Singapore Management University
# Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

# Finding small-bowel lesions: Challenges in endoscopy-image-based learning systems

Jungmo AHN
*Ajou University*

Loc Nguyen HUYNH
*Singapore Management University*, nlhuynh.2014@phdis.smu.edu.sg

Rajesh Krishna BALAN
*Singapore Management University*, rajesh@smu.edu.sg

Youngki LEE
*Singapore Management University*, YOUNGKILEE@smu.edu.sg

JeongGil KO
*Ajou University*

**DOI:** https://doi.org/10.1109/MC.2018.2381116

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Medical Sciences Commons, and the Software Engineering Commons

# Finding Small-Bowel Lesions: Challenges in Endoscopy-Image-Based Learning Systems

**Jungmo Ahn,** Ajou University

**Huynh Nguyen Loc,** Singapore Management University

**Rajesh Krishna Balan,** Singapore Management University

**Youngki Lee,** Singapore Management University

**JeongGil Ko,** Ajou University

*Capsule endoscopy identifies damaged areas in a patient's small intestine but often outputs poor-quality images or misses lesions, leading to either misdiagnosis or repetition of the lengthy procedure. The authors propose applying deep-learning models to automatically process the captured images and identify lesions in real time, enabling the capsule to take additional images of a specific location, adjust its focus level, or improve image quality. The authors also describe the technical challenges in realizing a viable automated capsule-endoscopy system.*

Capsule endoscopy is an effective means of diagnosing lesions (damaged areas) in the small intestine. Unlike the stomach or large intestine, which can be reached using wired endoscopes through the mouth or anus, investigating the small intestine using a wired endoscope is challenging due to the organ's position and length. Capsule endoscopy overcomes this problem by having the patient swallow a tiny pill-like device with a camera and wireless radio—for example, Given Imaging's PillCam SB series (www .givenimaging.com/en-int/Innovative -Solutions/Capsule-Endoscopy/Pages /default.aspx) or IntroMedic's Miro-Cam(www.intromedic.com/item/item _010100.asp)—that periodically takes images of the small bowel and sends them to an external storage device while moving along the intestine. Physicians manually analyze the collected images to identify lesions, which can be signs of various diseases such as obscure gastrointestinal (GI) bleeding, small-bowel tumors, and Crohn's disease.

However, state-of-the-art capsule endoscopy suffers from two critical shortcomings. First, it might not take clear pictures of lesions or could even completely miss them in many cases. Ideally, the capsule should take more images when moving through possible areas of damage and take fewer images in nonproblematic areas. However, due to the capsule's tiny size and limited battery, it defaults to taking pictures at regular intervals (for example, 2–4 fps) without any intelligent sampling. Second, doctors must manually go through a large number of images (approximately 50,000 images per patient) to make a proper diagnosis. This process is highly labor-intensive, requiring about 4

hours of doctor's time per patient, and thus highly costly.

To address these challenges, we propose an autonomous feedback-based capsule endoscopy system that uses *lesions-aware adaptive sampling* and *intelligent image-review interfaces*. The former lets the capsule take more high-quality pictures near potential lesion areas, which significantly increases the possibility of accurate diagnosis. The latter helps doctors quickly find images of lesions without having to spend multiple hours manually going through all of the captured images.

The key idea is to utilize deep-learning algorithms—for example, a convolutional neural network (CNN) and a generative adversarial network (GAN)—to accurately classify images as "normal" or "(potential) lesion" in real time and control the in-body capsule based on classification results. We first train our deep-learning models using images gathered from previous patients diagnosed with different small-bowel-related diseases. We pre-load the model on the external computing platform (typically worn at the patients' waist throughout the capsule-endoscopy procedure), which then uses the model to classify images in real time. When the image is recognized as a lesion (with high probability),

the system sends a control signal to the capsule device to temporarily increase its image quality and frame rate. If no further lesion images are detected, the capsule reduces its sampling rate to the normal 2–4 fps to conserve energy. Furthermore, the systems tags the classification results to the collected images, which it then organizes in a classified timeline to help doctors quickly find images of interest.

