HHS Public Access

Author manuscript

Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2021 October 11.

Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2021 April ; 2021: 329–333. doi:10.1109/isbi48211.2021.9433982.

VISUALIZING MISSING SURFACES IN COLONOSCOPY VIDEOS USING SHARED LATENT SPACE REPRESENTATIONS

Shawn Mathew[1,*], **Saad Nadeem**[2,*], **Arie Kaufman**[1]

[1]Stony Brook University, Department of Computer Science, USA

[2]Memorial Sloan Kettering Cancer Center, Department of Medical Physics, USA

Abstract

Optical colonoscopy (OC), the most prevalent colon cancer screening tool, has a high miss rate due to a number of factors, including the geometry of the colon (haustral fold and sharp bends occlusions), endoscopist inexperience or fatigue, endoscope field of view. We present a framework to visualize the missed regions per-frame during OC, and provides a workable clinical solution. Specifically, we make use of 3D reconstructed virtual colonoscopy (VC) data and the insight that VC and OC share the same underlying geometry but differ in color, texture and specular reflections, embedded in the OC. A lossy unpaired image-to-image translation model is introduced with enforced shared latent space for OC and VC. This shared space captures the geometric information while deferring the color, texture, and specular information creation to additional Gaussian noise input. The latter can be utilized to generate one-to-many mappings from VC to OC and OC to OC. The code, data and trained models will be released via our Computational Endoscopy Platform at https://github.com/nadeemlab/CEP.

Keywords

Colonoscopy; Virtual Colonoscopy; CycleGAN; Shared latent space; Directional discriminator

1. INTRODUCTION

More than 15 million colonoscopies are performed in the US every year [1, 2]. During these procedures, 22-28% of polyps and 20-24% adenomas are missed [3]. There are no commercial or automated tools available to assist endoscopists in gauging the amount of colon surface missed during OC procedures. The main culprit in substandard coverage during OC are the sharp bends and haustral folds, as depicted in Fig. 1a. Even though the endoscope tip can be flexed to look behind folds and sharp bends, beginner or tired endoscopists do not use this option wisely and may have a high miss rate. This miss rate can be reduced if endoscopists have a visualization tool to identify and investigate areas occluded by haustral folds.

*Equal Contribution

6.COMPLIANCE WITH ETHICAL STANDARDS

We used OC and VC retrospective human subjects with IRB approval. Also, OC video sequences and mesh results of Ma et al. [5] were used with their permission and prior IRB approval.

Recently, Ma et al. [5] and Freedman et al. [6] have presented approaches to quantify colon surface coverage. Ma et al. [5] reconstruct 3D mesh from contiguous chunk of colonoscopy video frames using training data generated from shape-from-motion. The missed surface is visualized as holes in the reconstructed mesh. The method, however, assumes cylindrical topology (endoluminal) views, smooth camera movements and masked-out specular reflection, making the method less practical in general colonoscopy scenarios. In contrast, Freedman et al. [6] have used a deep learning approach to estimate percentage coverage value directly for given colonoscopy video segments but do not provide any means for visualizing the missed colon surface.

In this paper, we present a deep learning model for realtime visualization of missed colon surfaces directly on the colonoscopy video frames without doing any prior offline 3D reconstruction using contiguous sets of frames. Specifically, we make use of prior 3D reconstructed virtual colonoscopy (VC) [7, 8] data, created from a computed tomography (CT) scan, to produce training data for missing surface visualization (Fig. 1b–d). This is used in conjunction with OC data for the same patient to drive an unpaired image-to-image translation with a modified lossy CycleGAN [4] and a new enforced shared OC and VC latent space representation. The lossy CycleGAN [4] by itself overfits due to the sparse training data for the missing surface task (most OC frames have no or few missing surface green pixels as opposed to the dense depth maps for which the lossy CycleGAN was originally proposed) and can easily hallucinate structures which do not exist, as shown in Fig. 1. Adding a shared latent space forces the network to preserve structures (and avoid hallucination) when translating between domains. With added Gaussian noise, we also show that the same framework with shared latent space representations can be used to generate realistic one-to-many mappings from VC to OC and OC to OC for augmenting OC datasets in computer-aided detection and classification pipelines.

