

**Open Access**

Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy

Joonmyeong Choi¹, Keewon Shin¹, Jinhoon Jung², Hyun-Jin Bae², Do Hoon Kim³, Jeong-Sik Byeon³ and Namkug Kim^{1,4}¹Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, ²Promedius, Inc., Seoul, ³Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, ⁴Department of Radiology, Asan Medical Center, Seoul, Korea

Recently, significant improvements have been made in artificial intelligence. The artificial neural network was introduced in the 1950s. However, because of the low computing power and insufficient datasets available at that time, artificial neural networks suffered from overfitting and vanishing gradient problems for training deep networks. This concept has become more promising owing to the enhanced big data processing capability, improvement in computing power with parallel processing units, and new algorithms for deep neural networks, which are becoming increasingly successful and attracting interest in many domains, including computer vision, speech recognition, and natural language processing. Recent studies in this technology augur well for medical and healthcare applications, especially in endoscopic imaging. This paper provides perspectives on the history, development, applications, and challenges of deep-learning technology. **Clin Endosc** 2020;53:117-126

Key Words: Artificial intelligence; Convolutional neural network; Deep learning; Endoscopic imaging; Machine learning

INTRODUCTION

Artificial intelligence (AI) is the software or systems that mimic the intellectual tasks performed by the human brain, using computers to understand the natural language used by humans, make logical inferences, and learn from past experience. The terminology “AI” was first used at the Dartmouth Conference in 1955, where the participants included computer pioneers, scientists, and cognitive psychologists. The proposal for this conference included an assertion: “every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.”¹ Today, almost six decades later, AI is characterized by its perception

and cognitive ability in computer vision, including object classification, detection, segmentation, and generation (Fig. 1). All these methods have been studied for practical applications, which were intended for use in different fields.

Classification refers to classifying discrete input values into multiple predefined classes. The role of the most common classification is to classify the elements of a given set into two groups. Others are multi-class and are able to categorize an item into one of the multiple categories. Classification algorithms are applied to email spam filtering, document categorization, image recognition, handwriting recognition, and speech-to-text translation.²⁻⁷ Object detection involves recognizing the boundary of instances from a particular class in an image, which has a wide range of applications in a variety of areas, including medical image analysis, robotics, surveillance, and human-computer interaction. Current methods have succeeded in the computer vision field but have several limitations when the problem is too complex or large. Image segmentation has become an increasingly critical process in computer vision.⁸⁻¹⁴ Segmentation refers to identifying parts of the image and understanding which object they belong to, pixel by pixel. This involves dividing the visual input into

Received: February 21, 2020 **Revised:** March 10, 2020

Accepted: March 13, 2020

Correspondence: Namkug Kim

Department of Radiology, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea

Tel: +82-2-3010-6573, **Fax:** +82-2-2045-3426, **E-mail:** namkugkim@gmail.com

ORCID: <https://orcid.org/0000-0002-3438-2217>

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

various partitions with similar imaging or semantic image properties.¹⁵⁻²¹ Segmentation depends on object detection and classification. In addition, there has been remarkable progress in image generation with the emergence of generative adversarial networks (GANs).²²

With an exceptional advance in AI, almost every kind of conventional computer vision task has been addressed or replaced. Therefore, deep learning (DL) has attracted great attention in the medical imaging science and engineering communities as a promising solution for full automation, fast, and accurate image analysis. Convolutional neural networks (CNNs) and their variants with DL models have become the most preferred and widely used methods in medical image analysis.

In the main text of this paper, we started presenting the history and the state-of-the-art techniques intended for application in endoscopic imaging using DL. The rest of this paper is structured as follows. First, we summarize the history of machine learning (ML), DL, and technological advances in AI-based medical imaging. Next, we review the technological

advances of AI in endoscopic imaging such as classification, detection, segmentation, and generation. Lastly, we discuss the perspective and limitations of AI in endoscopic imaging in more detail. This discussion includes a brief overview of medical imaging with AI, general facts and figures, and a detailed review of the proposed methods found in the literature.

MACHINE LEARNING AND DEEP LEARNING

Rise of AI

ML and DL are two subsets of AI that have recently garnered substantial attention worldwide. ML is a set of algorithms that learn a predictive model from data that enables predictions to be made without explicit programmed instructions for the task. DL is a specialized ML method and refers to algorithms inspired by the structure and function of a brain, called artificial neural networks (ANNs).^{23,24} DL is currently gaining considerable attention for its usefulness with big data

