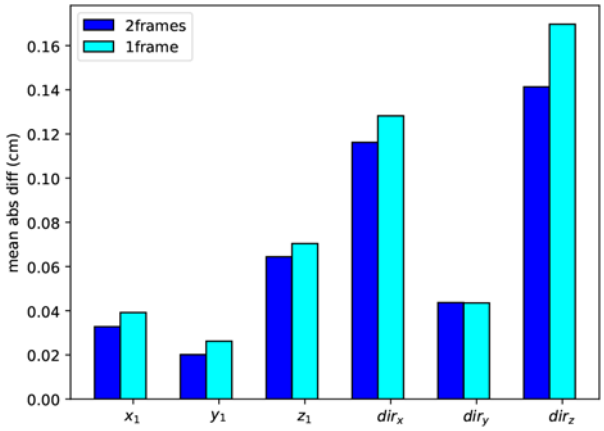

Fig. 5 *Comparison between the network when it trained with a pair of consecutive frames versus one frame to estimate CCL parameters, here (x_1, y_1, z_1) represent centreline points and dir is for the direction of CCL*

where it has moved closer to the wall but retains orientation towards what appears as the centre, whereas, in [7] the estimated centreline is shifted significantly as the darkest region in the image is partially occluded shifting the centre-of-dark region, showing the limitations of this method. In Fig. 3, comparative examples of visibility maps generated by both methods are presented for two short sequences. For the rightmost visibility maps, our method shows fewer artefacts, the corresponding video (see supplementary results) shows a fold which can be seen in our map but is lost in a less consistent map by Armin *et al.* [7]. In the left maps, our method shows a region that is uncovered in this sequence due to the downward tilt of the camera (see supplementary results), but this is not shown by Armin *et al.* [7]. This reflects a more accurate estimation of the centreline.

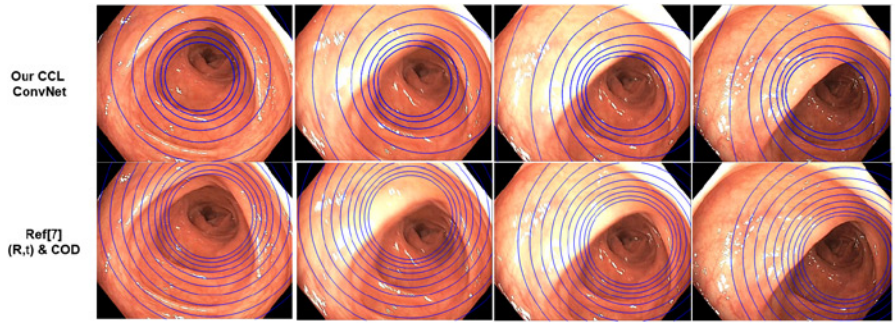

Fig. 6 *Projection of the cylinder onto a short sequence of real frames. The first row shows projection using CCL information estimated by our ConvNet and the second row presents results from [7]. When the camera moves from middle to left, the position of the projected cylinder using [7] changes with respect to the change of darkest region, but in our method, CCL remains consistent and keeps cylinder projection following the true centreline position*

4. Discussion and conclusion: This Letter presented a ConvNet approach that learns the CCL and its direction from optical colonoscopy frames. This is used to roughly fit a collection of concentric circles to the colon segment in 3D, and project them onto the image to unroll the image to form a band image. Stitching band images can provide a visibility map which can show any uncovered regions. This can help to improve awareness of uncovered areas, particularly for junior clinicians, and so improve the quality of colonoscopy.

In comparison to existing methods [7, 9], which were based on computing camera pose, 3D reconstruction and estimating the centre-of-dark region to estimate the centreline, we showed that the colon direction and centreline point could be directly learned by a ConvNet. Our results presented in Fig. 4 indicated the learning and generalisation of the CCL by a ConvNet. Further our results indicated that the estimation learned from simulated images can generalise to real colonoscopy video frames, which has not been shown previously for a ConvNet approach in optical colonoscopy.

Our approach used two frames, enabling the possibility to learn optical flow to have an indication of structure, while also having access to appearance parameters such as folds and dark regions in estimating the centreline. We speculate that ConvNet is able to incorporate these features to gain a more effective model than the previous more heuristic approach. The ablation study shows a clear gain from using multiple frames. However, we will investigate the performance of our ConvNet method using domain adaptation methods as in [16].