In summary, the contributions of this paper are as follows:

- We are the first to present a model to visualize missing surface regions for individual colonoscopy frames in realtime.
- We introduce shared OC and VC latent space representations to get more consistent geometry for missing surface inference task.
- Using additional Gaussian noise input, the model can also generate realistic OC images (one-to-many mapping) with different specular reflections, lighting and texture for a given OC or VC frame.

2. DATA

The OC and VC datasets were acquired for 10 patients at Stony Brook University Hospital. 2000 images from 5 patients were used for training, while 800 images from 2 patients were used for validation and 1200 images from 3 patients were used for testing. Even though VC and OC are captured for the same patient, there is no ground truth since the shape of the colon differs considerably between the two procedures. The borders in the OC images were cropped to exclude the fisheye lens artifacts. 3D meshes were reconstructed from CT scans in VC, using a pipeline similar to Nadeem and Kaufman [9].

In order to create training data for per-frame missing surface visualization, the opacity of the 3D colon mesh is lowered such that the more opaque regions indicate the missed surfaces, which are colored green in Fig. 1c. The per-frame missing surface data is generated through Blender and example videos are provided[1]. Fig. 1 shows a typical colon anatomy along with the haustral folds and the pictorial representation of a missed surface for a certain endoscope camera position. To aid the model with the image-to-image domain translation task, we added the missing surface information in green channel on top of the VC rendering of the colon (Fig. 1d).

3. METHOD

The presented approach focuses on training two generator networks, G_{oc} and G_{vc} and two discriminator networks. To train these networks, an objective function $\mathscr{L}$ composed of three parts is used. The first part, $\mathscr{L}_{trans}$ focuses on learning a proper image-to-image translation. $\mathscr{L}_{adv}$ produces realistic images, and $\mathscr{L}_{noise}$ helps the network utilize noise input for OC image generation:

$$\mathscr{L} = \mathscr{L}_{trans} + \mathscr{L}_{adv} + \mathscr{L}_{noise} \tag{1}$$

In order to learn the image-to-image translation, a cycle consistency loss, $\mathscr{L}_{cyc}$, and an extended cycle consistency loss, $\mathscr{L}_{excyc}$ [4] are used. $\mathscr{L}_{excyc}$ allows for a one-to-many translation by making comparisons in a common domain. The common domain between OC and VC is VC, since OC has additional textures, lighting and specular reflections. The cycle consistency and extended cycle consistency losses are as follows:

$$\mathscr{L}_{cyc}(G_a, G_b, A) = \mathbb{E}_{y \sim p(A)} \| y - G_a(G_b(y)) \|_1 \tag{2}$$

$$\mathscr{L}_{excyc}(G_a, G_b, A) = \mathbb{E}_{y \sim p(A)} \| G_b(y) - G_b(G_a(G_b(y))) \|_1 \tag{3}$$

where $y \sim p(A)$ is the data distribution of domain A and $\|\cdot\|_1$ represents the L1 norm.

To make the translation more robust, we add a shared latent space loss. Each generator, G_A, is composed of an encoder, En_A, and decoder, De_A.

The encoder brings the input image into latent space, while the decoder takes the latent space into the image domain. We propose that OC and VC share the same latent space, as the latent space stores geometric information which should be consistent between the two domains. This is depicted in Fig. 2 and the following equation:

$$\mathscr{L}_{SLS}(En_B, G_B, En_A, A) = \mathbb{E}_{y \sim p(A)} \| En_B(y) - En_A(G_B(y)) \|_1, \tag{4}$$

[1]Supplementary Video: https://youtu.be/x1-wwCiYeC0

Lastly, we add an identity loss, as described by Zhu et al. [10], in order to ensure consistent shading, when translating between the two domains. This is not done in the OC domain as it restricts the network from using the noise input properly.