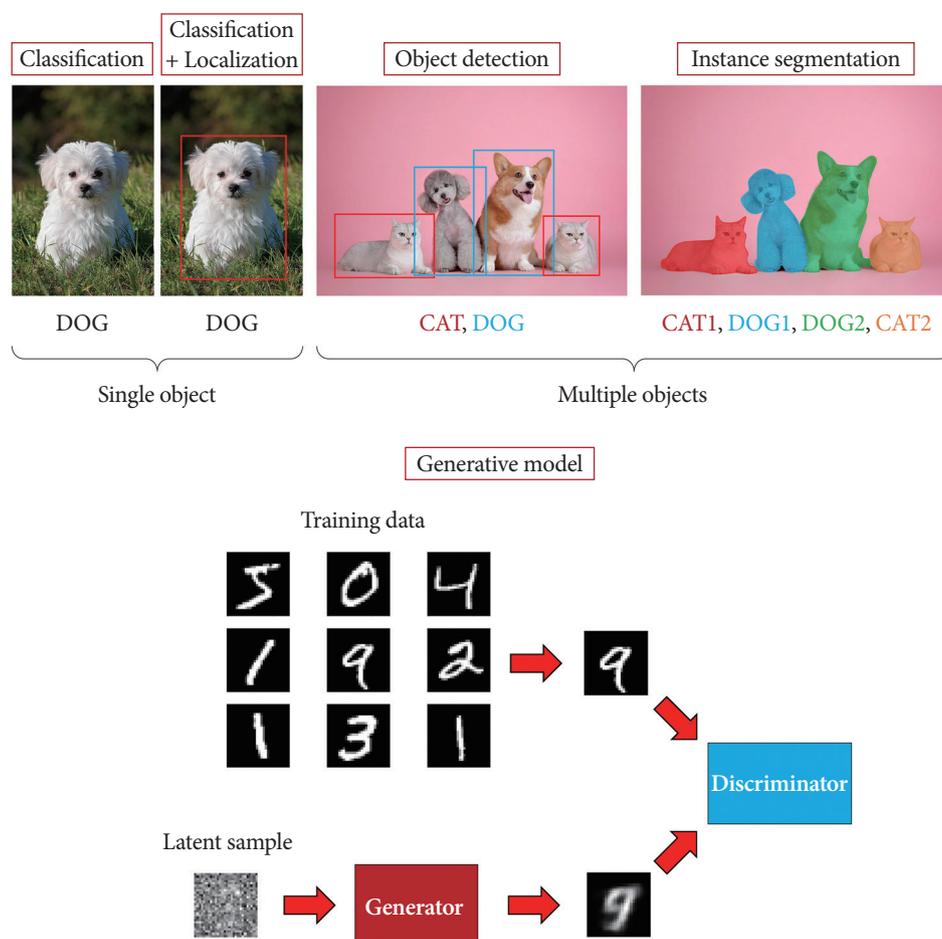


Fig. 1. Artificial intelligence in computer vision can be categorized as classification, detection, segmentation, and generation.

in healthcare. The concept of ANNs was first introduced in the 1950s; however, there were critical limitations that needed to be overcome over the following 50 years to cope with real-world problems. From the concept of the perceptron to CNNs, AI researchers have had several breakthrough moments during those 50 years (Fig. 2). In 1957, Rosenblatt, a psychologist, proposed the concept of the perceptron.²⁵ This is a one-layer neural network with a very simple configuration, but at the time, a room full of punch-card-type computers were required to realize it. Rosenblatt's perceptrons were first simulated on an IBM 704 computer at the Cornell Aeronautical Laboratory.²⁶ In 1960, Widrow and Hoff suggested the stochastic gradient descent method from the output error in a two-layer neural network without a hidden layer, under the name of the Widrow–Hoff delta rule method.²⁷ In 1967, Amari developed a neural network with three or more layers with a hidden layer.²⁸ In 1969, Minsky et al. published a paper on the limitations of neural networks.²⁹ Ironically, owing to their reputation, the study of neural networks had suffered. However, LeCun et al. developed a neural network architecture for recognizing hand-written digits using CNNs.³⁰ This was a basic cornerstone in the field which has been essential in making DL more efficient. They proposed the concept of learning the convolution kernel coefficients using a back-propagation algorithm directly from images of hand-written numbers.³⁰

In this way, learning became fully automated and performed better than with a manual parameter design. Subsequently, its feasibility was proved for a broader range of image recognition problems. In 2006, Hinton et al. published a paper on how to efficiently train multi-layer neural networks.³¹ This method, the auto-encoder, became the basis for explosive research in DL. In 2011, Microsoft began to use neural networks for language recognition. Since then, language recognition and machine translation, as well as image recognition, have become application fields of DL. In July 2012, Google Brain was initially established by Jeff Dean and Andrew Ng. Their main interest was neural network research for recognizing cats and making use of it for language recognition. In October of the same year, two students of Hinton won the ImageNet challenge, defeating the runners-up by a significant margin, by using a DL method.⁵ In December 2015, a Microsoft Corporation team outperformed humans in the ImageNet contest, and in March 2016,³² AlphaGo defeated Sedol Lee, the Go World Champion.³³ With conceptual and engineering breakthroughs including big data, large computing power, and new algorithms that have made deep neural networks (DNNs) a critical component of computing, DL is now a common technology in computer science, robot vision, autonomous driving, and medical imaging.