In the rest of this article, we detail the design of our autonomous feedback-based capsule endoscopy system and its challenges. In partic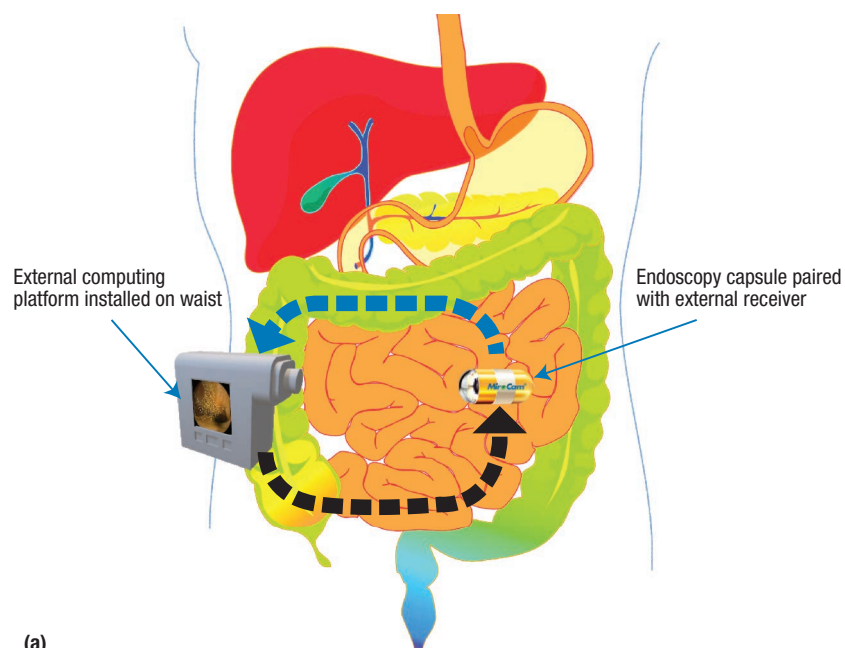ular, using an endoscopy dataset of 133,000 images collected from three patients, we present our preliminary deep-learning model design for lesion classification. We then outline remaining challenges and research directions in implementing the system.

## CAPSULE ENDOSCOPY

The capsule-endoscopy device includes a tiny camera, LED, transceiver, and battery to capture images of the gastrointestinal tract. As the blue arrow in Figure 1a shows, the capsule takes images and sends them to a waist-worn embedded device. In current capsule-endoscopy practices, images are simply stored until the capsule exits the body. Upon the completion of image collection, the doctor reviews

> THE KEY IDEA IS TO UTILIZE DEEP-LEARNING ALGORITHMS TO ACCURATELY CLASSIFY IMAGES AS "NORMAL" OR "(POTENTIAL) LESION" IN REAL TIME.

**(a)**



**(b)**

**FIGURE 1.** Autonomous feedback–based capsule endoscopy. (a) The blue arrow shows the transmission of endoscopy images from the capsule to the external computing platform, and the black arrows shows the platform sending feedback based on image classification results. (b) Flowchart of the feedback process; the "sb" flag indicates whether or not the capsule entered the small bowel.

the images to identify lesions on intestine walls.

While considered an effective way to capture small-bowel lesions, current capsule endoscopy has limitations.

First, unlike colonoscopy or gastroscopy, where doctors have full control to observe specific locations, the capsules take images usually at a fixed frame rate and do not have any adaptive control to focus on particular areas. Thus, in many cases, getting enough images of specific areas requires multiple rounds of lengthy and costly procedures. Second, doctors must spend hours manually reviewing all the taken images to find anomalies.

## Autonomous feedback-based capsule endoscopy

To overcome these challenges, we propose the inclusion of a deep-learning-based classification feedback loop, indicated by the black arrow in Figure 1a. Specifically, we plan to classify lesions in real time using the images that the capsule sends to the external embedded device and use this result to dynamically change the capsule's frame rate and/or image resolution. This process can provide doctors with enough image samples of lesions to make an accurate diagnosis with a single capsule-endoscopy procedure. Nevertheless, given that the capsule-endoscopy process takes as long as 8 to 10 hours to complete, the system must ensure that the frame-rate and image-resolution changes do not cause early shutdown of the capsule.