$$\begin{aligned}\mathscr{L}_{trans} &= \lambda_c\left[\mathscr{L}_{excyc}(G_{oc}, G_{vc}, I_{oc}) + \mathscr{L}_{cyc}(G_{vc}, G_{oc}, I_{vc})\right] \\ &+ \lambda_{SLS}[\mathscr{L}_{SLS}(En_{oc}, G_{oc}, En_{vc}, I_{vc}) \\ &+ \mathscr{L}_{SLS}(En_{vc}, G_{vc}, En_{oc}, I_{oc})] \\ &+ \lambda_{iden}\mathscr{L}_{iden}(I_{vc}),\end{aligned} \tag{5}$$

In our experiments, we set λ_c as 10, λ_{SLS} as 1, and λ_{iden} as 1.

Adversarial losses have shown success in producing realistic images. Specifically, we add a directional discriminator D_{dir}, to differentiate between the different directions of translation and an OC discriminator to restrain $G_{oc}(G_{vc}(I_{oc}))$, as described by [4].

$$\begin{aligned}\mathscr{L}_{dir}(G_a, G_b, D, A, B) &= \mathbb{E}_{y \sim p(A)}[\log(D(y, G_b(y))] \\ &+ \mathbb{E}_{x \sim p(B)}[\log(1 - D(G_a(x), x)]\end{aligned} \tag{6}$$

$$\mathscr{L}_{GAN}(G, D, A, B) = \mathbb{E}_{y \sim p(A)}[\log(D(y)] + \mathbb{E}_{x \sim p(B)}[\log(1 - D(G(x))] \tag{7}$$

These two losses compose the adversarial losses:

$$\mathscr{L}_{adv} = \mathscr{L}_{dir}(G_{oc}, G_{vc}, D_{dir}, I_{oc}, I_{vc}) + \mathscr{L}_{GAN}(G_{oc}, D_{oc}, I_{oc}, G_{oc}(G_{vc}(I_{oc}))) \tag{8}$$

We note that OC and VC share the same underlying geometric information, while OC has additional color, texture, and specular reflections. In order to reflect this additional information, we add a noise input to De_{oc} to drive the one-to-many mapping between VC and OC. $\mathscr{L}_{noise}$ is used to ensure a minimum distance between images with different noises, otherwise the noise vector is ignored:

$$\mathscr{L}_{noise}(De, N, L) = \mathbb{E}_{z_1, z_2 \sim p(N), l \sim p(L)} max(0, \|De(l, z_1) - De(l, z_2)\| - \alpha), \tag{9}$$

L is a latent space domain, N is a noise domain, and α is a variable to determine how much the images should differ. We set α to 0.1 in our experiments and draw our noise input from a normal distribution.

With L_{noise}, we can take a latent space variable along with various samples from the noise domain to produce OC images with different specular reflections, lighting, and texture. The latent space can be produced from both OC and VC images. With the addition of the noise input, we create a one-to-many mapping between VC-to-OC and OC-to-OC (Fig. 4).

We follow the same generator architecture described in CycleGAN [10] but instead of using 9 ResNet blocks, we use 10, 5 dedicated to encoder and the remaining 5 for decoder.

4. RESULTS

In Fig. 3, we show our results on OC frames from Ma et al. [5]. Ma et al. can produce 3D meshes from given OC video segments and can visualize the missing surfaces as holes in the reconstructed mesh. We highlight some holes in their reconstructed mesh that are detected by our model. Our results are similar for neighboring frames, without the use of any temporal connections for smoothing.

There is no ground truth for missed surfaces in OC data. To make a quantitative analysis, we texture a VC colon to create ground truth missed surface data. This is done by taking texture from OC frames, and mapping them on the VC colon mesh. Our method achieved an average per-pixel accuracy score of 81% and a Dice coefficient of .667 for the textured VC frames, despite the fact that surface area occluded by deep folds is difficult to predict in a single frame without additional information such as colon topology. Per-pixel accuracy is computed after converting the images into binary images based areas classified as missed:

$$Acc = \frac{TP + TN}{d}, \tag{10}$$

where TP and TN are the number of true positive and true negative pixels and d is the number of pixels in the image. The textured VC results are shown in the last three rows of Fig. 3. Complete videos are included in the supplement[2].