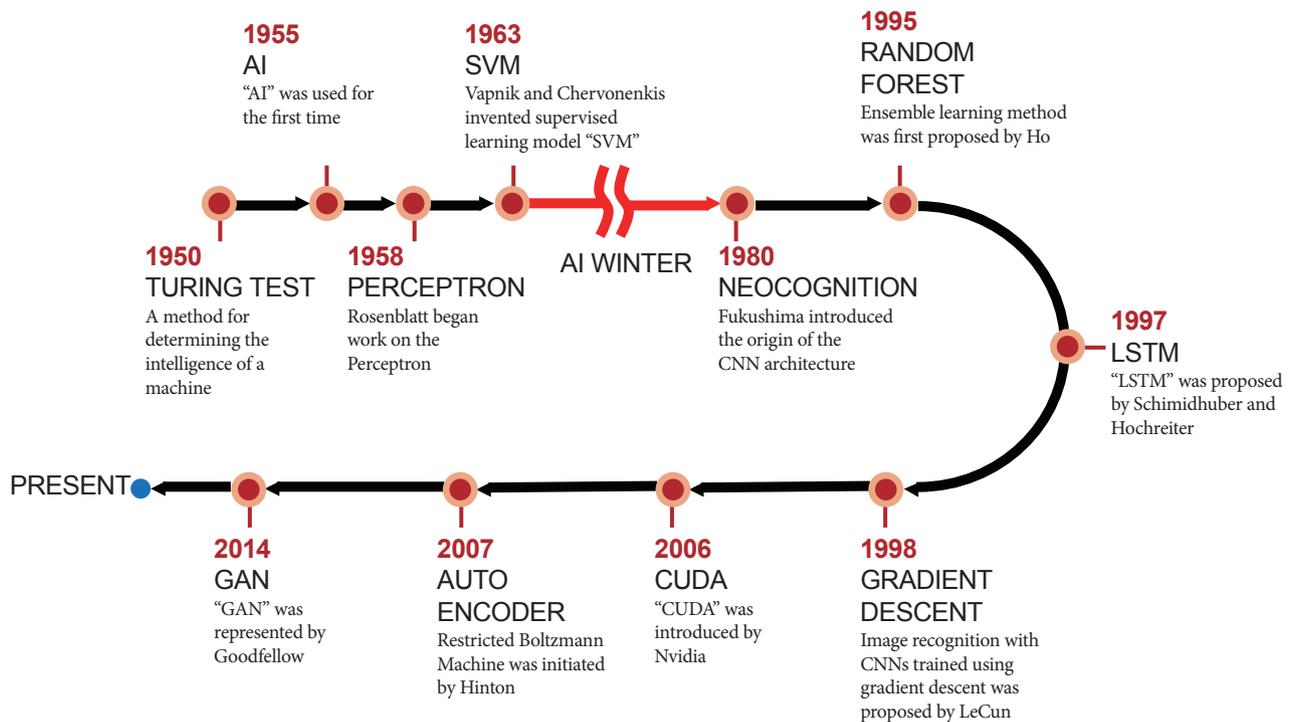


Fig. 2. Brief history of artificial intelligence. AI, artificial Intelligence; CNN, convolutional neural network; CUDA, compute unified device architecture; GAN, generative adversarial network; LSTM, long short term memory; SVM, support vector machine.

Deep neural network

DNNs are a special case of ANN, constructed by stacking layers of neural networks.³⁴ DNNs have recently been shown to have learning capabilities in a wide range of applications. ML researchers are expanding their DL horizons by investigating future applications in a variety of other areas. Deep networks require large amounts of annotated data for training. Since the 1980s, several ML algorithms for classification tasks have been developed for a variety of implementations, mathematical bases, and logical models. A DNN is an ANN with multiple layers (called hidden layers) between the input and output layers. This is an advantage when working with com-

plex data modeling with fewer nodes. In 2012, Google tried to create a human brain using a network of 16,000 computers based on the concept of a DNN. The network performed well with 84% accuracy, 10% higher than that achieved in the previous year in the ImageNet Large Scale Visual Recognition Challenge 2012. However, for image data CNNs are more accurate with less computation power required compared to DNNs.

Convolutional neural network

A CNN is used for analyzing images. Convolutional layers convolve the input and pass its result to the next layer. This