Figure 1b shows a flowchart of our proposed feedback process. Upon the arrival of images, the external computing platform operates a deep-learning model based on a CNN, customized to produce results with low latency. Our endoscopic image dataset is unique in that images have high similarities and small features represent the differences (for example, red bleeding dots or scars on the bowel surface). This makes the use of most existing (ImageNet-trained) CNN-based image-classification models difficult because they focus on classifying significantly different objects
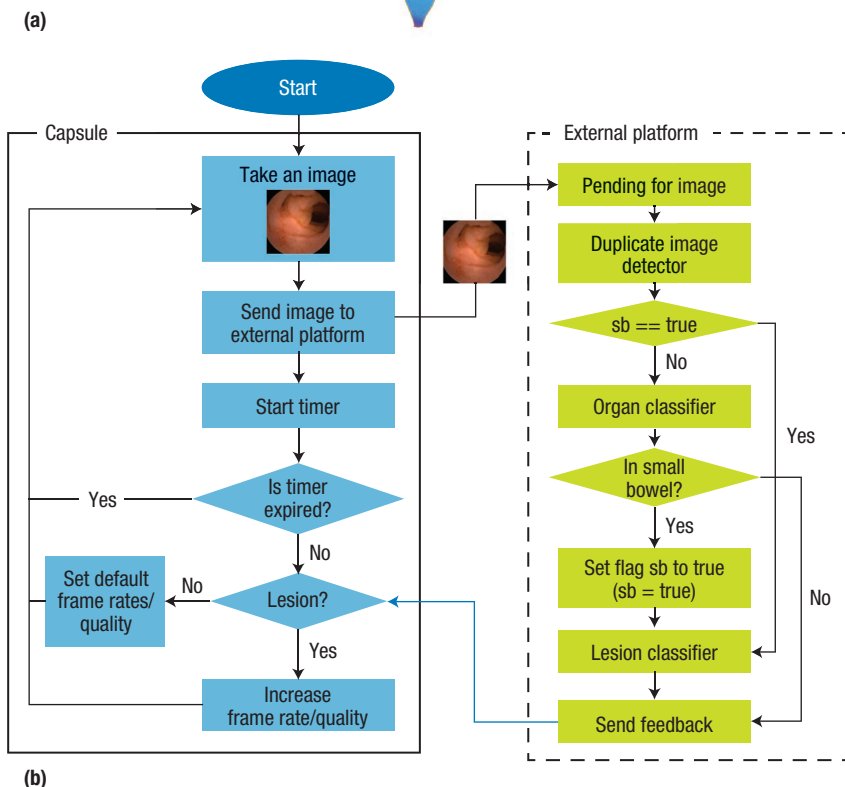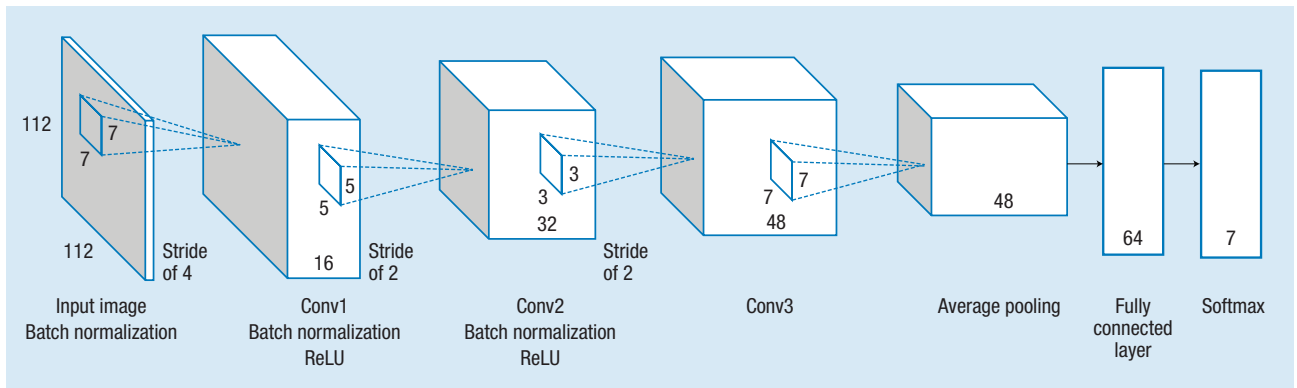
**FIGURE 2.** Customized three-layer convolutional neural network (CNN) model for lesion classification. We structure our model to achieve high accuracy with low computational latency.

such as cats and dogs. To overcome this limitation, we empirically design the model and select the hyperparameters by training the network on a training dataset from scratch until it converges, and we continuously modify the model based on a validation dataset.

Figure 2 shows our CNN, which consists of three convolutional layers. The first two layers are followed by batch normalization and rectified linear unit (ReLU) activation, respectively. ReLU is a frequently used nonlinear activation function, allowing the neural network to detect nonlinear features, and batch normalization normalizes the input data to avoid gradient vanishing.[1] The third convolutional layer is followed by global average pooling, and we add a final softmax layer for classification. Table 1 presents details of each network layer.