We added a noise loss to our model to generate realistic OC images from given OC or VC images. The generated images have the same underlying geometry as the input image but vary in lighting, specular reflections and texture, as shown in Fig. 4. The first two rows show VC to OC images. The last two rows show results for our OC to OC mapping. Just like the OC input, the model generates different lighting, specular reflections and texture. Note that the texture changes are more subtle than the changes in specular reflection and lighting. This will be addressed in future work.

5. LIMITATIONS AND FUTURE WORK

The missed polyps and anomalies are mostly occluded by the haustral folds. Even though our model in general works well for deep as well as shallow haustral folds, there are instances where only a partial missed surface area is highlighted for deep folds. This is understandable given the fact that we are not taking into account additional information, such as the overall colon topology. In the future, we will incorporate this information by inferring the colon centerline to improve the overall performance.

The current OC image generation creates a rather sparse distribution of images, especially in the texture space. To improve this, we will split texture, specular reflection, and lighting into three separate noise vectors which will provide finer control over these aspects and can potentially force the model to generate a more diverse set of OC images.

[2]Supplementary Video: https://youtu.be/x1-wwCiYeC0

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

We Dr. Sarah McGill (UNC Chapel Hill) for OC videos and to Ruibin Ma for comparisons. This research was funded in part by NIH/NCI Cancer Center Support Grant P30 CA008748 and NSF grants CNS1650499, OAC1919752, and ICER1940302.

8. REFERENCES

[1]. Joseph DA, Meester RGS, Zauber AG, Manninen DL, Winges L, Dong FB, Peaker B, and van Ballegooijen M, "Colorectal cancer screening: estimated future colonoscopy need and current volume and capacity," Cancer, vol. 122, no. 16, pp. 2479–2486, 2016. [PubMed: 27200481]

[2]. Seeff LC, Richards TB, Shapiro JA, Nadel MR, Manninen DL, Given LS, Dong FB, Winges LD, and McKenna MT, "How many endoscopies are performed for colorectal cancer screening? results from cdc's survey of endoscopic capacity," Gastroenterology, vol. 127, no. 6, pp. 1670–1677, 2004. [PubMed: 15578503]

[3]. Leufkens AM, Van Oijen MGH, Vleggaar FP, and Siersema PD, "Factors influencing the miss rate of polyps in a back-to-back colonoscopy study," Endoscopy, vol. 44, no. 05, pp. 470–475, 2012. [PubMed: 22441756]

[4]. Mathew S, Nadeem S, Kumari S, and Kaufman A, "Augmenting colonoscopy using extended and directional cyclegan for lossy image translation," arXiv preprint arXiv:2003.12473, 2020.

[5]. Ma R, Wang R, Pizer S, Rosenman J, McGill SK, and Frahm J-M, "Real-time 3D reconstruction of colonoscopic surfaces for determining missing regions," Med Image Comput Comput Assist Interv, pp. 573–582, 2019. [PubMed: 34113926]

[6]. Freedman D, Blau Yo, Katzir L, Aides A, Shimshoni I, Veikherman D, Golany T, Gordon A, Corrado G, Matias Y, and Rivlin E, "Detecting deficient coverage in colonoscopies," arXiv preprint arXiv:2001.08589, 2020.

[7]. Hong L, Muraki S, Kaufman A, Bartz D, and He T, "Virtual voyage: Interactive navigation in the human colon," 24th Conf Comput Gr & Interact Tech, pp. 27–34, 1997.

[8]. Pickhardt PJ, Choi JR, Hwang Inku, Butler JA, Puckett ML, Hildebrandt HA, Wong RK, Nugent PA, Mysliwiec PA, and Schindler WR, "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults," N Engl J Med, vol. 349, no. 23, pp. 2191–2200, 2003. [PubMed: 14657426]

[9]. Nadeem S and Kaufman A, "Computer-aided detection of polyps in optical colonoscopy images," SPIE Medical Imaging, vol. 9785, pp. 978525, 2016.

