

Realistic endoscopic image generation method using virtual-to-real image-domain translation

Masahiro Oda¹ ✉, Kiyohito Tanaka², Hirotugu Takabatake³, Masaki Mori⁴, Hiroshi Natori⁵, Kensaku Mori^{1,6}

¹Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

²Department of Gastroenterology, Kyoto Second Red Cross Hospital, 355-5, Haruobi-cho, Kamigyo-ku, Kyoto, Kyoto 602-8026, Japan

³Department of Respiratory Medicine, Sapporo-Minami-Sanjo Hospital, Nishi-6-chome, Minami-3-jo, Chuo-ku, Sapporo, Hokkaido 060-0063, Japan

⁴Department of Respiratory Medicine, Sapporo-Kosei General Hospital, Higashi-8-chome, Kita-3-jo, Chuo-ku, Sapporo, Hokkaido 060-0033, Japan

⁵Department of Respiratory Medicine, Keiwakai Nishioka Hospital, 1-52, 4-jo 4-chome, Nishioka, Toyohira-ku, Sapporo, Hokkaido 062-0034, Japan

⁶Research Center for Medical Bigdata, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

✉ E-mail: moda@mori.m.is.nagoya-u.ac.jp

Published in Healthcare Technology Letters; Received on 18th September 2019; Accepted on 2nd October 2019

A realistic image generation method for visualisation in endoscopic simulation systems is proposed in this study. Endoscopic diagnosis and treatment are performed in many hospitals. To reduce complications related to endoscope insertions, endoscopic simulation systems are used for training or rehearsal of endoscope insertions. However, current simulation systems generate non-realistic virtual endoscopic images. To improve the value of the simulation systems, improvement of the reality of their generated images is necessary. The authors propose a realistic image generation method for endoscopic simulation systems. Virtual endoscopic images are generated by using a volume rendering method from a CT volume of a patient. They improve the reality of the virtual endoscopic images using a virtual-to-real image-domain translation technique. The image-domain translator is implemented as a fully convolutional network (FCN). They train the FCN by minimising a cycle consistency loss function. The FCN is trained using unpaired virtual and real endoscopic images. To obtain high-quality image-domain translation results, they perform an image cleansing to the real endoscopic image set. They tested to use the shallow U-Net, U-Net, deep U-Net, and U-Net having residual units as the image-domain translator. The deep U-Net and U-Net having residual units generated quite realistic images.

1. Introduction: Endoscopic diagnosis and treatment are commonly performed in hospitals. An endoscope is inserted into a patient's body through an incised part or a natural orifice to capture interior views of the body. Endoscopes such as a colonoscope, gastro-endoscope, and bronchoscope are used to diagnose digestive organs, trachea, and bronchi. Laparoscope and thoracoscope are used in surgeries. In endoscopic diagnosis and treatment, a surgeon or a physician needs to understand the condition of target organs or lesions from endoscopic images. On account of the limitation of the field of view and difficulty of understanding 3D structure, endoscopic procedures are difficult to perform. Complications occur in endoscopic diagnosis and treatment due to their difficulty. In colon diagnosis and treatment, some inexperienced physicians cause colon perforation using colonoscope. In laparoscopic surgeries, some inexperienced surgeons injure organs by the laparoscope or other surgical tools outside of the viewing fields of the laparoscope.

Complications caused related to insufficient recognitions of organ structures by physicians or surgeons during endoscopic procedures need to be reduced. To reduce such complications, training of endoscopic procedures using simulation systems are performed [1–4]. Endoscopic simulation systems are useful because physicians or surgeons can improve their manipulation ability without any trouble with patients. Such simulation systems are effective for less-experienced physicians or surgeons. When pre-operative 3D images of a patient are available, a patient-specific simulation or rehearsal can be performed using an endoscopic simulation system. Such patient-specific simulation systems are effective

not only for less-experienced but also experienced physicians or surgeons.

One problem of endoscopic simulation systems is the lack of reality on their generated virtual endoscopic images. Commonly, textures of the organ surfaces and light reflection on the virtual endoscopic images are different from the real images. The simulation systems commonly generate virtual images from the surface or volumetric model of organs. The surface or volumetric model represents anatomical structures coarsely compared to the real organ shapes. Detailed anatomical structures such as tiny blood vessels on organ surfaces are not included in the models. Such tiny blood vessels affect colour and texture of the organ surfaces. Colour and textures of the organ surfaces in real images need to be considered in image generations. To improve reality of virtual images, the texture mapping technique is used in simulation systems. However, preparation of good texture images for the texture mapping is difficult. Furthermore, the light reflection model employed by the simulation systems cannot simulate the real light reflection due to lack of the light reflection/absorption property information of the organ surfaces. The appearance of light reflection/absorption on the organ surfaces is not realistic on the generated images. In summary, (i) regardless of real textures of the organ surfaces and (ii) lack of the light reflection/absorption property information of the organ surfaces reduce reality of the generated virtual endoscopic images.