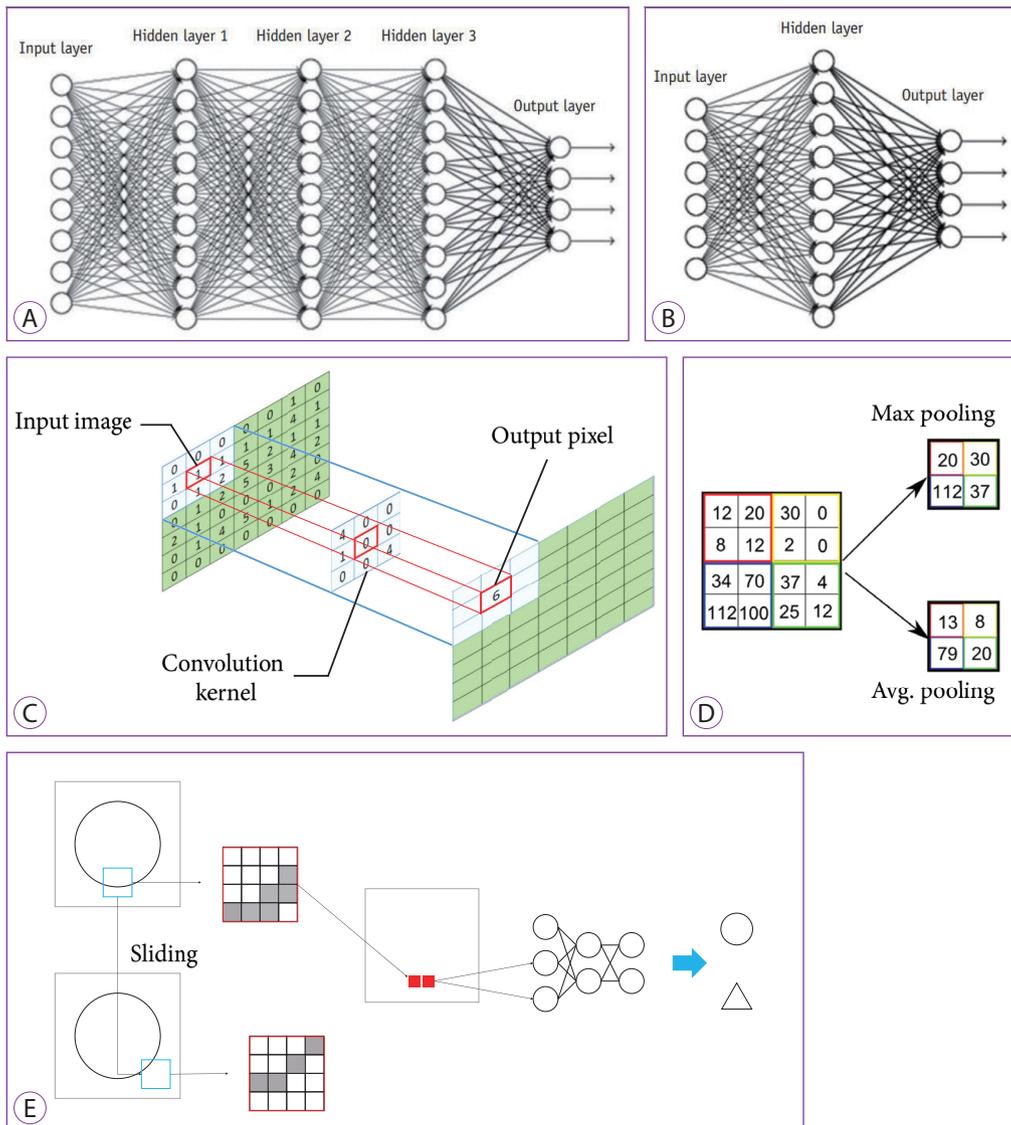


Fig. 3. Concept diagrams of deep learning (DL) and convolutional neural networks. (A) Typical DL neural network with three deep layers between input and output layers. (B) Typical artificial neural network with one layer between input and output layers. (C) Convolution method. (D) Max and average pooling methods. (E) The workflow of a convolutional neural network with one convolutional layer and one max pooling layer. Each pixel (red rectangle) of a region of interest (ROI, blue rectangle) extracted from an image are input to the neural network with the two classes of a circle and a triangle. Moving ROIs in the image were pooled with the maximum value (solid red rectangle).

was inspired by the response of a neuron in the human visual cortex to a specific stimulus.³⁵ A well-trained CNN consists of a hierarchy of information such as an edge, a corner, a part of an object, and a structure of an object in image classification.³⁶ A single CNN architecture consists of a series of convolutional layers, pooling layers, followed by a fully connected layer. The main purpose of a convolutional layer is to extract learnable features such as characteristic local motifs from images. The parameters of a special filter operator, called a convolution, are trained, and the mathematical operation takes two inputs, such as an image matrix and a kernel. By learning consequential kernels, the visual features can be effectively extracted. This is like the mechanism of a visual cortex. With the use of a filter bank, where each filter is a squared mesh that moves over the original image, the convolutional process can be performed. The moving grid image (pixel value) is summed using the filter weights, and multiple filters in convolutional layers are applied to generate multiple functional maps. Convolution, an important component of CNN, is essential to the success of image processing tasks. Pooling layers with the maximum or average are used to effectively reduce the functional map size. They also preserve the object shapes and location of the semantic features detected in the image. Therefore, pooling makes the convolutional layer less susceptible to small shifts or distortions of the object. In most cases, maximum pooling is used empirically. Inserting a pooling layer periodically between successive convolutional layers of the CNN architecture is common. After several convolutional and pooling layers, the high-level inference of the neural network takes place through fully connected layers, and all functional responses are integrated from the entire image to generate the final result.

Recently, deep CNN architectures (Fig. 3A) have been developed and compared to the traditional ANN (Fig. 3B). The parameters of a convolution are learned, which represents the multiplication of the local neighborhood of each pixel by a small array of learned parameters (kernels) (Fig. 3C). By pooling pixels together to capture an increasingly large field of view, the functional map is progressively reduced spatially (Fig. 3D). A simple task to determine whether a written figure is a circle

or triangle with a CNN is shown in Fig. 3E. The figure shows the black part of the edge of the circle passes as input. A small area called a filter (4×4 area in the red rectangle) is extracted from the image; this is compressed as a single feature by pooling, and the process is repeated while sliding the area. As a result, an input image of a fully connected layer is created.

CNNs generated significantly better classification in large image repositories, such as ImageNet, than traditional ANNs. As a deep CNN architecture typically includes many neural network layers, there can be a tremendous number of hyperparameters to estimate. Training and tuning the models requires many data samples. For example, in the ImageNet dataset, training a DL architecture from scratch in a class requires more than 1,000 cases per class. Generally, the minimum data size requirements depend on the intended use and information of the medical image, which must be investigated individually. However, there are other ways to circumvent data size standards, including the reuse of pre-trained networks with transfer learning and data augmentation techniques, which deliver reasonable results with approximately 100 cases for each class.³⁷

The biggest drawback of a CNN is the length of training time needed in addition to the complex implementation because training a CNN is computationally intensive and data intensive. Training with large datasets can take days or even weeks. Due to the large number of floating-point operations and the relatively low rate of data transfer in all of the training steps, this task can only be done with a graphics processing unit (GPU). GPUs are a major technical advance in the computer industry that work with big data and enhanced computing power. The main advantage of a GPU over a central processing unit (CPU) is the relatively low cost and high computational throughput achieved by the large-scale parallel architecture. This reduces training time, speeds recruitment, and increases efficiency.

Computer-aided detection/diagnosis

Several computer-aided detection (CAD) and diagnosis systems in radiology were developed and introduced into clinical workflows in the early 2000s. Unlike initial expect-

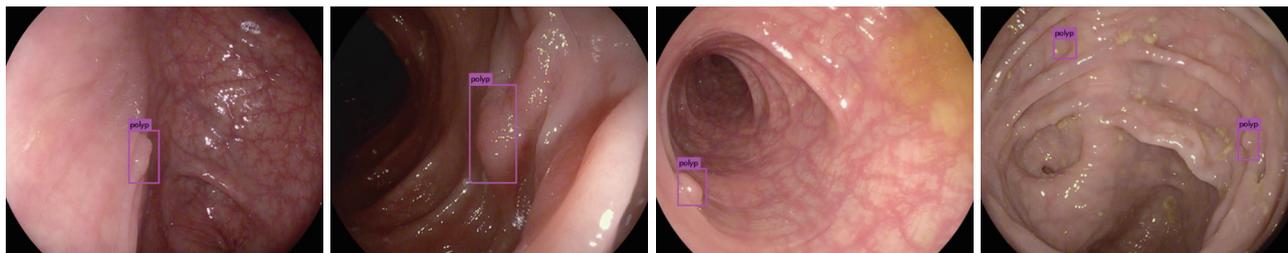


Fig. 4. Examples of polyp detection with convolutional neural network in colonoscopy.

tations, some shortcomings of the CAD systems have been revealed through several studies.^{38,39} Key among these was that the CAD systems performed poorly compared to expert radiologists, resulting in longer evaluation times and additional biopsies. Therefore, there was little or no net profit from using CAD. Modern DL technology can help increase productivity by overcoming the technical difficulties of the early CAD systems, increasing detection accuracy, and enabling physicians to use AI for mediocre repetitive tasks. DL is recognized as the most suitable method for analyzing big data to extract meaningful information in healthcare. AI has now emerged as a promising technique for differential diagnosis, automatic lesion detection, and the generation of preliminary reports.