Our method currently projects the colon as a collection of concentric circles with the same radius as a rough estimation, and therefore in our future work, we are aiming to train a network to learn additional parameters of a colon segment such as radii for each chamber, along with colon structure and camera parameters. Using VGG has been shown to be effective; other networks such as ResNet and DenseNet along with a bigger dataset will be investigated and compared with the ConvNet used in this Letter in future work.

In summary, our ConvNet-based method shows promising results for CCL estimation and generating a map of the internal colon surface. More investigations need to be performed to generate a real-time visibility map with high precision. This can help clinicians to make better decisions while inspecting a colon.

5 References

[1] Prema T., Michael S., Jane Y., *ET AL.*: 'Colorectal cancer in Australia – cancer guidelines wiki'. Available at https://wiki.cancer.org.au/australia/Guidelines:Colorectal_cancer/Colorectal_Cancer_in_Australia#Introduction

[2] 'World health organization (WHO)'. Available at https://www.who.int/news-room/fact-sheets/detail/cancer

[3] Imperiale T.F., Glowinski E.A., Juliar B.E., *ET AL.*: 'Variation in polyp detection rates at screening colonoscopy', *Gastrointest. Endosc.*, 2009, **69**, (7), pp. 1288–1295. Available at http://www.sciencedirect.com/science/article/pii/S001651070703180X

[4] Edakkanambeth Varayil J., Enders F., Tavanapong W., *ET AL.*: 'Colonoscopy: what endoscopists inspect under optimal conditions', *Gastroenterology*, 2011, **140**, (5), p. S-718. Available at http://dx.doi.org/10.1016/S0016-5085(11)62982-X

[5] Tajbakhsh N., Shin J.Y., Gurudu S.R., *ET AL.*: 'Convolutional neural networks for medical image analysis: full training or fine tuning?', *IEEE Trans. Med. Imaging*, 2016, **35**, (5), pp. 1299–1312. Available at http://ieeexplore.ieee.org/document/7426826/

[6] Hong D., Tavanapong W., Wong J., *ET AL.*: '3D reconstruction of virtual colon structures from colonoscopy images', *Comput. Med. Imaging Graph.*, 2013, **38**, (1), pp. 22–33. Available at http://linkinghub.elsevier.com/retrieve/pii/S0895611113001523

[7] Armin M.A., Chetty G., De Visser H., *ET AL.*: 'Automated visibility map of the internal colon surface from colonoscopy video', *Int. J. Comput. Assist. Radiol. Surg.*, 2016, **11**, (9), pp. 1599–1610. Available at http://link.springer.com/10.1007/s11548-016-1462-8

[8] Mahmood F., Durr N.J.: 'Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy', *Med. Image Anal.*, 2018, **48**, pp. 230–243. Available at http://www.sciencedirect.com/science/article/pii/S1361841518303761

[9] Armin M.A., De Visser H., Chetty G., *ET AL.*: 'Visibility map: a new method in evaluation quality of optical colonoscopy'. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany, 2015, vol. 9349, pp. 396–404. Available at http://link.springer.com/10.1007/978-3-319-24553-9_49

[10] Yin Z., Shi J.: 'GeoNet: unsupervised learning of dense depth, optical flow and camera pose'. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 2018. Available at http://arxiv.org/abs/1803.02276

[11] Armin M.A., Barnes N., Khan S.: 'Unsupervised learning of endoscopy video frames correspondences from global and local transformation'. Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Granada, Spain, 2018 (*LNCS*), pp. 108–117

[12] Halier S., Angenent S., Tannenbaurn A., *ET AL.*: 'Nondistorting flattening maps and the 3-D visualization of colon CT images', *IEEE Trans. Med. Imaging*, 2000, **19**, (7), pp. 665–670

[13] Sudarsky S., Geiger B., Chefdhotel C., *ET AL.*: 'Colon unfolding via skeletal subspace deformation'. Medical Image Computing and Computer Assisted Intervention MICCAI 2008, New York, USA, 2008 (*LNCS*), pp. 205–212

[14] Ilg E., Mayer N., Saikia T., *ET AL.*: 'Flownet 2.0: evolution of optical flow estimation with deep networks'. Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, 2017, pp. 2462–2470. Available at http://openaccess.thecvf.com/content_cvpr_2017/html/Ilg_FlowNet_2.html

[15] Abadi M., Agarwal A., Barham P.: 'Tensorflow: large-scale machine learning on heterogeneous distributed systems', arXiv:160304467 [cs], 2016. Available at http://arxiv.org/abs/1603.04467

[16] Rau A., Edwards P.E., Ahmad O.F., *ET AL.*: 'Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy', *Int. J. Comput. Assist. Radiol. Surg.*, 2019, **14**, (7), pp. 1167–1176