[10]. Zhu J-Y, Park T, Isola P, and Efros AA, "Unpaired image-to-image translation using cycle-consistent adversarial networks," IEEE Int Conf Computer Vision, pp. 2223–2232, 2017.

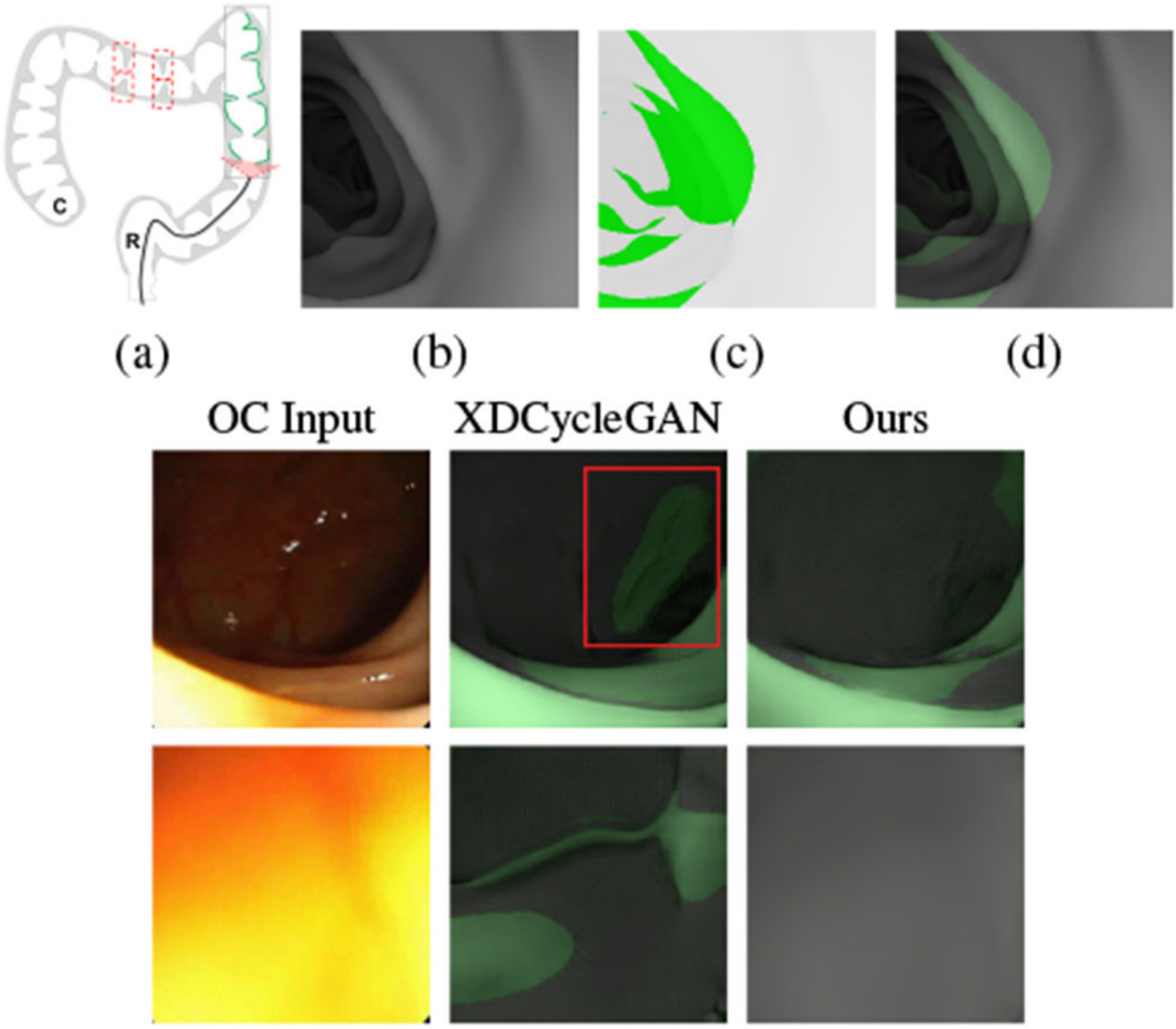

Fig. 1.