To improve the value of endoscopic simulation systems, improvement of the reality of their generated virtual endoscopic

images is necessary. Simulation systems having realistic image renderers greatly contribute skill improvement of surgeons or physicians in training or rehearsals of endoscope manipulation.

Recently, generative adversarial networks (GANs) are utilised for realistic image generation. Shrivastava *et al.* [5] improved the reality of simulated images using a framework of GAN to generate realistic images having annotation information. They applied their method to eye images and distance images of hand pose. Engelhardt *et al.* [6] proposed a method to improve reality of endoscopic images taken from surgical phantoms. They used the tempCycleGAN to convert endoscopic images. Gil *et al.* [7] and Esteban-Lansaque *et al.* [8] improved reality of virtual bronchoscopic images using GANs. Their ideas are mainly based on CycleGAN [9]. CycleGAN enables establishment of image translator from unpaired data. CycleGAN-based methods were proposed not only for realistic image generation but also segmentation [10] and image modality conversion [11, 12]. The previous method [6] introduced temporal information to CycleGAN to achieve temporally consistent results. Also, some methods [10–12] introduced original loss functions to impose constraint conditions to CycleGAN. CycleGAN uses an image-domain translation network to perform image conversion. The structure of the network is strongly related to its ability of learning representations from images. Quality of image-domain translation is significantly affected by the network. However, the relationship between image-domain translation quality and network structure has not been investigated.

We propose a realistic image generation method for endoscopic simulation systems. We generate virtual endoscopic images from a CT volume of a patient. We translate the virtual endoscopic images into the real endoscopic image domain using a virtual-to-real image-domain translation technique. The image-domain translator is build based on a large number of virtual and real endoscopic images as a fully convolutional network (FCN). Our method can be trained using unpaired image data. To obtain high-quality image-domain translation, we perform an image cleansing to the real endoscopic image set. The quality of translated images depends on the FCN network structure of the image-domain translator. We employed network structures including the shallow U-Net, U-Net, deep U-Net, and U-Net having residual units as the image-domain translator. Relationships between the network structures and translated image quality are investigated in our experiments. Our method can automatically generate realistic textures on the organ surfaces in the virtual endoscopic images.

The contributions of this paper are (i) proposal of realistic virtual endoscopic image generation method and (ii) investigation of relationships between image-domain translation network structures and quality of translated images.

2. Realistic endoscopic image generation method

2.1. Overview: The proposed method converts virtual endoscopic images into realistic endoscopic images. We use sets of the virtual and real endoscopic images. These images are obtained from different patients (unpaired data). In the real endoscopic image set, some images which are not suitable to establish translation from virtual to a real domain. Samples of such images and the reason why they are not suitable are given in Section 2.2.2. Therefore, we apply ‘a data cleansing’ process to improve the quality of image-domain translation. We make an image-domain translation network based on the CycleGAN [9] technique. The image-domain translation network is trained to minimise cycle consistency loss. We used four FCNs as the image-domain translator to investigate relationships between network structures and image-domain translation quality.

2.2. Data preparation and cleansing

2.2.1 Virtual endoscopic images: We generate virtual endoscopic images from CT volumes using a volume rendering method [13].

These images were taken during manual fly-through in hollow organs. Samples of images taken in the colon are shown in Fig. 1. The set of virtual endoscopic images are denoted as V .

2.2.2 Real endoscopic images: We capture real endoscopic images during endoscope insertions to patients (Fig. 2). Among these images, we exclude images that are not suitable for the image-domain translation. This process is called as ‘data cleansing’ process. This process contributes to improve the quality of image-domain translation. To obtain better image-domain translation results, both of the images in the source and target domains should capture similar objects, such as the organs. However, some real endoscopic images contain parts of endoscope, surgical tools, faeces, and fluid. Also, some real endoscopic images are taken in special imaging modes of endoscopes such as the narrow-band imaging and magnification modes. Samples of such images are shown in Fig. 3. They never appear in images in the virtual domain. Therefore, we remove these images from the set of real endoscopic images used to train the image-domain translator. The set of the remaining real endoscopic image is denoted as R .

