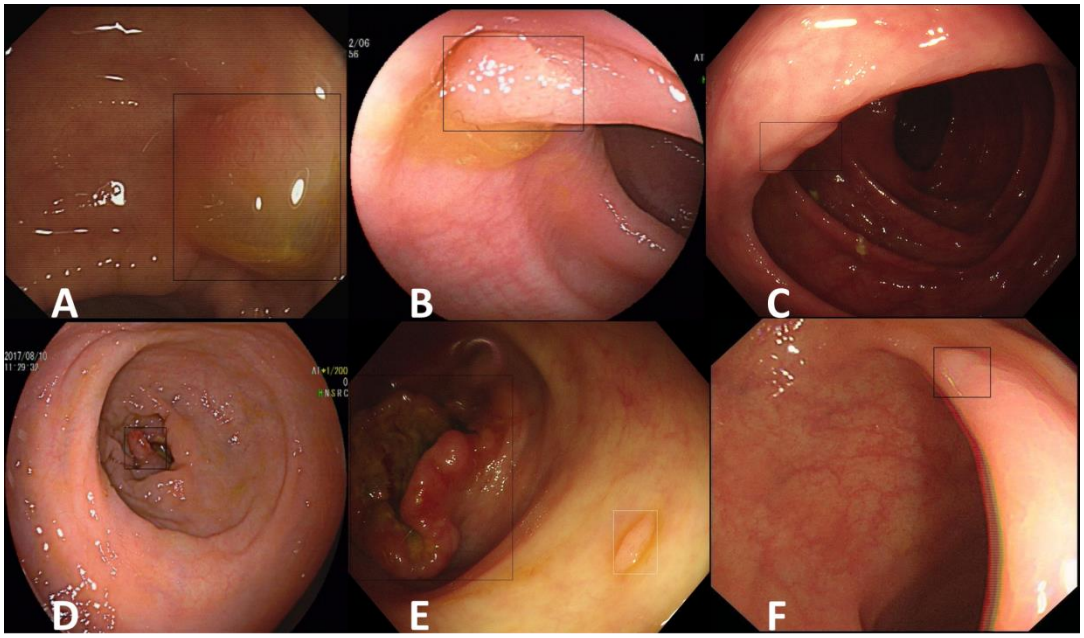
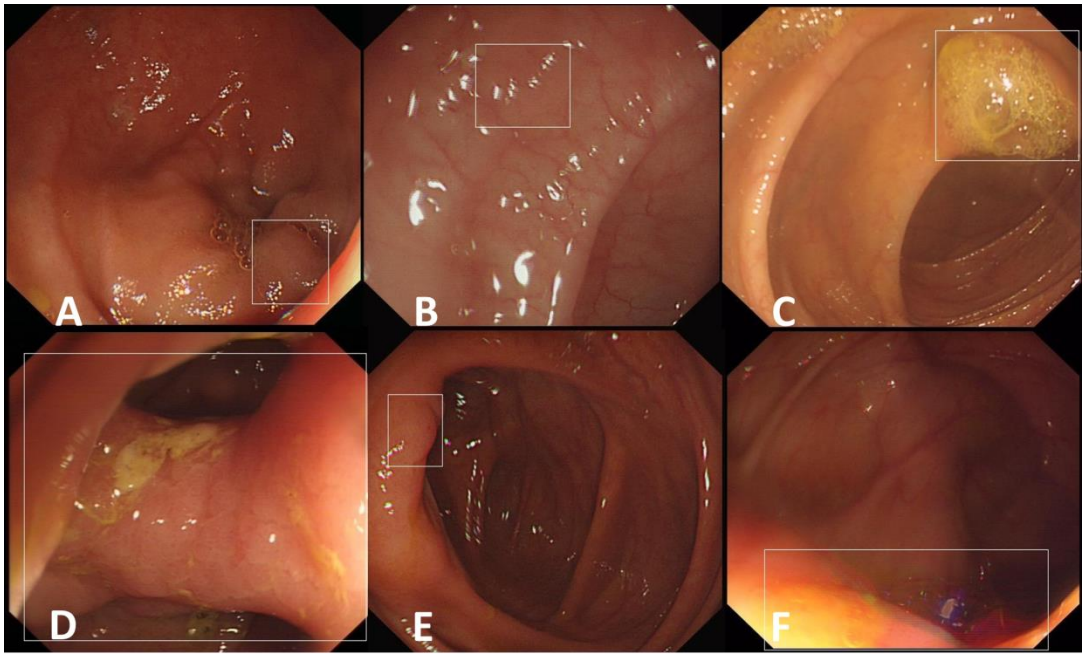



**Supplementary Figure 1 Examples of false-negative images and estimated causes.** A: Fecal fluid; B: Reflections of light; C: Difficult angle; D: Far distance; E: Shadow; F: Fuzzy image.