After running this model, the embedded computing platform outputs a classification result on whether or not the image possesses potential lesions. Although the hardware specifications can vary, the embedded computing platform is typically a resource-constrained device—for example, equipped with an ARM Cortex-A class processor (www.arm.com/products/processors/cortex-a). Nevertheless, for running deep-learning models, integrating embedded GPUs (such as NVIDIA Jetson-series processors) can be a plausible option. The following subsections present details on the components of our proposed system.

**TABLE 1.** Network architecture for endoscopy lesion detection with hyperparameter details.

| Type/stride | Filter shape | Input shape |
|---|---|---|
| Conv1/s4 | 7 × 7 × 3 × 16 | 112 × 112 × 3 |
| Batch normalization and rectified linear unit (ReLU) activation | | |
| Conv2/s2 | 5 × 5 × 16 × 32 | 28 × 28 × 16 |
| Batch normalization and ReLU activation | | |
| Conv3/s2 | 3 × 3 × 32 × 48 | 13 × 13 × 32 |
| Global average pool/s1 | Pool 7 × 7 | 7 × 7 × 48 |
| Fully connected | 64 | 1 × 48 |
| Softmax | Classifier | 1 × 64 |

### Duplicate image detector

Typically, capsule-endoscopy images have many duplicates. For example, the device might not (or will very slowly) move while navigating curves or encountering contents in the intestines, producing multiple images of the same scene. Quantitatively, previous work suggests that removing duplicate images from a capsule-endoscopy image set can reduce the set size by about 68 percent.[2]

In our design, it is also important to filter out these duplicate images at the external computing device for two major reasons. First, filtering duplicate images minimizes resource usage. By not classifying duplicate images, we can reduce the latency for lesion recognition sufficiently to provide real-time feedback to the capsule

device.[3] Second, given that doctors must review the entire image set to identify and confirm the (non)existence of lesions, minimizing duplicate images can help save time to provide high-quality care to many patients. In doing so, previous works estimated camera motions or designed similarity computation algorithms such as SIFT and SURF.[4,5] Our system also aims to identify images with high similarity by exploiting pixel-level correlations across a series of subsequent images.

### Designing low-latency lesion classifiers

The key design goal of our CNN-based classifier is low latency to provide real-time feedback to the capsule device (see Figure 2). Currently, the model parameters are tuned empirically with

the aim of detecting images for obscure GI bleeding, Crohn's disease, surveillance of polyps, and detection of small-bowel tumors. We note that endoscopic images for these different lesions show small but noticeable differences (for example, red or rash-shaped surfaces), which makes it promising to design a deep-learning-based lesion identifier/classifier.

Re-training and fine-tuning the deep-learning model with domain-specific datasets is a common practice to increase its accuracy (for example, VGG-16[6] and ResNet-151[7]). However, prior domain-specific models are not optimized for latency or memory usage, which are critical requirements in our scenario. A few recently proposed systems accelerate deep-learning-based classifiers on mobile devices by applying various optimization strategies such as layer decomposition and caching to improve latency. For example, DeepX[8] classifies an image using AlexNet at 500 ms latency, while Deep-Mon[9] further enhances latency to 260 ms by leveraging mobile GPUs. Nevertheless, prior systems are limited in supporting high-frame-rate image analysis in real time. Our optimized inference model contains 32,000 parameters and computes 4.7 million multiply–accumulate operations. As a reference, AlexNet contains 61 million parameters and computes 721 million multiply–accumulate operations, suggesting that our model is relatively lean. In terms of latency, our prototype model classifies a single image within about 1.14 ms on the Jetson Tegra K1 embedded GPU.[10]

## Increasing lesion-detection probabilities
A major drawback of capsule endoscopy is that doctors cannot maneuver the capsule to make detailed observations of target locations. Consequently, images can be blurred or only parts of the lesion might be present in an image. Under such circumstances, the only option is to redo the procedure in the hopes of getting a better view of the lesion or to diagnose based on limited information.

Our feedback-based system tries to overcome such limitations by forcing the capsule to perform two different actions. First, we can configure the capsule to take images at higher frame rates, which allows context-based image-quantity adjustment. Second, we can have the capsule take higher-quality images. Due to battery limitations, the capsule takes images at low resolution (320 × 320 pixels). While capturing higher-quality images during the entire process can stress the battery, defining short bursts of high-resolution images improves the chances of a better diagnosis on the first capsule-endoscopy trial. As a preliminary evaluation, we conducted experiments to distinguish normal images from erosion images (which take up 40 percent of the entire lesion dataset). We were able to detect erosion images with 86 percent accuracy and successfully classify normal images with 94.4 percent accuracy.