2.3. Virtual-to-real image-domain translation

2.3.1 Image-domain translation method: We make virtual-to-real image-domain translation by using the sets of images V and R . The virtual-to-real image-domain translator performs mapping $G: V \rightarrow R$. We achieve mapping G in adversarial training frameworks. An adversarial discriminator D_R , which discriminates between real images r and translated high-quality images $G(v)$. v is a virtual image in V and r is a real image in R . Based on adversarial loss [14], we define an objective function for mapping G and discriminator D_R as

$$L_{\text{gan}}(G, D_R, V, R) = E_{r \in R} \{ \log D_R(r) \} + E_{v \in V} \{ \log (1 - D_R(G(v))) \}. \quad (1)$$

When real images are provided, the function gives likelihood $D_R(r) \in [0, 1]$. When virtual images are provided, the images are translated as $G(v)$, and the function gives likelihood $1 - D_R(G(v)) \in [0, 1]$. The objective function gives higher



Fig. 1 Samples of virtual endoscopic images taken in colon

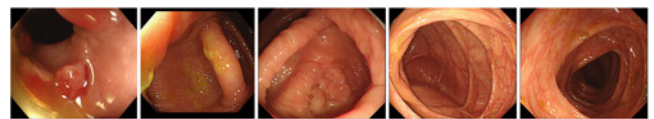


Fig. 2 Samples of real endoscopic images taken in colon

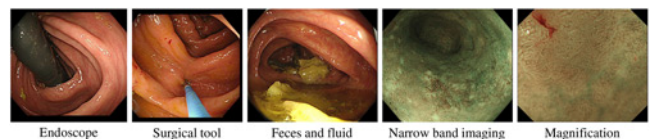


Fig. 3 Samples of real endoscopic images that are not used to train image-domain translator. These images contain endoscope, surgical tool, faeces, and fluid. Also, images taken in narrow-band imaging and magnification modes of endoscopes are not used to train

values if D_R discriminates real and translated real images correctly. Also, the objective function gives lower values if G generates good translated real images that fool the discriminator. G and D_R are trained to achieve $\arg \min_G \max_{D_R} L_{\text{gan}}(G, D_R, V, R)$. Similarly, we define an inverse mapping $F: R \rightarrow V$. An objective function for mapping F and discriminator D_V , which discriminates between v and $F(r)$, is also defined as $L_{\text{gan}}(F, D_V, R, V)$. However, use of the objective function is not enough to limit range of distributions of images in a target domain. Appropriate limitation to the distribution of a target domain makes translation results better. To achieve better mapping G , we introduce a cycle consistency loss as proposed in the cycle-consistent adversarial networks [9]. The loss implies an image-translation result of an input image by G and F is similar to the input image. This means $v \rightarrow G(v) \rightarrow F(G(v)) \approx v$. The cycle consistency loss is defined as

$$L_{\text{cyc}}(G, F) = E_{v \in V} \{ \|F(G(v)) - v\|_1 \} + E_{r \in R} \{ \|G(F(r)) - r\|_1 \}. \quad (2)$$

We define overall objective function as

$$L(G, F, D_V, D_R) = L_{\text{gan}}(G, D_R, V, R) + L_{\text{gan}}(F, D_V, R, V) + \lambda L_{\text{cyc}}(G, F), \quad (3)$$

where λ is the weight of the cycle consistency loss. We train G, F, D_V, D_R to achieve $\arg \min_{G, F} \max_{D_V, D_R} L(G, F, D_V, D_R)$.

For the training, virtual endoscopic images $v_i (i = 1, \dots, I) \in V$ and real endoscopic images $r_j (j = 1, \dots, J) \in R$ are used. I and J are the numbers of virtual and real endoscopic images. Both the discriminators D_R and D_V are implemented as convolutional neural networks. Both G and F are implemented as FCNs. Details of the network structures of G and F are described in Section 2.3.2. After the training, we obtain G as the virtual-to-real endoscopic image-domain translator.

2.3.2 Network structure of image-domain translator: The quality of images generated by the image-domain translator is affected by network structures of them. Therefore, we investigate relationships between the network structures and image-domain translation results. We use networks including the *shallow U-Net*, *U-Net*, *deep U-Net*, and *U-Net having residual units* as G and F . Each structure is explained below.

The *U-Net* used here is a modified version of the original U-Net [15], which has the instance normalisation [16] after each convolution layer.

The *shallow U-Net* is made by removing layers in the U-Net. A convolution layer is removed from the encoding and decoding paths of the U-Net, respectively.

The *deep U-Net* is made by adding layers to the U-Net. A convolution layer is added to the encoding and decoding paths of the U-Net, respectively.

The *U-Net having residual units* is made by replacing convolution layers in the U-Net with residual units [17].

3. Experiments: We performed realistic endoscopic image generation using the proposed method. We used virtual and real colonoscopic images in this experiment. We evaluated generated images qualitatively.