**Supplementary Figure 2 Examples of false-positive images and estimated causes.** A: Folds; B: Reflections of light; C: Bubble; D: Colonic valve; E: Ileocecal valve; F: Fuzzy image.

**Supplementary Table 1 Datasets from 20 digestive endoscopy centers in China**

| Digestive endoscopy centers | No. of Images | No. of lesions |
|---|---|---|
| Fuzhou General Hospital of Nanjing command | 716 | 56 |
| Jilin Guowen Hosipital | 13,562 | 962 |
| Dalian Municipal Central Hospital | 617 | 36 |
| First Affiliated Hospital of Kunming Medical University | 557,946 | 8,167 |
| Second Affiliated Hospital of Kunming Medical University | 831 | 37 |
| The Second Affiliated Hospital of Nanjing Medical University | 56,581 | 2,864 |
| Second Affiliated Hospital of Xuzhou Medical University | 15,856 | 494 |
| Xuzhou Central Hospital | 63,648 | 2,847 |
| Huai'an First People's Hospital | 30,316 | 1,140 |
| Huai'an Second People's Hospital | 27,955 | 2,697 |
| Magang Hospital | 49,796 | 1,008 |
| The Second Hospital of Hebei Medical University | 1,023,727 | 34,728 |
| Changhai Hospital | 1,104,118 | 50,598 |
| San Er Ling Yi Hospital | 752 | 59 |
| Xi'an Central Hospital | 21,565 | 6,182 |
| The Second Affiliated Hospital of Xi'an Jiaotong University | 1,156 | 902 |
| Weinan Central Hospital | 2,871 | 143 |
| Shaanxi Province People's Hospital | 29,951 | 1,570 |
| The First Affiliated Hospital of Xi'an Jiaotong University | 26,123 | 89 |
| 323 Hospital of People's Liberation Army | 2,479 | 69 |
| Total | 3,030,566 | 114,648 |

**Supplementary Table 2 Baseline characteristics of polyps in 47 colonoscopy videos**

| | | Polyps, no. |
|---|---|---|
| **Pathology** | adenomatous | 56 |
| | Nonneoplastic | 30 |
| **Location** | Proximal | 48 |
| | Distal | 38 |
| **Morphology** | ≤5 mm | 53 |
| | 6-9 mm | 27 |
| | ≥10 mm | 6 |
| **Size** | Flat/sessile | 71 |
| | Semi-pedunculated | 11 |
| | Pedunculated | 4 |
| **Detection difficulty** | Challenging | 13 |
| | Easy | 73 |
| **Overall** | | 86 |

**Supplementary Table 3 Characteristics of lesions in test dataset**

| | All | Non-adenomatous polyps | Adenomatous polyps | Adenocarcinomas |
|---|---|---|---|---|
| **Normal** | 9958 | - | - | - |
| **Lesions** | 1412 | 487 | 640 | 285 |
| **Size** | | | | |
| ≥ 10 mm | 303 | 3 | 15 | 285 |
| 6-9 mm | 396 | 112 | 284 | 0 |
| <5 mm | 713 | 372 | 341 | 0 |
| **Paris Classification** | | | | |
| Ip | 180 | 3 | 10 | 167 |
| Is | 581 | 189 | 357 | 35 |
| IIa | 525 | 275 | 250 | 0 |
| IIb | 30 | 19 | 11 | 0 |
| IIc | 2 | 0 | 0 | 2 |
| III | 32 | 0 | 0 | 32 |
| Unclassified | 62 | 1 | 12 | 49 |