## Minimizing battery usage with organ classification
While not yet part of our system, we plan to add an organ classifier to identify the capsule's location. This is important in conserving battery resources given that it takes more than an hour for a capsule to pass through the stomach after swallowing,[11] which translates to approximately one eighth of its expected lifetime in the human body. Suppressing image taking (or at least reducing frame rate) before the capsule enters the small bowel would improve the lifetime. As Figure 1b shows, organ classification would be a preliminary screening phase. Once the capsule is determined to be in the small bowel, a flag is set so that the images received at the external platform can bypass this preliminary filter. Previous work shows that such classification is possible using a CNN variant,[12] and we plan to take a similar approach.

## TECHNICAL CHALLENGES
Despite our efforts to date, there are numerous technical (and nontechnical) hurdles we still must overcome before designing a fully autonomous feedback-based capsule endoscopy system.

### System-level challenges
Some challenges relate to building the overall system.

**Low latency for the feedback process.** The feedback process to autonomously control the capsule requires multiple message transmissions (including image transmissions) and computation (for example, image classification). Thus, the overall process can encounter layers of delay. Long delays in the feedback can lead to the capsule moving to a new location before receiving control commands, making low-latency feedback an important requirement. Typically, commercial capsule-endoscopy products take images at about 2 fps, which suggests that an image of a new location is expected to be taken every 500 ms.[11] We calculate that the capsule moves at about 0.56 mm/s, as the small bowel is some 8 meters in length and the capsule can exit the small bowel in about 4.5 hours. Therefore, it is important that the feedback procedures occur

prior to the capsule moving significantly far away from a suspected lesion. The MiroCam transmits its images at 6 Mbps, which takes about 140 ms for a standard 100-Kbyte image (for example, a 320 × 320 pixel image in raw Bayer pattern). This unavoidable transmission latency further challenges the system design. Note that these numbers represent the tightest timing limits for the capsule-endoscopy system and assume that the capsule does not take any reverse actions (such as go back up the bowel).

**Image pre-filtering.** To achieve real-time image classification, we must reduce the number of images to be processed by the deep-learning models by pre-filtering images that are certain to not have lesions (due to location or image quality) or duplicate images similar to previously processed samples. Figure 3a shows images that can be pre-filtered due to image quality. Specifically, these images captured variants of bubble-shaped intestinal fluids, which, even when passed through a properly trained model, will not be useful in detecting target lesions. Figure 3b shows duplicate images taken by the capsule device, which also require filtering.

**Low energy consumption.** Once swallowed, the capsule-endoscopy device should monitor the entire small bowel. Starting from the throat, the capsule will pass through the stomach and enter the small bowel; this time is called the gastric transit time (GTT). The GTT can differ among people by as little as 30 minutes to as long as multiple hours. The capsule device is active for the entire GTT and continuously takes images, lighting the scene with its embedded LED and transmitting/
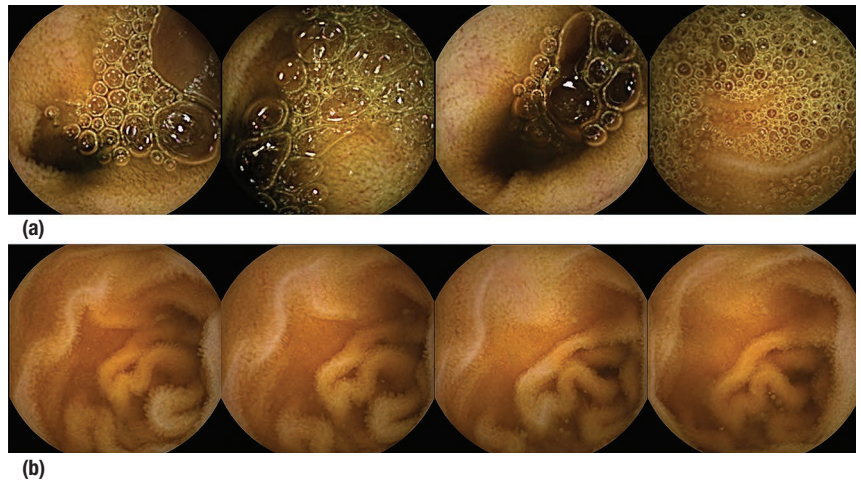


**FIGURE 3.** Endoscopy image pre-filtering. (a) Images with bubble-shaped intestinal fluids taken from the small bowel. (b) Duplicate images. Ideally, both types of images should be filtered prior to classification to conserve limited resources.

receiving data from the external embedded device. It is particularly vital that the capsule remains active during the entire small-bowel transit time (SBTT). The mean and standard deviation for the SBTT is 4.1 and 2.2 hours, respectively.[13] A recent study reports that the completion rate of capsules with 12-hour lifetimes are about 9 percent higher than capsules with 8-hour lifetimes.[14] This lengthy investigation time makes it vital that our proposed adaptive image-taking algorithms are as energy efficient as possible.

