**CLINICAL ENDOSCOPY**

# Application of Artificial Intelligence in Capsule Endoscopy: Where Are We Now?

Youngbae Hwang[1*], Junseok Park[2*], Yun Jeong Lim[3] and Hoon Jai Chun[4]

[1]Intelligent Image Processing Research Center, Korea Electronics Technology Institute (KETI), Seongnam, [2]Digestive Disease Center, Institute for Digestive Research, Department of Internal Medicine, Soonchunhyang University College of Medicine, Seoul, [3]Department of Internal Medicine, Dongguk University Ilsan Hospital, Dongguk University College of Medicine, Goyang, [4]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea

Unlike wired endoscopy, capsule endoscopy requires additional time for a clinical specialist to review the operation and examine the lesions. To reduce the tedious review time and increase the accuracy of medical examinations, various approaches have been reported based on artificial intelligence for computer-aided diagnosis. Recently, deep learning–based approaches have been applied to many possible areas, showing greatly improved performance, especially for image-based recognition and classification. By reviewing recent deep learning–based approaches for clinical applications, we present the current status and future direction of artificial intelligence for capsule endoscopy. **Clin Endosc 2018;51:547-551**

**Key Words:** Capsule endoscopy; Deep learning; Artificial intelligence; Lesion detection

## INTRODUCTION

Capsule endoscopy (CE) has been developed to obtain endoscopic imaging of the entire small bowel.[1] Since its introduction, clinical practice guidelines have been established for several conditions, including unexplained obscure gastrointestinal (GI) bleeding, small bowel Crohn's disease, small bowel tumors, and other miscellaneous abnormalities.[2,3] Because endoscopy images are acquired by imaging sensors, most computer vision technologies[4] can be applied directly. Most of all, in the field of CE, various approaches based on artificial intelligence for computer-aided diagnosis have been undertaken

to reduce the long review time. Here, we summarize deep learning–based works for CE and present possible direction of artificial intelligence for CE.

## ADVENT OF DEEP LEARNING−BASED COMPUTER AIDED DIAGNOSIS

Since the early 2000s, the computer-aided decision support system (CADSS) has been researched extensively; as such, endoscopes can now take digital pictures.[5] CADSS has been designed to improve diagnostic accuracy by classifying abnormalities. In addition, a supportive system instead of direct decision-making support has emerged, including image quality enhancement, depth information extraction, and endoscopy localization. The ratio of the number of publications between conventional flexible endoscopy and CE regarding CADSS has been similar since 2007.[5]

Based on artificial intelligence, specifically computer vision and machine learning methodologies, various computational methods including algorithms for detecting hemorrhage and lesions, reducing review time, localizing capsules or lesions, and enhancing video quality have been proposed to improve

efficiency and diagnostic accuracy.[6] For detecting hemorrhages and lesions, color and texture information is usually used as a distinctive feature.[7-10] Image features extracted from endoscopic images can be classified into target class using machine learning algorithms such as a support vector machine (SVM),[7-9] neural network,[10] or binary classifier.[11] Although previous methods based on machine learning classifiers with invariant features showed concordant results for detecting various lesions, they have limitations related to insufficient training and testing databases and problems with specific feature design.

In early computer vision, image features such as corner[12] and edge[13] corresponding to scene structures were used to infer the class or 3D geometry of the target object. However, these primitive features are not invariant to various imaging condition changes (camera rotation and translation, illumination changes, occlusion, background clutter, and so on). To handle various changes, invariant features have been proposed that are robust to scale,[14,15] affine,[16] and local shape changes.[17] These handcrafted features have shown good performance for image-based recognition.[18,19] However, their results for very large-scale datasets are insufficient for practical application.[20] Since deep learning–based methods have shown much improved recognition performance,[21-24] most computer vision and machine learning problems have been approached using deep learning.