To train the virtual-to-real image-domain translator, we generated $I = 8072$ virtual colonoscopic images from six cases of abdominal CT volumes taken for CT colonography diagnosis (air is insufflated to the colon). The virtual colonoscopic images were captured during manual fly-through in the colons. We collected 18,775 real colonoscopic images, which were captured during colonoscope insertions to patients. Two colonoscopists operated a colonoscope. The real colonoscopic images were captured at random intervals

during colonoscope insertions (the images are not consecutive video frames). The data cleansing process was applied to the images. 5369 images were removed from the images in the process. The remaining 13,406 images were obtained as the data cleansed images. These images were used to train the image-domain translator. The CT volumes and the real colonoscopic images were taken from different patients. The networks were trained in 100 epochs. 100 epochs were enough for training the networks because the large number of images were provided in one epoch in training. The mini-batch size was 20. The size of all the images used in the proposed method was 256×256 pixels.

We performed training of the networks on an NVIDIA Quadro P6000 because major amounts of GPU memory were necessary to load images in mini-batches. After the training, we performed image-domain translations using the trained networks on an NVIDIA TITAN V.

To confirm how the data cleansing process contributes to improve the quality of the image-domain translation results, we trained the image-domain translator using following two datasets.

Cleansing: 13,406 real colonoscopic images (data cleansing was applied), and 8072 virtual colonoscopic images.

No cleansing: 18,775 real colonoscopic images (data cleansing was not applied) and 8072 virtual colonoscopic images.

We compared the two image-domain translators trained using the two datasets.

4. Results: By using the trained image-domain translator trained using the ‘Cleansing’ dataset, we converted virtual colonoscopic images. Processing times of training and inferences are shown in Table 1. Virtual colonoscopic images and corresponding image-domain translation results generated using four network structures are shown in Fig. 4. Image-domain translation results of consecutive images in virtual colonoscopic videos are shown in Fig. 5.

As shown in Fig. 4, virtual colonoscopic images were translated to realistic images. Textures and light reflections on the colon surfaces in the translated images were similar to real colonoscopic images. The arrows in Fig. 4 indicate unnatural parts in the translated images. As shown in Fig. 5, the translated images generated from consecutive images in videos have temporal smoothness.

The translated images generated using the deep U-Net and the U-Net having residual units were quite similar to real colonoscopic images. Textures and light reflections (specular) on the colon surfaces were resembled to real images. The translated images generated using the U-Net were also similar to real images. However, an unnatural part was observed in the images (indicated by an arrow in Fig. 4). In the translated images generated using the shallow U-Net, many unnatural parts were observed (indicated by arrows in Fig. 4). Furthermore, textures and light reflections in these images were close to those of the virtual colonoscopic images.

We also converted virtual colonoscopic images by using the image-domain translator trained using the ‘No cleansing’ dataset. The U-Net was used as the image-domain translator. Virtual colonoscopic images and corresponding image-domain translation results are shown in Fig. 6. Also, other image-domain translation results are shown in Fig. 7.

Table 1 Processing times of training and inference. Trainings were performed on a NVIDIA Quadro P6000 and inferences were performed on a NVIDIA TITAN V

Network structure	Training	Inference (1 image)
shallow U-Net	31 h 30 min	0.0039 s
U-Net	34 h 44 min	0.0042 s
deep U-Net	35 h 40 min	0.0046 s
U-Net having residual units	46 h 47 min	0.0052 s