**Capsule localization.** Identifying the location of the capsule is very useful to ensuring that it remains active during the SBTT. To do this, we propose using a simple image-processing technique to determine the capsule's transition from the stomach to the duodenum, which is where the small bowel begins. In addition, while less related to detecting lesions, knowing the capsule's location can also help physicians perform better operations

when lesions are detected. To do this effectively will require new solutions (for example, use of inertial sensors), as using just images will not achieve good localization accuracy due to the similarity of images.

## Challenges for designing learning models

Other challenges pertain to designing the classification model.

**Dataset asymmetry.** By nature, deep-learning model training requires a large volume of labeled data. Also, ideally, the training dataset should not be overly biased toward a certain class. However, it is difficult to secure an equal balance of data samples for normal and lesion images. For example, a dataset of capsule-endoscopy images from six patients with small-bowel-related diseases holds more than 133,000 images, while the count of images with lesions is only about 100. This is inevitable given that the entire small bowel needs to be

monitored to capture a few problematic instances. Magnifying the problem, these 100 images include images for six different types of diseases (bleeding, erosion, ulcer, tumor, and so on), which further reduces the number of samples for each category. Such biased training data makes it difficult to design an accurate deep-learning model for capsule-endoscopy images.

**Gathering "labeled" ground-truth data.** It is well known that hospitals are stacked with massive amounts of patient data. However, the main problem that many researchers face

**Achieving both low latency and high classification accuracy.** To achieve real-time feedback-based capsule control, we must perform low-latency classification on the external platform. However, the model must also achieve low false-negative and false-positive rates, which impacts the system's clinical reliability and the capsule's energy efficiency, respectively. These are usually conflicting tradeoffs; thus, it is important that the system be well tuned to meet both requirements.

**Untrained cases.** While there are a limited number of diseases that occur

Unfortunately, collecting additional samples is costly, time-consuming, or sometimes impossible. Fortunately, new unsupervised learning tools and techniques such as autoencoders and GANs can help generate "fake" yet meaningful samples. In particular, these fake images are structurally similar to yet different enough from real images to build accurate predictive models with less overfitting issues. However, a key limitation is that fake data generation is not always contextually meaningful and the core features that represent a lesion could be absent in some cases.

We currently use the VAE-GAN model[16] with a network consisting of an encoder, a decoder, and a discriminator. The encoder and decoder act as the image generator, and the discriminator tries to determine which images are real. Figure 4a shows original images and Figure 4b shows visually similar images generated by the GAN using features extracted from the original images.

> ## THE MODEL MUST ACHIEVE LOW FALSE-NEGATIVE AND FALSE-POSITIVE RATES, WHICH IMPACTS THE SYSTEM'S CLINICAL RELIABILITY AND THE CAPSULE'S ENERGY EFFICIENCY.

is the lack of ground-truth labeling. For example, for capsule-endoscopy images, all of the images are labeled as "with lesion" if the particular patient has issues "somewhere" in the dataset. This does not necessarily mean that all of the images in the set include lesions. Tagging individual images is labor-intensive and costly, taking more than four hours for each capsule-endoscopy image set. Nevertheless, previous work by Google[15] shows that computer-based medical-image classification, with expert guidance, is feasible. Their system, which aims to detect diabetic retinopathy using approximately 130,000 images labeled by 54 ophthalmologists, shows 96 percent sensitivity using the Inception-v3 network.

in the small bowel, symptoms can be diverse. For example, a bleeding symptom detected at the intestine can have different spreading patterns. A tumor or ulcer can form in various shapes. Unlike clinically trained humans, deep-learning models cannot easily classify unique variants with high accuracy.

## RESEARCH DIRECTIONS
We now outline interesting future research directions.

### Artificial data generation
The heavy asymmetry between positive and negative samples within clinical datasets makes it challenging to build an accurate model.

We believe that with more lesion images for training, GANs potentially offer higher-quality "lesion-like" image samples. Note that doctors still must verify these "generated" images before they can be used. Overall, we believe research in generating realistic artificial clinical images can help in the development of many medical learning/classification systems.