TECHNOLOGICAL ADVANCES OF AI IN ENDOSCOPIC IMAGING

AI-based classification in endoscopy

Recently, CNNs have been widely studied in the field of endoscopy with promising results, including esophagogastroduodenoscopy, colonoscopy, and wireless capsule endoscopy (WCE). The initial study of classification in endoscopic imaging reported that the classification of the mucosal pit patterns by magnifying the endoscopy not only corresponded to the transition from a gastric to an intestinal phenotype in Barrett's epithelium but was also associated with an increase in cell cycles.⁴⁰ With the initial success in endoscopy, the classification of features has proved to be useful for the endoscopic diagnosis of Barrett's esophagus. In addition, finding events such as entering the next digestive organs and detecting active bleeding is of major interest to the endoscopists reviewing WCE videos. To reduce the long review time, an algorithm for event boundary detection in WCE videos was proposed.⁴¹ An energy-based extraction of the proposed event detection feature showed a recall of 76% and a precision of 51%. The performance of diagnosis methods for early gastric cancers was investigated based on mucosal surface pattern classification in enhanced-magnification endoscopy by comparing them with conventional magnification endoscopy (magnification chromoendoscopy). In their study, mucosal surface patterns were classified into five types. The study indicated that the classification of types IV–V was strongly correlated with the presence of early gastric cancer (sensitivity of 100% and specificity of 89.7%).⁴² Stehle et al. proposed a multi-stage system for the classification of colon polyps.⁴³ They achieved a correct classification rate of 90% using their datasets, including 56 polyps with histologically confirmed ground truth, which corresponds to a sensitivity of 91.9% and a specificity of 84.2%. Recently, we reported that our DL model classifying

the mucosal surface patterns of colon polyps achieved an overall diagnostic accuracy of 81.3%–82.4% in real-time histological diagnosis of colon polyps, which was comparable with that of expert endoscopists (82.4%–87.3%) and higher than that of trainee endoscopists (63.8%–71.8%) (Fig. 4).⁴⁴ The transfer of learning in CNNs showed great potential for celiac disease classification based on endoscopic images.⁴⁵ The authors used four CNNs that were pre-trained on the ImageNet database. Three different transfer learning strategies were explored to classify the endoscopic images of the celiac disease. Full fine-tuning of the CNNs achieved the highest classification accuracies, although the small amount of training data available led to overfitting. As a result, the VGG-F network (an eight-layer deep CNN) showed the best performance for celiac disease classification. The fully fine-tuned VGG network outperformed the four state-of-the-art image representations. Transfer learning for low-level CNN features from various non-medical source tasks using deep CNN representations showed good performance for the automatic detection and classification of colorectal polyps.^{44,46,47} The proposed diagnostic model, which minimized the time-consuming pre-processing, outperformed the previous state-of-the-art methods.

AI-based detection in endoscopy

Screening colonoscopy is one of the most explored domains in the AI-based detection field, as the detection of colorectal polyps without missing any is of paramount importance in colorectal cancer screening. A previous study showed that the risk of interval colorectal cancer decreased by 3.0% as the adenoma detection rate increased by 1.0%.⁴⁸ Despite the importance of polyp detection, the missing rate for polyps is still high (6%–27%).⁴⁹ There were several recent studies that suggested a DL approach to decrease the adenoma miss rate.^{50–52} They showed that it is possible to detect polyps in real-time with reasonable accuracy. For instance, Zhang et al. achieved an 84% F1-score at speeds of 50 frames-per-second with a modified Single Shot MultiBox Detector.^{12,53} In addition, a study by Urban et al.⁵¹ suggested that a CNN is capable of detecting polyps regardless of their morphological type (polypoid or nonpolypoid): in general, a nonpolypoid polyp is challenging to detect because of its shape. Zheng et al. proposed a real-time polyp detector based on You-Only-Look-Once in endoscopic images.⁵²

Another promising area for CNN technology applications is WCE.⁵⁴ It is more comfortable and less invasive than conventional endoscopy with a light tube.⁵⁵ However, analyzing the video from WCE is challenging because of the long review time and the low quality of the images. Therefore, an automated detection and classification system is required, and CNNs can be a solution. Fan et al.⁵⁶ explored this possibility and they

developed a system that detected small intestinal ulcers and erosion in WCE images.⁵⁷ They made use of the popular CNN architecture, AlexNet,⁵ and achieved a high accuracy of approximately 95%.