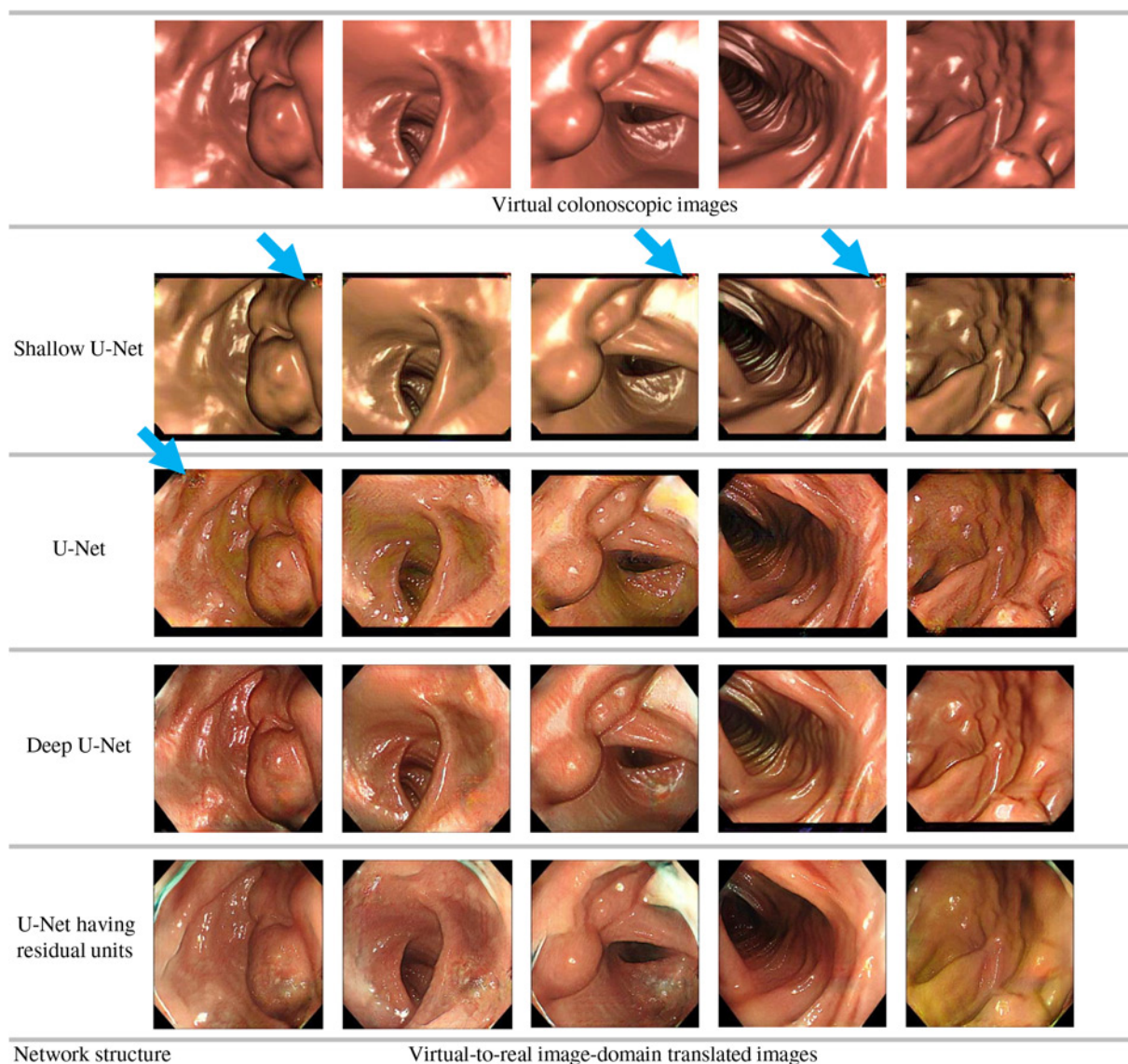


Fig. 4 Top row shows virtual colonoscopic images. These images were given to image-domain translator trained using 'Cleansing' dataset. Remaining rows show image-domain translation results obtained using shallow U-Net, U-Net, deep U-Net, and U-Net having residual units. Unnatural parts in translated images are indicated by arrows

As shown in Figs. 6 and 7, many unnatural parts such as black spots, blue haustral folds, and green textured scenes were observed in the translated images.

5. Discussion: We obtained very realistic images using the deep U-Net and the U-Net having residual units as image-domain translators trained using the 'Cleansing' dataset. Textures and light reflections on the colon surfaces in these images were similar to those of real colonoscopic images. Current endoscopic simulation systems have a problem with their visualisation. Their visualisations are not realistic because of the regardless of real textures of the organ surfaces and lack of the light reflection/absorption property information of the organ surfaces. Current simulation systems improve reality of visualisation using the texture mapping technique. However, making good texture images for the texture mapping is difficult. Our proposed realistic endoscope image generation method reduces the problems of the current simulation systems. Realistic textures on the colon surfaces in the translated images greatly improved image reality. The realistic textures were automatically generated. Also, the appearance of light reflections in the translated images is quite similar to that in the real colonoscopic images. Furthermore, the image-domain translators

generated 'black corners' that commonly appear in real endoscopic images. The image-domain translator learned to generate the 'black corners' from real endoscopic images. The proposed method brings great improvement of reality on visualisation of endoscopic simulation systems. Simulation systems having the realistic visualisation scheme greatly contribute skill improvement of surgeons or physicians in training or rehearsals of endoscope manipulation.

The translated images generated using the shallow U-Net were not realistic. Textures and light reflections on the colon surfaces in these images were not changed so much from those in virtual colonoscopic images. That is because the layer number of the shallow U-Net was not enough to represent the change of appearances between virtual and real images. Furthermore, black noises were observed in the translated images (black noises are indicated by arrows in Fig. 4). The black noises were caused in the decoding process of images from feature values in the shallow U-Net. The shallow U-Net cannot recover spatial information sufficiently in the decoding process. Because of the lack of spatial information, parts of the 'black corners' were generated at wrong positions in the translated images. They were observed as black noises. The deeper network works well as the image-domain translator.