### Accurate capsule localization
As mentioned above, a remaining key challenge is accurately locating the capsule inside the small bowel. However, adding additional sensors is difficult due to power, weight, and interference concerns. A better solution is to locate the capsule using the captured images. Unfortunately, this is difficult due to

the lack of properly labeled location data and the similarity of small-bowel images at different locations. Creating an accurate model would require technical and clinical staff help to gather large amounts of labeled image data and verify the network model.

## Zero false-negative classifier

It is vital that our solution, even with high base accuracy levels, does not misclassify any images with symptoms as normal images (generate false negatives). While false positives can be filtered out manually, even a single false-negative case can lead to critical errors in the clinical domain. Unfortunately, training a low or zero false-negative model is a challenging task, in particular due to the asymmetry of the dataset. However, another problem is the lack of a well-defined cost function to penalize the neural network when making false-negative inferences. Commonly used cost functions, such as categorical cross-entropy, penalize false-negative and false-positive results simultaneously. Therefore, research on identifying proper cost functions and generating usable images of various lesions can help in designing an effective neural network for autonomous feedback-based capsule-endoscopy systems.

I n this article, we described our initial efforts at designing an enhanced capsule-endoscopy system to address the two key challenges of current state-of-the-art technology: poor-quality or missed images of lesions, and a labor-intensive manual review process. In particular, we presented a low-latency, accurate, deep-learning-based autonomous feedback mechanism that recognizes possible images of lesions
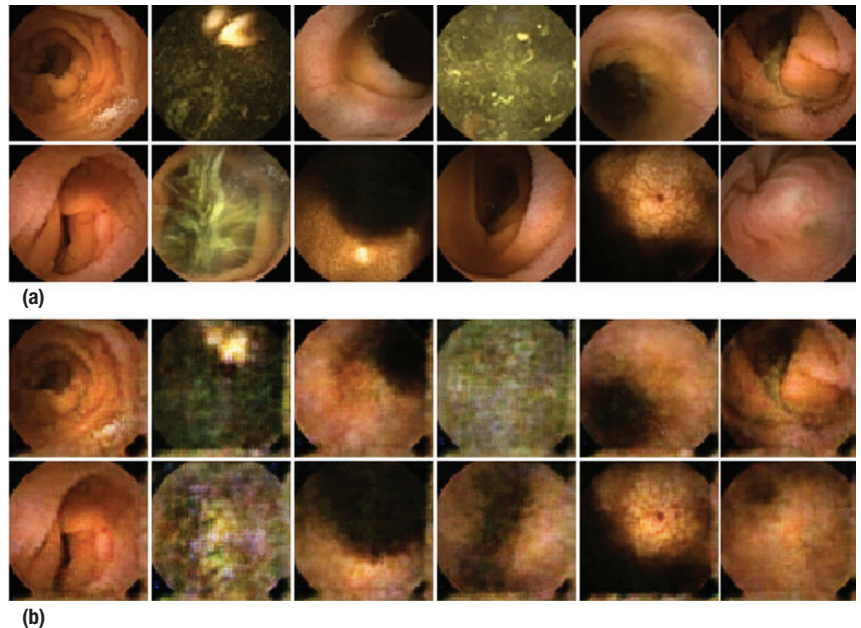


**FIGURE 4.** Artificially generated endoscopy images using a generative adversarial network (GAN): (a) original images and (b) artificially generated images. Artificial images could be used to overcome the asymmetry of positive and negative datasets in the training process.

in real time to dynamically increase the sampling rate or image resolutions. We also reviewed various system-level and deep-learning-related challenges, such as real-time feedback, energy constraints, and image quality, that must be addressed for autonomous feedback-based capsule-endoscopy systems to become viable. Finally, we are designing an intelligent image-review interface that uses our lesion classifier to help doctors focus on a much smaller set of possible lesion images. ⬛

## REFERENCES

1. Jiuxiang Gu et al., "Recent Advances in Convolutional Neural Networks," *Pattern Recognition*, vol. 77, May 2018, pp. 354–377

2. D.K. Iakovidis and A. Koulaouzidis, "Software for Enhanced Video Capsule Endoscopy: Challenges for Essential Progress," *Nature Reviews Gastroenterology & Hepatology*, vol. 12, no. 3, 2015, pp. 172–186.