(a) Pictorial representation of a colon illustrating missed colon surface (green outline) when the endoscope (black line) traverses from rectum (R) to cecum (C) and back. (b) A VC rendering for a mesh reconstructed from a CT scan. (c) The missing surfaces (green regions) can be rendered by casting rays from virtual camera positions and marking mesh faces not directly intersecting the rays. (d) Missing colon surface visualization with a virtual colonoscopy rendering is used for training the model. At the bottom are examples of XDCycleGAN [4], proposed for scale-consistent depth map inference across colonoscopy video frames, overfitting for missing surface inference task due to sparse training data, as shown with the red bounding box and the predicted missed structures (bottom) when the camera is occluded.

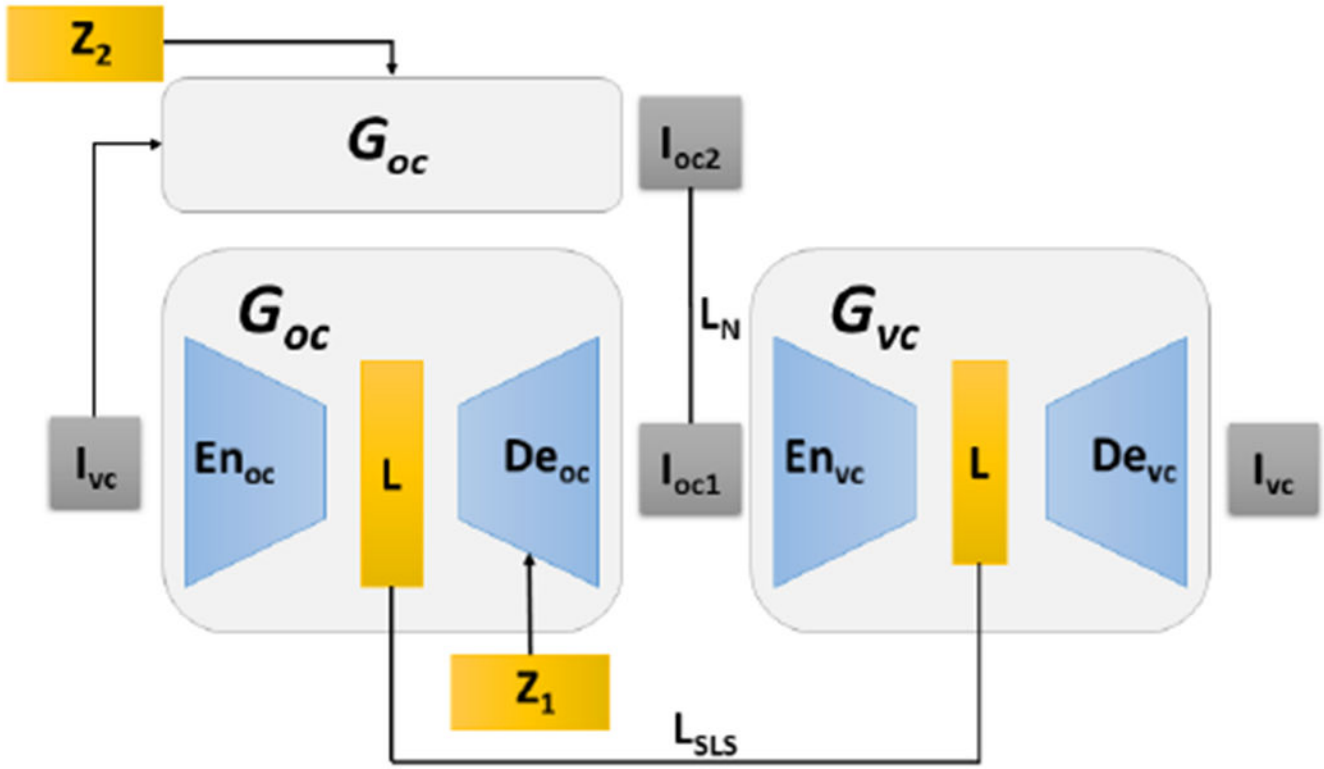

Fig. 2.