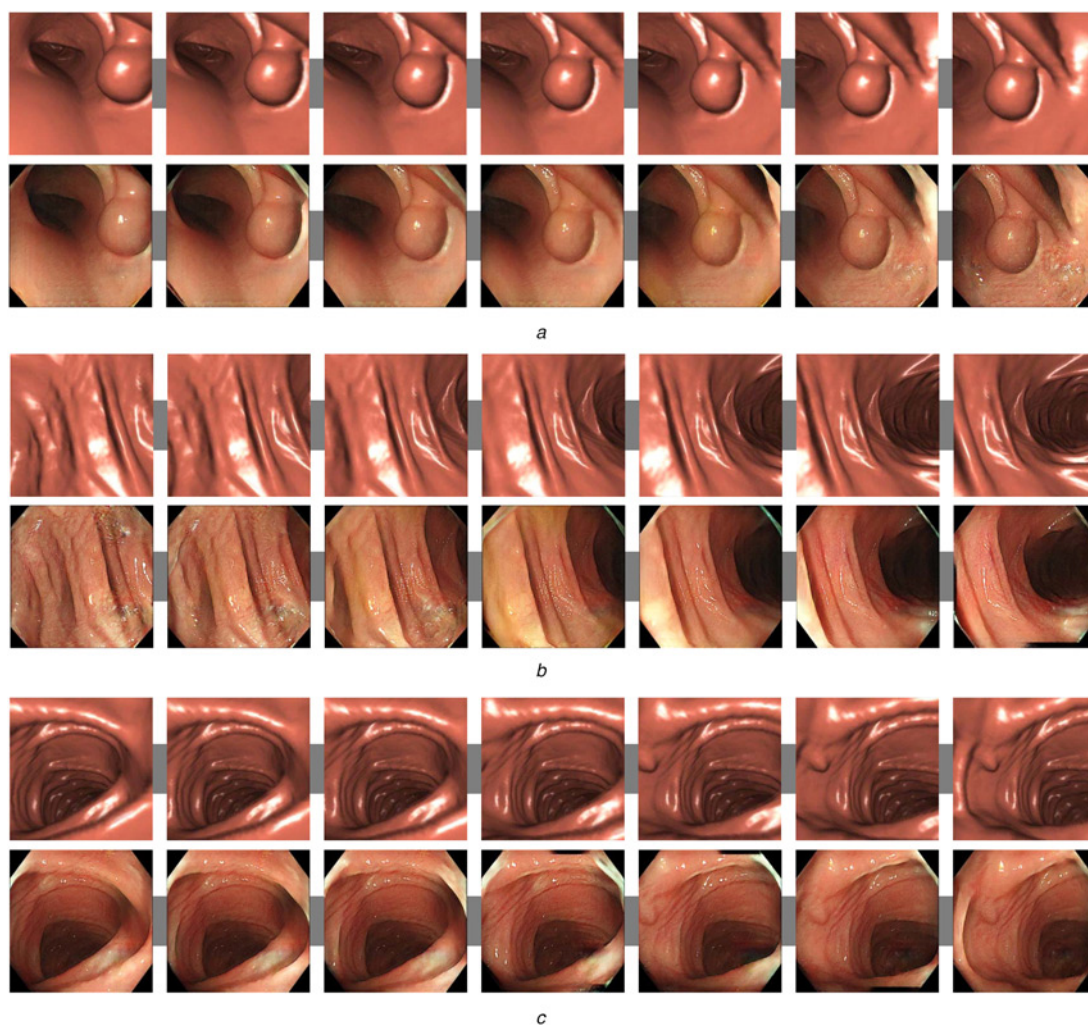


Fig. 5 Image-domain translation results of consecutive images in virtual colonoscopic videos. Top and bottom rows show virtual colonoscopic images and corresponding image-domain translation results using U-Net having residual units
a, b, c show different scenes in videos