AI-based segmentation in endoscopy

Polyp segmentation had been considered a relatively simple task, as the automatic algorithm only needed to analyze certain polyp frames to find and localize the existing polyps. An earlier study on segmentation in endoscopic imaging in 2000 was initiated through automatic segmentation of the colon for virtual colonoscopy.⁵⁸ It was not an AI-based study but was meaningful as a computer vision method for automatic segmentation. The algorithm was designed to segment the colon from computer tomography volumes based on the features of colon geometry. A decade later, polyp segmentation in narrow band imaging colonoscopy was proposed in 2009.⁵⁹ At this stage, the segmentation task for endoscopic imaging was still dependent on conventional edge detection methods such as the canny edge detector. Therefore, noise reduction was always the greatest concern among researchers, and the first trial for polyp segmentation in WCE videos was proposed. In 2010, Hwang et al. proposed a geometric feature-based method for WCE images.⁶⁰ Gabor texture features and k-means clustering were adapted in watershed segmentation with an initial marker selection.⁶⁰ In 2010, Arnold et al. presented a method for the segmentation of specular highlights based on color-image-thresholding, nonlinear filtering, and an efficient inpainting method; a visually favorable result was obtained by altering the specular regions using their algorithm.⁶¹ The segmented regions were characterized using a curvature-based geometric feature to determine the final polyp candidates. Preliminary experiments demonstrated that the proposed method can detect polyps with 100% sensitivity and over 81% specificity. Subsequently, a similar method was attempted by other research groups.⁶²⁻⁶⁴ The first DL-based method was presented in 2018,⁶⁵ and the study focused on polyp segmentation in colonoscopy. Three models with fully convolutional architectures were trained for three epochs by using 342 color images of a colon polyp. All images were obtained from the CVC-Colon database (Machine Vision Group, Computer Vision Centre, Barcelona, Spain) and validated using 38 images from the same database. A low dice score may occur when there is a heavy class imbalance, such as in this dataset, where ~1% of pixels are in the polyp class compared to ~99% in the background class. A similar approach was reported by Brandao et al.⁶⁶ They obtained a relatively high segmentation accuracy and a detection precision and recall of 73.61% and 86.31%, respectively. Consequently, polyp segmentation in endoscopy using DL remains a challenging issue.

Generative model for endoscopy

Many applications using CNNs have been introduced in medical imaging, but it is often difficult to obtain high-quality, balanced datasets in the medical domain.⁶⁷ To overcome this problem, several studies have used GANs to create high-quality synthetic medical images.^{68,69} The GAN has excellent strengths in synthesis showing very impressive results associated with generated content and realistic visual content. In addition, it can generate information that we lack. There are several applications that apply these two advantages of GANs to endoscopies, such as image augmentation, domain adaptation, and 3D reconstruction, to help diagnose the endoscope. Previous studies have shown that the synthesis of lesion images can significantly improve the detection accuracy in endoscopic and capsule endoscopic images.⁷⁰⁻⁷² The gastric cancer detection model was significantly improved when a portion of the synthesized images with realistic lesions was added to the training dataset. In addition, the variation autoencoder-GAN model was used to overcome the imbalance of positive and negative datasets.⁷³ The style transfer with cycle-GAN can improve the performance of detection generalization performance.⁷⁴ As there is no depth information in endoscopy images or videos, there have been many studies on domain adaptation to actual endoscopy images by learning the depth map through virtual endoscopy. Intra-operative medical data to gather realistic virtual endoscopy has been explored using an in-depth learning approach to transfer the content preserving style of intra-operative medical data.⁷⁵ Specifically, it has been proven that the generation model outperforms while predicting depth from colonoscopy images in terms of accuracy and robustness to domain changes.⁷⁶ However, the endoscope does not have many applications, because the performance of generating endoscopic images is poor, relative to other medical modalities. In addition, as endoscopic images do not have a formal protocol or structure, the generation of endoscopic images using GAN in an endoscope is a difficult task. Nevertheless, improved GAN models will be used as tools to produce more realistic endoscopic images that can deceive the endoscopist, thereby helping to aid in computer-aided diagnosis, CAD, and endoscopy training.

DISCUSSION

At present, endoscopists suffer from a heavy workload as the number of patients increases. Such a burden makes it difficult to achieve effective diagnosis and therapy for patients. However, the new DL technology is expected to be helpful to endoscopists by enabling more accurate lesion detection, diagnosis, and prediction by providing quantitative analysis

of suspicious lesions. Moreover, it may also shorten times for certain clinical tasks such as recognition assistance with attention, automated classification, prediction, and automatic report generation, which are benefits that AI can provide to the clinical workflow. DL has already shown superior performance over humans in some audio recognition and computer vision tasks including detection and recognition. Endoscope manufacturers such as Olympus, Fuji, Pentax, and others have begun research on DL applications in endoscopic imaging. Through these technological innovations, some major changes in endoscopic practices may soon occur. When we consider the implementation of AI in endoscopic imaging, however, we predict this technological innovation will assist in a reduction of the burden and distraction from repetitive and labor-intensive tasks, rather than replacing endoscopists at least in the near future, because the use of DL and AI in endoscopy is currently in its infancy.

To develop AI and its practical applications in endoscopy further, a good mutual understanding and cooperation between both endoscopists and engineers is important. Additionally, large and fully annotated databases (e.g., ImageNet) are required to facilitate the development of AI in medical imaging. This is one of the most important issues in terms of training the DL network, as well as for its evaluation. The active participation of endoscopists is also indispensable to establishing a large medical image dataset for both training and meticulous clinical validation of a developed AI system. Finally, the ethical, regulatory, and legal issues raised while using patient clinical imaging data in the development of AI must be carefully considered.

Even with many favorable results from early studies, there are several concerns that should be addressed before the implementation of DL methods in endoscopic imaging, some of which are listed as follows. First, the dependency on training data, especially the quality and amount of it and the overfitting issue should be carefully considered. Regarding the variations in disease prevalence, medical imaging modalities, and practice patterns among hospitals worldwide, we need to confirm that the trained networks are generally useful through external validation. Second, the black box nature of the current DL technique should be highlighted. Even when the DL-based method demonstrates excellent results, on many occasions, it is difficult or almost impossible to explain the technical and logical basis of the system. Additionally, legal liability issues would be raised if we were to adopt a DL system in daily endoscopy practices. Unlike our expectations, however, none of the systems can be perfect, and who or what should take responsibility for an error and misinformation should be addressed by experts from various fields, such as medicine, computer science, ethics, and law.

CONCLUSIONS

With the rapid advancement of AI technology, medical doctors, including endoscopists, must gain technical knowledge to understand the capabilities of AI and how AI can change endoscopy practices in the near future. We believe that these ML-based analytical tools will eventually be adopted in diagnosis practice. Nonetheless, this prediction does not directly imply the complete replacement of medical doctors. This “selective replacement” is not really the ultimate replacement but a strong complement to irreplaceable and incredible human skills; thus, it can boost overall effectiveness.