A VC image, I_{vc}, is brought into latent space with En_{oc}, and brought to the OC domain through De_{oc}. The OC image, I_{oc} returns to latent space through En_{vc}. The latent vector produced by En_{oc} and En_{vc} should be the same, since it stores the same geometric information. This is enforced by $\mathscr{L}_{SLS}$. To generate different OC images, De_{oc} takes noise input, Z, to create the additional OC information. In order to ensure the network uses the noise, we apply a loss, $\mathscr{L}_N$, between OC images with different noise vectors.

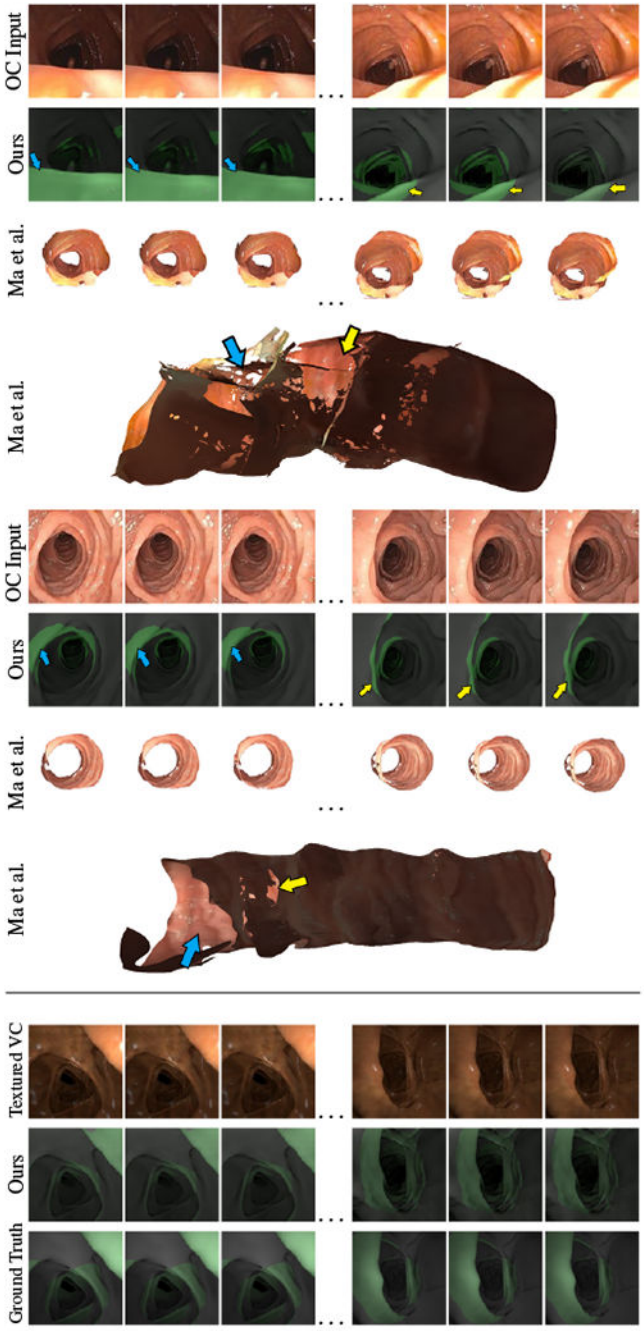

Fig. 3.

The top portion shows results on video sequences from Ma et al. [5]. We indicate the missing regions predicted by our model on the meshes reconstructed by their pipeline using blue and yellow arrows. The missing regions are visualized as holes on the reconstructed meshes. The bottom portion shows our results on textured VC input along with the ground truth.