## LESION DETECTION AND CLASSIFICATION FOR CE

Deep learning–based lesion detection and classification methods for flexible endoscopy have recently been presented.[25] The ability of computer-assisted image analysis with a deep learning–based method, more specifically convolutional neural networks (CNN), has been tested to detect polyps, a surrogate for adenoma detection rate. With 8,641 labeling datasets from the colonoscopies of over 2,000 patients, the method showed an accuracy of 96.4%. For polyp detection, the binary classification task (whether an input image contains at least one polyp) has been performed using CNN architectures such as VGG (Visual Geometry Group from Oxford)[22] and ResNet.[24] For polyp localization, a task involving localizing the polyps in the image, a variation of Darknet has been used.[26]

Several methods based on deep learning have been proposed for CE in Table 1.[27-34] Zou et al. proposed a CNN-based method to solve the classification problem of digestive organs in CE.[27] The problem has three possible classes: stomach, small intestine, and colon. Compared to conventional scale invariant feature transform (SIFT)–[14] and SVM-based approaches (90.31%), the proposed method showed an accuracy of 95.52% for 15K images from 25 patients. Similarly, Seguí et al. proposed a classification method of motility events such as turbid, bubbles, clear blob, wrinkle, and wall.[28] They obtained an accuracy of 96.01% for 100K and 20K training and testing

**Table 1.** State-of-the-Art Deep Learning Based Methods for Capsule Endoscopy

| Study | Class | No. of training/testing images | No. of patients or videos | Features | Accuracy | Sensitivity/ Specificity |
|---|---|---|---|---|---|---|
| Zou et al. (2015)[27] | Localization[a] | 60K/15K | 25 patients | Alexnet | 95.5% | No info. |
| Seguí et al. (2016)[28] | Scene classification[b] | 100K/20K | 50 videos | CNN | 96.0% | No info. |
| Jia et al. (2016)[29] | Bleeding | 8.2K/1.8K | No info. | Alexnet | 99.9% | 99.2%/No info. |
| Li et al. (2017)[30] | Haemorrhage | 9,672/2,418 | No info. | LeNet AlexNet GoogLeNet VGG-Net | 100% | 98.7%/No info. |
| Yuan et al. (2017)[32] | Polyp | 4,000 (No info.) | 35 patients | SSAE | 98.0% | No info. |
| Iakovidis et al. (2018)[34] | Various lesions[c] | 465/233 | 1,063 volunteers | CNN | 96.3% | 90.7%/88.2% |
| | | 852/344 | No info. | | 77.5% | 36.2%/91.3% |
| He et al. (2018)[33] | Hookworm | 400K/40K | 11 patients | CNN | 88.5% | 84.6%/88.6% |
| Leenhardt et al. (2018)[31] | Angiectasia | 600/600 | 200 videos | CNN | No info. | 100%/96% |

CNN, convolutional neural networks; SSAE, stacked sparse autoencoder.
[a]Localization, Localization of stomach, small intestine, colon.
[b]Scene classification, Scene classification of Bubble, wrinkle, turbid, wall, clear.
[c]Various lesions, Gastritis, Cancer, bleeding, ulcer.

dataset images, respectively. The previous approach of combining handcraft features such as gist,[35] SIFT,[14] and color only achieved 82.8% accuracy.

For detecting bleeding or hemorrhaging, deep learning–based approaches have demonstrated 99.9% accuracy for 2,850 positive images[29] and 100% accuracy for 390 positive images.[30] The sensitivity is over 99%. Because the color cue of hemorrhages is obvious, the accuracy and sensitivity are much higher than those for classifying problems of digestive organs[27] or motility events.[28] Among various deep learning networks, there is a non-negligible discrepancy between the highest performance network (GoogLeNet, 100%) and the lowest performance network (LeNet, 97.44%).[30] To detect GI angiectasia, a CNN-based sematic segmentation algorithm was proposed.[31] From 200 capsule endoscopies, 20,000 normal frames and 2,946 frames with vascular lesions were extracted. To avoid overfitting, they used 600 images for training and another 600 images for testing by excluding successive frames of the same lesions. This work obtained a sensitivity of 100% and a specificity of 96%.