**Supplementary Table 4 The diagnostic performance of CADe in 47 colonoscopy videos**

| | | PPV, % | Sensitivity, % | NPV, % | specificity, % |
|---|---|---|---|---|---|
| **Location** | Proximal | 59.5[58.9-60.0] | 84.9[84.4-85.3] | 98.0[97.9-98.0] | 92.7[92.5-92.8] |
| | Distal | 82.5[81.9-83.0] | 96.1[95.8-96.4] | 100.0[99.9-100.0] | 99.7[99.7-99.7] |
| **Morphology** | Flat | 67.6[67.2-68.0] | 88.1[87.8-88.4] | 98.7[98.6-98.7] | 95.4[95.4-95.5] |
| | Non-flat | 87.8[87.1-88.5] | 99.3[99.0-99.4] | 99.9[99.8-99.9] | 97.7[97.6-97.8] |
| **Size** | ≤ 5 mm | 61.0[60.5-61.5] | 83.8[83.4-84.2] | 98.2[98.1-98.2] | 94.2[94.1-94.3] |
| | ＞5 mm | 87.8[87.2-88.3] | 99.5[99.3-99.6] | 99.9[99.9-100.0] | 98.6[98.5-98.6] |
| **Pathology** | Adenomatous | 69.3[68.9-69.7] | 94.1[93.8-94.3] | 99.1[99.0-99.1] | 93.7[93.6-93.8] |
| | Nonneoplastic | 75.5[74.8-76.2] | 80.4[79.7-81.1] | 98.4[98.3-98.4] | 97.9[97.8-97.9] |
| **Challenging polyps** | | 84.5[83.7-85.2] | 66.2[65.3-67.1] | 94.5[94.3-94.7] | 97.9[97.8-98.1] |
| **Overall** | | 65.8[65.5-66.1] | 92.2[91.9-92.4] | 98.9[98.8-98.9] | 93.6[93.6-93.7] |

**Supplementary Table 5 Baseline characteristics of patients and colonoscopies**

| **Characteristics** | |
|---|---|
| **Age, mean (SD)** | 50.8 (12.9) |
| **Male, no. (%)** | 112 (53.6%) |
| **Indications for colonoscopy, no. (%)** | |
| Screening | 66 (31.6%) |
| Surveillance | 42 (20.1%) |
| Diagnosis | 101 (48.3%) |
| **Number of colonoscopists, no. (%)** | |
| ≥3000 | 4 (20.0%) |
| 1000-3000 | 10 (50.0%) |
| ＜1000 | 6 (30.0%) |
| **Number of colonoscopies, no. (%)** | |
| ≥3000 | 46 (22.0%) |
| 1000-3000 | 102 (48.8%) |
| < 1000 | 61 (29.2%) |
| **Colonoscope type, no. (%)** | |
| CF-Q290 | 171 (81.8%) |
| CF-Q260 | 26 (12.4%) |
| CF-Q240 | 12 (5.7%) |
| **BBPS, mean (SD)** | 6.1 (1.4) |
| **BBPS≥6, no. (%)** | 157 (75.1%) |
| **Withdrawal time, mean (SD), min** | 5.8 (2.4) |
| **Withdrawal time≥6 min, no. (%)** | 102 (48.8%) |

SD, standard deviation; BBPS, Boston bowel preparation score

**Supplementary Table 6 Number of polyps detected by colonoscopists or CADe**

| | Col Only | Col and CADe | CADe Only |
|---|---|---|---|
| **Polyps/Adenomas** | 3/0 | 168/67 | 17/5 |
| **Age** | | | |
| <50 | 1/0 | 42/14 | 6/1 |
| ≥50 | 2/0 | 126/53 | 11/4 |
| **Sex** | | | |
| Male | 3/0 | 104/46 | 11/3 |
| Female | 0/0 | 64/21 | 6/2 |
| **Location** | | | |
| Proximal | 0/0 | 74/27 | 8/2 |
| Distal | 3/0 | 94/40 | 9/3 |
| **Size** | | | |
| ≥10 mm | 0/0 | 13/13 | 0/0 |
| 6-9 mm | 1/0 | 37/23 | 1/0 |
| ≤5 mm | 2/0 | 118/31 | 16/5 |
| **Morphology** | | | |
| Flat | 3/0 | 137/44 | 15/4 |
| Subpedunculated | 0/0 | 24/15 | 2/1 |
| Pedunculated | 0/0 | 7/6 | 0/0 |
| **Indications** | | | |
| Screening | 2/0 | 47/22 | 6/3 |
| Surveillance | 0/0 | 50/19 | 6/1 |
| Diagnosis | 1/0 | 71/26 | 5/1 |
| **Colonoscopes** | | | |
| CF-Q290 | 3/0 | 131/58 | 14/5 |
| CF-Q260 | 0/0 | 27/6 | 3/0 |
| CF-Q240 | 0/0 | 10/3 | 0/0 |
| **Experience** | | | |
| >3000 | 1/0 | 57/23 | 0/0 |