Conflicts of Interest

The authors have no financial conflicts of interest.

ORCID

Joonmyeong Choi: <https://orcid.org/0000-0003-3420-7125>
 Keewon Shin: <https://orcid.org/0000-0002-5028-5716>
 Jinhoon Jung: <https://orcid.org/0000-0003-3255-1774>
 Hyun-Jin Bae: <https://orcid.org/0000-0001-5134-5517>
 Do Hoon Kim: <https://orcid.org/0000-0002-4250-4683>
 Jeong-Sik Byeon: <https://orcid.org/0000-0002-9793-6379>

REFERENCES

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Mag* 2006;27:12-14.
2. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273-297.
3. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 1997;55:119-139.
4. Perronnin F, Sánchez J, Mensink T. Improving the Fisher kernel for large-scale image classification. In: 11th European Conference on Computer Vision; 2010 Sep 5-11; Heraklion, Greece. Berlin: Springer Berlin Heidelberg; 2010. p. 143-156.
5. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25 (NIPS 2012) 2012:1097-1105.
6. Wu Z, Wang X, Jiang YG, Ye H, Xue X. Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. In: *Proceedings of the 23rd ACM International Conference on Multimedia*; 2015 Oct; Brisbane, Australia. New York (NY): Association for Computing Machinery; 2015. p. 461-470.
7. Jia Y, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding. *ArXiv.org* 2014. Available from: <https://arxiv.org/abs/1408.5093>.
8. Kobatake H, Yoshinaga Y. Detection of spicules on mammogram based on skeleton analysis. *IEEE Trans Med Imaging* 1996;15:235-245.
9. Sung K, Poggio T. Example-based learning for view-based human face detection. *IEEE Trans Pattern Anal Mach Intell* 1998;20:39-51.
10. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Trans Pattern Anal Mach Intell* 2010;32:1627-1645.
11. Dollár P, Wojek C, Schiele B, Perona P. Pedestrian detection: an evaluation of the state of the art. *IEEE Trans Pattern Anal Mach Intell*

- 2012;34:743-761.
12. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. ArXiv.org 2016. Available from: <https://arxiv.org/abs/1512.02325>.
 13. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. ArXiv.org 2016. Available from: <https://arxiv.org/abs/1506.02640>.
 14. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. ArXiv.org 2018. Available from: <https://arxiv.org/abs/1708.02002>.
 15. Deng J, Dong W, Socher R, Li L, Kai L, Li FF. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20-25; Miami (FL), USA. Piscataway (NJ): IEEE; 2009. p. 248-255.
 16. Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM. Selective search for object recognition. Int J Comput Vis 2013;104:154-171.
 17. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. ArXiv.org 2014. Available from: <https://arxiv.org/abs/1311.2524>.
 18. Dai J, He K, Sun J. Instance-aware semantic segmentation via multi-task network cascades. ArXiv.org 2015. Available from: <https://arxiv.org/abs/1512.04412>.
 19. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. ArXiv.org 2018. Available from: <https://arxiv.org/abs/1703.06870>.
 20. Arnab A, Torr PHS. Pixelwise instance segmentation with a dynamically instantiated network. ArXiv.org 2017. Available from: <https://arxiv.org/abs/1704.02386>.
 21. Bolya D, Zhou C, Xiao F, Lee YJ. YOLACT: real-time instance segmentation. ArXiv.org 2019. Available from: <https://arxiv.org/abs/1904.02689>.
 22. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. ArXiv.org 2014. Available from: <https://arxiv.org/abs/1406.2661>.
 23. Chen TCT, Liu CL, Lin HD. Advanced artificial neural networks. Algorithms 2018;11:102.
 24. Chen YY, Lin YH, Kung CC, Chung MH, Yen IH. Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes. Sensors (Basel) 2019;19:E2047.
 25. Rosenblatt F. The perceptron, a perceiving and recognizing automaton: (Project Para). Buffalo (NY): Cornell Aeronautical Laboratory; 85-460-1; 1957.
 26. Marcus G. Hying artificial intelligence, yet again [Internet]. New York (NY): The New Yorker; c2013 [cited 2020 Jan 29]. Available from: <https://www.newyorker.com/tech/elements/hying-artificial-intelligence-yet-again>.
 27. Hanson SJ. A stochastic version of the delta rule. Physica D 1990;42:265-272.
 28. Amari S. A theory of adaptive pattern classifiers. IEEE Trans Comput 1967;EC-16:299-307.
 29. Minsky ML, Papert SA. Perceptrons: an introduction to computational geometry. Cambridge (MA): Institute of Technology; 1969.
 30. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1:541-551.
 31. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313:504-507.
 32. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. ArXiv.org 2015. Available from: <https://arxiv.org/abs/1512.03385>.
 33. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484-489.
 34. Fiesler E, Beale R. Handbook of neural computation. Bristol: Institute of Physics Pub; 1997.
 35. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. J Physiol 1968;195:215-243.
 36. Katole AL, Yellapragada KP, Bedi AK, Kalra SS, Chaitanya MS. Hierarchical deep learning architecture for 10K objects classification. ArXiv.org 2015. Available from: <https://arxiv.org/abs/1509.01951>.
 37. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: general overview. Korean J Radiol 2017;18:570-584.
 38. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007;356:1399-1409.
 39. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175:1828-1837.
 40. Endo T, Awakawa T, Takahashi H, et al. Classification of Barrett's epithelium by magnifying endoscopy. Gastrointest Endosc 2002;55:641-647.
 41. Lee J, Oh J, Shah SK, Yuan X, Tang SJ. Automatic classification of digestive organs in wireless capsule endoscopy videos. In: Proceedings of the 2007 ACM symposium on Applied computing; 2007 Mar; Seoul, Korea. New York (NY): Association for Computing Machinery; 2007. p. 1041-1045.
 42. Tanaka K, Toyoda H, Kadowaki S, et al. Surface pattern classification by enhanced-magnification endoscopy for identifying early gastric cancers. Gastrointest Endosc 2008;67:430-437.
 43. Stehle T, Auer R, Gross S, et al. Classification of colon polyps in NBI endoscopy using vascularization features. In: SPIE Medical Imaging; 2009 Feb 7-12; Lake Buena Vista (FL), USA. 2009.
 44. Song EM, Park B, Ha CA, et al. Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model. Sci Rep 2020;10:30.
 45. Wimmer G, Vécsei A, Uhl A. CNN transfer learning for the automated diagnosis of celiac disease. In: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA); 2016 Dec 12-15; Oulu, Finland. Piscataway (NJ): IEEE; 2016. p. 1-6.
 46. Zhang R, Zheng Y, Mak TW, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. IEEE J Biomed Health Inform 2017;21:41-47.
 47. Yoon HJ, Kim S, Kim JH, et al. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. J Clin Med 2019;8:E1310.
 48. Corley DA, Levin TR, Doubeni CA. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:2541.
 49. Ahn SB, Han DS, Bae JH, Byun TJ, Kim JP, Eun CS. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. Gut Liver 2012;6:64-70.
 50. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. Gastroenterology 2018;154:2027-2029.e3.
 51. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018;155:1069-1078.e8.
 52. Zheng Y, Zhang R, Yu R, et al. Localisation of colorectal polyps by convolutional neural network features learnt from white light and narrow band endoscopic images of multiple databases. Conf Proc IEEE Eng Med Biol Soc 2018;2018:4142-4145.
 53. Zhang X, Chen F, Yu T, et al. Real-time gastric polyp detection using convolutional neural networks. PLoS One 2019;14:e0214133.
 54. Li B, Meng MQ, Xu L. A comparative study of shape features for polyp detection in wireless capsule endoscopy images. Conf Proc IEEE Eng Med Biol Soc 2009;2009:3731-3734.
 55. Liaqat A, Khan MA, Shah JH, Sharif M, Yasmin M, Fernandes SL. Automated ulcer and bleeding classification from WCE images using multiple features fusion and selection. J Mech Med Biol 2018;18:1850038.
 56. Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys Med Biol 2018;63:165001.
 57. Alaskar H, Hussain A, Al-Aseem N, Liatsis P, Al-Jumeily D. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. Sensors (Basel) 2019;19:E1265.