For detecting polyps, Yuan and Meng proposed a stacked sparse autoencoder–based approach.[32] Compared to other machine learning–based previous works, they achieved an accuracy of 98% for 4,000 images from 35 patients. Their method also classified normal images into turbid, bubble, and clear types. For detecting hookworms, He et al. proposed a novel edge extraction network to capture their characteristics.[33] From 440K images of 11 patients, the accuracy and sensitivity of the proposed network were 88.5% and 84.6%, respectively. When previous networks such as Alexnet[21] and GoogLeNet[23] were applied, the accuracy was higher (96.0% and 93.7%, respectively); however, the sensitivity was much lower (48.1% and 77.1%, respectively). Therefore, a novel network design for the specific problem is very important to obtain relevant results.

Iakovidis et al. presented a three-phase approach: a weakly supervised CNN for abnormality classification, deep saliency detection to detect salient points, and iterative cluster unification to localize GI anomalies.[34] They tested their proposed method for two datasets: a larger dataset (D1) with 10K images from more than 1,000 volunteers and a smaller dataset (D2) with 2,352 images. Similar to Leenhardt et al.,[31] they used 465 and 233 images for D1 training and testing, respectively, and 852 and 344 images for D2 training and testing, respectively. Compared with other deep learning– and machine learning–based previous works, their method showed improved results for D1 because they designed a complex multi-stage architecture for CE. However, their results for D2 were comparably worse than other methods, especially the low sensitivity (36.2%). We believe that the D2 is too small for the proposed

deep learning network to be learned effectively.

Although recent works based on deep learning have shown better performance than previous works based on handcrafted features,[27,28,33,34] there are contrary cases in which the method on handcrafted features was better for other problems.[36] For medical modality classification, Harris corner[12] with SIFT[14] and local binary pattern[37] descriptors showed better classification results than the CNN-based method because the number of images in the dataset for various modalities is insufficient to train a deep architecture.[36] The need for a sufficiently large database is one of the limitations of deep learning–based approaches. Overfitting is another crucial issue for deep learning–based approaches. This issue becomes more severe in cases of small databases. Although there are several approaches to mitigate overfitting in deep learning,[38,39] it remains an important problem for practical use. Overfitting should be considered for medical applications, which can be reduced by cost function regularity, data augmentation, relevant data selection, and other factors.

## CONCLUSIONS

Here we reviewed recent deep learning–based approaches for CE that have been applied to various problems such as scene classification and the detection of bleeding/hemorrhage/angiectasia, polyp/ulcer/cancer, and hookworms. Using large datasets from dozens of patients, they achieved much higher accuracy and sensitivity rates, sometimes close to 100%, compared to precious machine learning–based methods.

Because collecting databases for CE is difficult, more effective and generalized methods with the cooperation of many physicians and artificial intelligence engineers are required. Similar to other areas such as computer vision and robotics, the deep learning–based methodology will become more convincing and widely used. For medical applications, however, there are other bottlenecks such as dataset gathering, determination in terms of clinical aspects, and practical usage of computer-aided methods. Several research topics for CE remain, such as capsule localization, image enhancement, and reducing review time. One of the main drawbacks for CE is the lack of prospective trials to verify the accuracy of the computer-aided diagnosis. Since retrospective studies have reported more meaningful outcomes such as real-time image-based analysis during colonoscopy,[40,41] prospective research using CE is also very important for clinical applications.

## REFERENCES

1. Iddan G, Meron G, Glukhovsky A, Swain P. Wireless capsule endoscopy. Nature 2000;405:417.

2. Fisher LR, Hasler WL. New vision in video capsule endoscopy: current status and future directions. Nat Rev Gastroenterol Hepatol 2012;9:392-405.

3. Kwack WG, Lim YJ. Current status and research into overcoming limitations of capsule endoscopy. Clin Endosc 2016;49:8-15.

4. Szeliski R. Computer vision: algorithms and applications. London: Springer-Verlag; 2011.

5. Liedlgruber M, Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. IEEE Rev Biomed Eng 2011;4:73-88.

6. Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. Nat Rev Gastroenterol Hepatol 2015;12:172-186.

7. Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. Gastrointest Endosc 2014;80:877-883.