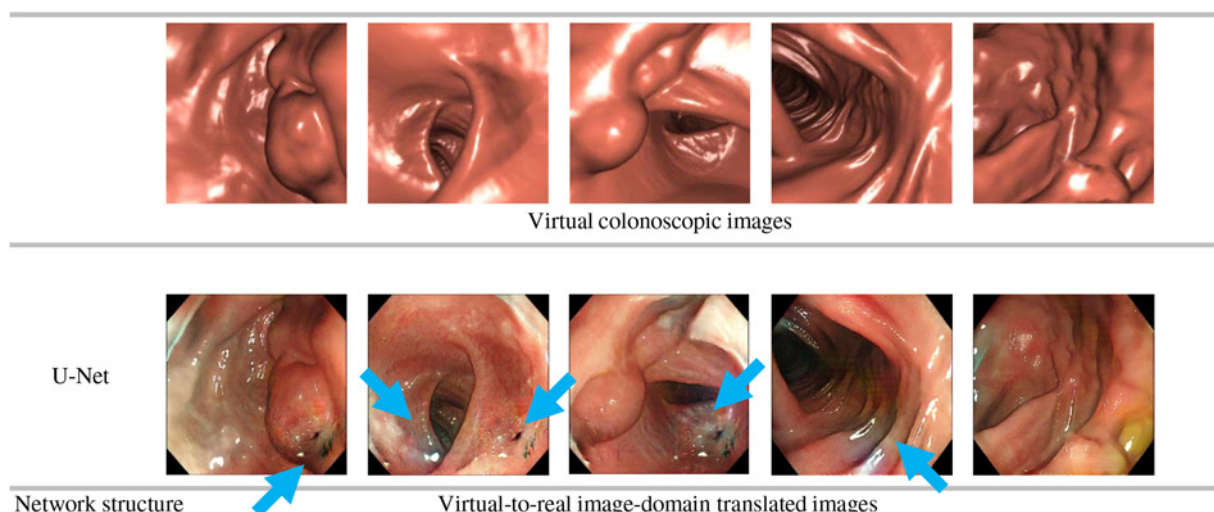


Fig. 6 Top row shows virtual colonoscopic images. They were given to image-domain translator trained using 'No cleansing' dataset. Bottom row shows image-domain translation results obtained using U-Net. Unnatural parts in translated images are indicated by arrows

We performed image-domain translation of consecutive images in virtual colonoscopic videos. As shown in Fig. 5, the image-domain translator generated images having temporal smoothness. This result is important when we utilise the proposed method as a

visualisation of endoscopic simulation systems. Even though the proposed method employs single 2D image-based process, the results have temporal smoothness. The U-Net having residual units successfully recovers spatial information in each image. It

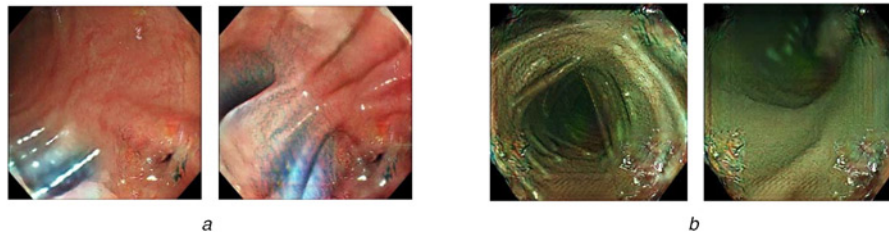


Fig. 7 Generated images by image-domain translator trained using 'No cleansing' dataset. U-Net is used as image-domain translator
 a Contains blue haustral folds. Haustral folds were confused with blue surgical tools (sample of surgical tool is shown in Fig. 3)
 b Shows green textured scenes. They were affected by narrow-band images (sample of image is shown in Fig. 3)

also resulted in generation of images having temporal smoothness. Furthermore, as shown in Table 1, inference times of the image-domain translators were rapid enough to process images in real-time. It means the proposed method can be used for real-time visualisation of simulation systems.

The data cleansing process contributed to improving the quality of image-domain translation. As shown in Figs. 6 and 7, the translated images had many unnatural parts such as black spots, blue haustral folds, and green textured areas. They were not realistic compared to the translation results generated by using the 'Cleansing' dataset. Images removed by the data cleansing process contain black endoscopes, blue surgical tools, and narrow-band images. Use of such images in training resulted in generations of black spots, blue haustral folds, and green parts. The data cleansing process is important to obtain better image-domain translation results.

Commonly, in the training of GANs and CycleGAN, mode collapse problem occurs sometimes. If this problem occurs, the generator network produces a limited variety of images. In our experiments, we had no mode collapse problem. In real colonoscopic images, there is no obvious difference of appearances among images taken by different endoscopists and images taken during different colonoscope insertions. It means feature distribution of the images is less biased. The images are less likely to cause mode collapse problem when they are used in training of GANs. Also, we used the relatively large mini-batch size in training. Parameters in the networks were updated using the averaged feedback of images in a mini-batch. It reduces convergence of the network parameters under effect of specific training image. The use of large mini-batch size might be contributed to reducing mode collapse problem.

6. Conclusions: We proposed a realistic endoscopic image generation method for endoscopic simulation systems. Current endoscopic simulation systems generate less-realistic images. Our proposed method generates realistic virtual images. We generate virtual endoscopic images of a patient from a CT volume. The virtual endoscopic images are translated to realistic images using a virtual-to-real image-domain translation technique. The image-domain translator is implemented as an FCN. The imagedomain translator is trained to minimise the cycle consistency loss. We employed the shallow U-Net, U-Net, deep U-Net, and U-Net having residual units as the image-domain translators. Experiments were performed to confirm relationships between the network structures of the image-domain translator and the quality of translated images. The deep U-Net and U-Net having residual units generated quite realistic images. However, the shallow U-Net could not increase reality of images. The deeper network is necessary to perform virtual-to-real image-domain translation.