58. Wyatt CL, Ge Y, Vining DJ. Automatic segmentation of the colon for virtual colonoscopy. *Comput Med Imaging Graph* 2000;24:1-9.
59. Gross S, Kennel M, Stehle T, et al. Polyp segmentation in NBI colonoscopy. In: Meinzer HP, Deserno TM, Handels H, Tolxdorff T, eds. *Bildverarbeitung für die medizin 2009*. Berlin: Springer, Berlin, Heidelberg; 2009. p. 232-256.
60. Hwang S, Celebi ME. Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing; 2010 Mar 14-19; Dallas (TX), USA. Piscataway (NJ): IEEE; 2010. p. 678-681.
61. Arnold M, Ghosh A, Ameling S, Lacey G. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *EURASIP J Image Video Process* 2010;2010:814319.
62. Ganz M, Xiaoyun Y, Slabaugh G. Automatic segmentation of polyps in colonoscopic narrow-band imaging data. *IEEE Trans Biomed Eng* 2012;59:2144-2151.
63. Condessa F, Bioucas-Dias J. Segmentation and detection of colorectal polyps using local polynomial approximation. In: Campilho A, Kamel M, eds. *Image analysis and recognition*. Berlin: Springer, Berlin, Heidelberg; 2012. p. 188-197.
64. Tuba E, Tuba M, Jovanovic R. An algorithm for automated segmentation for bleeding detection in endoscopic images. In: 2017 International Joint Conference on Neural Networks (IJCNN); 2017 May 14-19; Anchorage (AK), USA. Piscataway (NJ): IEEE; 2017. p. 4579-4586.
65. Boonpogmanee I. Fully convolutional neural networks for semantic segmentation of polyp images taken during colonoscopy. *Am J Gastroenterol* 2018;113: S1532.
66. Brandao P, Mazomenos E, Ciuti G, et al. Fully convolutional neural networks for polyp segmentation in colonoscopy. In: *SPIE Medical Imaging*; 2017 Mar 3; Orlando (FL), USA. 2017.
67. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access* 2018;6:9375-9389.
68. Salehinejad H, Valae S, Dowdell T, Colak E, Barfett J. Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018 Apr 15-20; Calgary, Canada. Piscataway (NJ): IEEE; 2018. p. 990-994.
69. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018;321:321-331.
70. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data* 2019;6:60.
71. Shin Y, Qadir HA, Balasingham I. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access* 2018;6:56007-56017.
72. Kanayama T, Kurose Y, Tanaka K, et al. Gastric cancer detection from endoscopic images using synthesis by GAN. In: Shen D, Liu T, Peters TM, et al., eds. *Medical image computing and computer assisted intervention – MICCAI 2019*. Cham: Springer, Cham; 2019. p. 530-538.
73. Ahn J, Loc HN, Balan RK, Lee Y, Ko J. Finding small-bowel lesions: challenges in endoscopy-image-based learning systems. *Computer* 2018;51:68-76.
74. Mahmood F, Chen R, Durr NJ. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Trans Med Imaging* 2018;37:2572-2581.
75. Esteban-Lansaque A, Sanchez C, Borrás A, Gil D. Augmentation of virtual endoscopic images with intra-operative data using content-nets. *bioRxiv* 2019:681825.
76. Rau A, Edwards PJE, Ahmad OF, et al. Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. *Int J Comput Assist Radiol Surg* 2019;14:1167-1176.