8. Lv G, Yan G, Wang Z. Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines. Conf Proc IEEE Eng Med Biol Soc 2011;2011:6643-6646.

9. Karargyris A, Bourbakis N. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. IEEE Trans Biomed Eng 2011;58:2777-2786.

10. Pan G, Yan G, Qiu X, Cui J. Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. J Med Syst 2011;35:1477-1484.

11. Mamonov AV, Figueiredo IN, Figueiredo PN, Tsai YH. Automated polyp detection in colon capsule endoscopy. IEEE Trans Med Imaging 2014;33:1488-1502.

12. Harris C, Stephens M. A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference 1988; 1988 Aug 31-Sep 2; Romsey, UK. Romsey: Roke Manor Research; 1988. p. 147-151.

13. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8:679-698.

14. Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision; 1999 Sep 20-27; Kerkyra, Greece. Piscataway (NJ): IEEE; 1999. p. 1150-1157.

15. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). Comput Vis Image Underst 2008;110:346-359.

16. Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. Int J Comput Vis 2004;60:63-86.

17. Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 2002;24:509-522.

18. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005 Jun 20-25; San Diego (CA), USA. Piscataway (NJ): IEEE; 2005. p. 886-893.

19. 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.

20. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211-252.

21. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems; 2012 Dec 3-6; Lake Tahoe (NV), USA. Red Hook (NY): Curran Associates, Inc.; 2012. p. 1097-1105.

22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv e-prints 2014. https://ui.adsabs.harvard.edu/#abs/2014arXiv1409.1556S.

23. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7-12; Boston (MA), USA. Piscataway (NJ): IEEE; 2015. p. 1-9.

24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NJ): IEEE; 2016. p. 770-778.

25. 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.

26. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NJ): IEEE; 2016. p. 779-788.

27. Zou Y, Li L, Wang Y, Yu J, Li Y, Deng WL. Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: 2015 IEEE International Conference on Digital Signal Processing (DSP); 2015 Jul 21-24; Singapore. Piscataway (NJ): IEEE; 2015. p. 1274-1278.

28. Seguí S, Drozdzal M, Pascual G, et al. Generic feature learning for wireless capsule endoscopy analysis. Comput Biol Med 2016;79:163-172.

29. Jia X, Meng MQH. A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016 Aug 16-20; Orlando (FL), USA. Piscataway (NJ): IEEE; 2016. p. 639-642.

30. Li P, Li Z, Gao F, Wan L, Yu J. Convolutional neural networks for intestinal hemorrhage detection in wireless capsule endoscopy images. In: 2017 IEEE International Conference on Multimedia and Expo (ICME); 2017 Jul 10-14; Hong Kong, China. Piscataway (NJ): IEEE; 2017. p. 1518-1523.

31. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest Endosc 2018 Jul 11 [Epub]. https://doi.org/10.1016/j.gie.2018.06.036.

32. Yuan Y, Meng MQ. Deep learning for polyp recognition in wireless capsule endoscopy images. Med Phys 2017;44:1379-1389.

33. He JY, Wu X, Jiang YG, Peng Q, Jain R. Hookworm detection in wireless capsule endoscopy images with deep learning. IEEE Trans Image Process 2018;27:2379-2392.

34. Iakovidis DK, Georgakopoulos SV, Vasilakakis M, Koulaouzidis A, Plagianakos VP. Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification. IEEE Trans Med Imaging 2018;37:2196-2210.

35. Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 2001;42:145-175.

36. Khan S, Yong SP. A comparison of deep learning and hand crafted features in medical image modality classification. In: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS); 2016 Aug 15-17; Kuala Lumpur, Malaysia. Piscataway (NJ): IEEE; 2016. p. 633-638.

37. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24:971-987.

38. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J

Mach Learn Res 2014;15:1929-1958.

39. Cogswell M, Ahmed F, Girshick R, Zitnick L, Batra D. Reducing overfitting in deep networks by decorrelating representations. ArXiv e-prints 2015. http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1511.06068.

40. Kominami Y, Yoshida S, Tanaka S, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016;83:643-649.

41. Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med 2018;169:357-366.