Future work includes quantitative evaluation of the results, applications to other organs, utilisation of other network structures as an image-domain translator, and introduction of temporal information in image-domain translation.

7. Funding and Declaration of Interests: Parts of this research were supported by the AMED grant nos. 18lk1010028s0401, 19lk1010036h0001, and 19hs0110006h0003, the MEXT/JSPS KAKENHI grant nos 26108006, 17H00867, and 17K20099, and the JSPS Bilateral International Collaboration Grants.

8 References

- [1] Miki T., Iwai T., Kotani K., *ET AL.*: 'Development of a virtual reality training system for endoscope-assisted submandibular gland removal', *J. Cranio-Maxillofacial Surg.*, 2016, **44**, (11), pp. 1800–1805
- [2] Khan R., Plahouras J., Johnston B.C., *ET AL.*: 'Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy', *Cochrane Database Syst. Rev.*, 2018, **8**, (8)
- [3] Triantafyllou K., Lazaridis L.D., Dimitriadis G.D.: 'Virtual reality simulators for gastrointestinal endoscopy training', *World. J. Gastrointest. Endosc.*, 2014, **6**, (1), pp. 6–12
- [4] Harpham-Lockyer L., Laskaratos F.-M., Berlingieri P., *ET AL.*: 'Role of virtual reality simulation in endoscopy training', *World. J. Gastrointest. Endosc.*, 2015, **7**, (18), pp. 1287–1294
- [5] Shrivastava A., Pfister T., Tuzel O., *ET AL.*: 'Learning from simulated and unsupervised images through adversarial training'. IEEE Conf. CVPR, Honolulu, USA, 2017
- [6] Engelhardt S., Simone R.D., Full P.M., *ET AL.*: 'Improving surgical training phantoms by hyperrealism: deep unpaired image-to-image translation from real surgeries'. Int. Conf. MICCAI, Granada, Spain, 2018, vol. 11070, pp. 747–755
- [7] Gil D., Esteban A., Sanchez C., *ET AL.*: 'Enhancing virtual bronchoscopy with intra-operative data using multi-objective GANs', *Int. J. CARS*, 2019, **14**, (1), pp. S7–S8
- [8] Esteban-Lansaque A., Sanchez C., Borrás A., *ET AL.*: 'Augmentation of virtual endoscopic images with intra-operative data using content-nets'. bioRxiv 681825, 2019
- [9] Zhu J.-Y., Park T., Isola P., *ET AL.*: 'Unpaired image-to-image translation using cycle-consistent adversarial networks'. arXiv:1703.10593, 2017
- [10] Huo Y., Xu Z., Bao S., *ET AL.*: 'Adversarial synthesis learning enables segmentation without target modality ground truth'. IEEE 15th Int. Symposium ISBI, Washington D.C., USA, 2018
- [11] Wolterink J.M., Dinkla A.M., Savenije M.H.F., *ET AL.*: 'Deep MR to CT synthesis using unpaired data'. Int. Workshop SASHIMI, Quebec, Canada, 2017, vol. 10557, pp. 14–23
- [12] Hiasa Y., Otake Y., Takao M., *ET AL.*: 'Cross-modality image synthesis from unpaired data using CycleGAN'. Int. Workshop SASHIMI, Granada, Spain, 2018, vol. 11037, pp. 31–41
- [13] Mori K., Suenaga Y., Toriwaki J.: 'Fast software-based volume rendering using multimedia instructions on PC platforms and its application to virtual endoscopy'. Proc. SPIE Medical Imaging, San Diego, USA, 2003, vol. 5031, pp. 111–122
- [14] Goodfellow I., Pouget-Abadie J., Mirza M., *ET AL.*: 'Generative adversarial nets'. NIPS, Montreal, Canada, 2014
- [15] Ronneberger O., Fischer P., Brox T.: 'U-Net: convolutional networks for biomedical image segmentation'. Int. Conf. MICCAI, Munich, Germany, 2015, vol. 9351, pp. 234–241
- [16] Ulyanov D., Vedaldi A., Lempitsky V.: 'Instance normalization: the missing ingredient for fast stylization'. arXiv:1607.08022, 2017
- [17] He K., Zhang X., Ren S., *ET AL.*: 'Deep residual learning for image recognition'. IEEE Conf. CVPR, Las Vegas, USA, 2016