3. B. Kang and H. Choo, "A Deep-Learning-Based Emergency Alert System," *ICT Express*, vol. 2, no. 2, 2016, pp. 67–70.

4. H.-G. Lee et al., "Reducing Redundancy in Wireless Capsule Endoscopy

## ABOUT THE AUTHORS

**JUNGMO AHN** is a PhD student in the Department of Software and Computer Engineering at the College of Information Technology at Ajou University and a member of the Ajou Embedded Intelligent Systems (AEIS) Laboratory. His research interests include embedded learning systems, mobile computing, and healthcare applications. Contact him at ajm100@ajou.ac.kr.

**HUYNH NGUYEN LOC** is a PhD student in the School of Information Systems at Singapore Management University. His current research focuses on mobile computing, parallel programming, and deep-learning applications. Contact him at nlhuynh.2014@phdis.smu.edu.sg.

**RAJESH KRISHNA BALAN** is a professor of information systems and director of the LiveLabs Urban Lifestyle Innovation Platform at Singapore Management University. His research interests include mobile systems, power management, and usability. Balan received a PhD in computer science from Carnegie Mellon University. Contact him at rajesh@smu.edu.sg.

**YOUNGKI LEE** is an assistant professor of information systems at Singapore Management University. His research focuses on mobile and sensor platforms to enable always available and highly enriched awareness of human behavior and contexts as well as building innovative life-immersive mobile applications in various domains such as daily well-being, childcare, and advertising in collaboration with domain experts. Lee received a PhD degree in computer science from KAIST. Contact him at youngkilee@smu.edu.sg.

**JEONGGIL KO** is an assistant professor in the Department of Software and Computer Engineering at the College of Information Technology as well as in the Department of Biomedical Informatics at the School of Medicine at Ajou University, where he also leads the AEIS Laboratory. His research interests are in the general area of developing web and cloud-based sensing systems with ambient intelligence for the Internet of Things and cyber-physical systems. Ko received a PhD in computer science from Johns Hopkins University. He is a Senior Member of IEEE. Ko is the corresponding author for this article. Contact him at jgko@ajou.ac.kr.

Videos," *Computers in Biology and Medicine*, vol. 43, no. 6, 2013, pp. 670–682.

5. H. Liu et al., "Wireless Capsule Endoscopy Video Reduction Based on Camera Motion Estimation," *J. Digital Imaging*, vol. 26, no. 2, 2013, pp. 287–301.

6. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," preprint, arXiv:1409.1556, 2014.

7. K. He et al., "Deep Residual Learning for Image Recognition," *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition* (CVPR 16), 2016, pp. 770–778.

8. N.D. Lane et al., "DeepX: A Software Accelerator for Low-Power Deep Learning Inference on Mobile Devices," *Proc. 15th Int'l Conf. Information Processing in Sensor Networks* (IPSN 16), 2016, article no. 23.

9. L.N. Huynh, Y. Lee, and R.J. Balan, "DeepMon: Mobile GPU-Based Deep Learning Framework for Continuous Vision Applications," *Proc. 15th Ann. Int'l Conf. Mobile Systems, Applications, and Services* (MobiSys 17), 2017, pp. 82–95.

10. NVIDIA Corp., *NVIDIA Tegra K1: A New Era in Mobile Computing*, v1.0, white paper, 2014; www.nvidia.com /content/PDF/tegra_white_papers /Tegra_K1_whitepaper_v1.0.pdf.

11. Z. Liao, C. Xu, and Z.-S. Li, "Completion Rate and Diagnostic Yield of Small-Bowel Capsule Endoscopy: 1 vs. 2 Frames per Second," *Endoscopy*, vol. 42, no. 5, 2010, pp. 360–364.

12. Y. Zou et al., "Classifying Digestive Organs in Wireless Capsule Endoscopy Images Based on Deep Convolutional Neural Network," *Proc. 2015 IEEE Int'l Conf. Digital Signal Processing* (DSP 15), 2015, pp. 1274–1278.

13. G. Ou et al., "Effect of Longer Battery Life on Small Bowel Capsule Endoscopy," *World J. Gastroenterology*, vol. 21, no. 9, 2015, pp. 2677–2682.

14. M. Rahman et al., "Comparison of the Diagnostic Yield and Outcomes between Standard 8 H Capsule Endoscopy and the New 12 H Capsule Endoscopy for Investigating Small Bowel Pathology," *World J. Gastroenterology*, vol. 21, no. 18, 2015, pp. 5542–5547.

15. V. Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, 2016, pp. 2402–2410.

16. A.B.L. Larsen et al., "Autoencoding beyond Pixels Using a Learned Similarity Metric," *Proc. 33rd Int'l Conf. Machine Learning* (ICML 16), 2016, pp. 1558–1566.