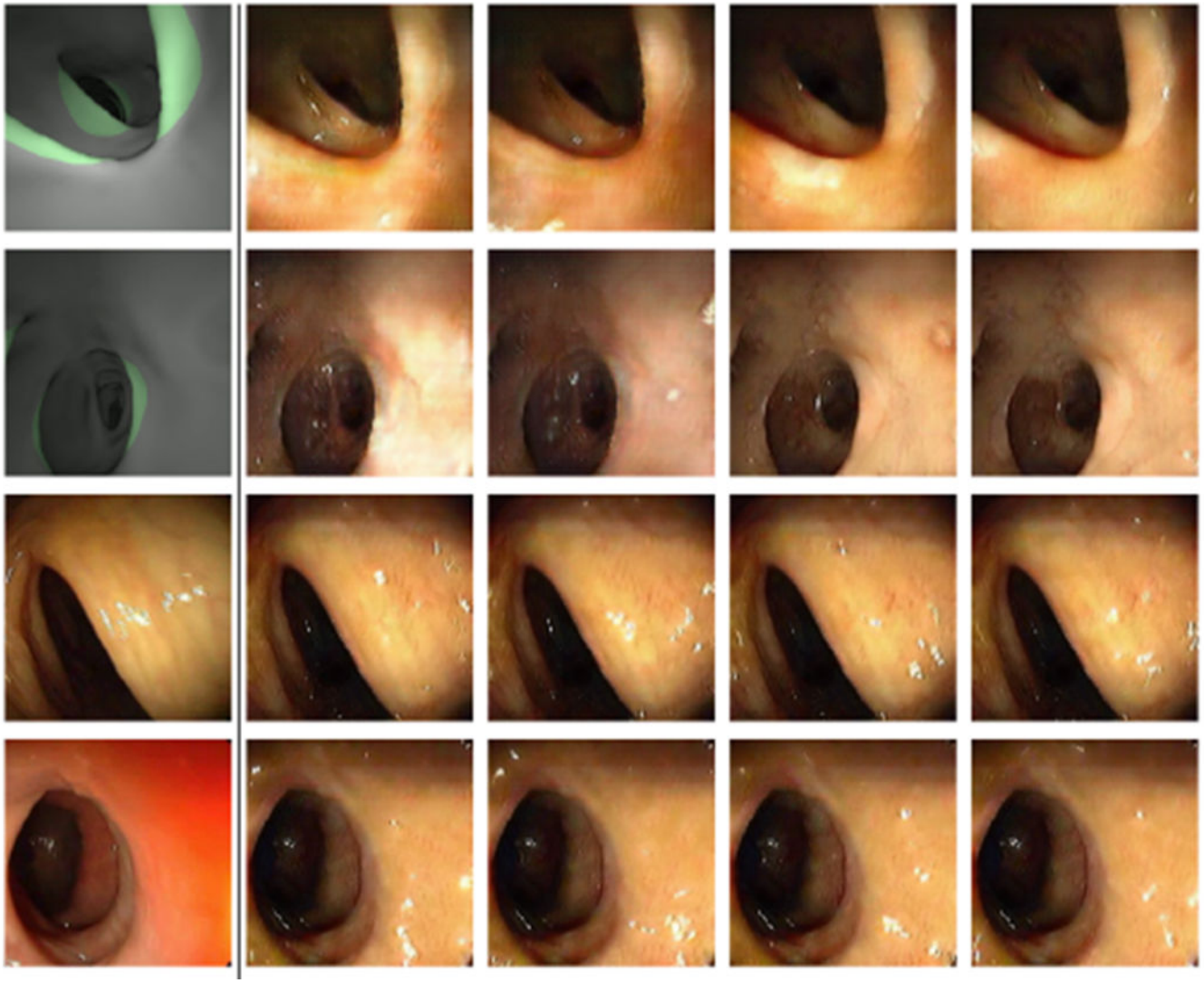

Fig. 4.

The first column is input for generating OC images with different lighting, specular reflections, and textures. The first two rows show VC to OC translation. The last two rows show our model results for OC to OC.