| | | | |
|---|---|---|---|
| 1000-3000 | 1/0 | 79/28 | 12/3 |
| <1000 | 1/0 | 32/16 | 5/2 |
| **BBPS** | | | |
| < 6 | 1/0 | 19/7 | 1/0 |
| ≥ 6 | 2/0 | 149/60 | 16/5 |
| **Withdrawal time** | | | |
| < 6 min | 1/0 | 56/24 | 2/1 |
| ≥ 6 min | 2/0 | 112/43 | 15/4 |

Col, colonoscopists; CADe, artificial intelligence assisted-diagnosis system; BBPS, Boston bowel preparation score

**Supplementary Video legends**

**Supplementary Video 1 The process of labeling images.**

**Supplementary Video 2 The process of detecting and classifying polyps.**

**Supplementary Materials**

**Appendix A**

**Screening and labeling of colonoscopy images**

**Figure 1-A** illustrates the workflow of preprocessing images. More than 3 million images from 67836 patients were retrospectively collected from 20 centers using the following brands of colonoscopes: OLYMPUS (CF-Q290, CF-190, CF-180, CF-Q260, CF-160, and CF-Q240), FUJINON (EC-530, EC-450, and EC-250), and PENTAX (EC38-210 and EC34-210) (**Supplementary Table 1**). The images were stored at resolutions ranging from 472×395-1276×1020 pixels with a median resolution of 764×572 pixels. Primary screening of images and image labeling were conducted to ensure image quality and to further confirm the pathology of the polyps. Only images of patients with one polyp and corresponding pathological report were included to ensure the accuracy of pathological results. A JAVA Script was used to screen for records with only one colonic polyp to ensure one-to-one correspondence between each polyp and histological report in the primary screening of images. Subsequently, the included images were hand-checked again by doctors to exclude images of multiple polyps, images with imaging enhancing techniques, and images of non-colonic polyps, inflammatory bowel disease, non-epithelial polyps, and polyps without a pathological diagnosis, and images with obvious halation, defocus, and fuzzy dots. Images of the ileocecal valve, colonic valve, anus, and fecal fluid or bubbles without colonic polyps were additionally included as negative materials for training and testing.

Regarding the image labeling, 55 certified colonoscopists were asked to independently confirm the presence of polyps. During the process, colonoscopists could outline the border of the polyps, label the pathology, location, size, and morphology of the polyps, and make revisions at any time in a specially designed labeling system (**Supplementary video 1**).

All the labeling procedures of colonoscopists were independently conducted with blindness and repetition. The classification of each image was annotated when the consensus was reached in randomly selected 3 colonoscopists.

When the consensus could not be reached, 2 colonoscopists with 10+ years' experience made the final determination. The location of each polyp was defined as the minimum rectangular area that included the regions delineated by the 3 colonoscopists. All labeled images were randomly divided into a training or testing dataset while images from the same polyp were ensured not to be divided. The numbers of images finally enrolled in the two groups were 59708 (8018 polyps) and 11370 (825 polyps), respectively (**Figure 1-A**).

**Appendix B**

**Screening and labeling of colonoscopy videos**

In order to comprehensively evaluate CADe's real-time diagnostic performance, 47 unaltered full-range colonoscopy videos of routine practice including 86 histologically confirmed polyps were consecutively collected from routine practice of Changhai endoscopy center.

Notably, both easily-detecting and challenging polyps were included in the video datasets. All included videos were complete continuous withdrawing process after reaching the cecum, with the mean withdrawing time (excluding the time for biopsy and polypectomy) more than 6 minutes. The frames of biopsy, polypectomy and post-polypectomy wound were strictly excluded from analysis. Every frames of polyps appearing from the cecum to anus were included. These colonoscopies were performed by three colonoscopists (no. of diagnostic colonoscopy > 1000) of Changhai endoscopy center. To validate the diagnostic performance of CADe in real-time colonoscopy videos, two colonoscopists with ADR > 30% were invited to labeled the images (with or without polyp) frame by frame, whose results served as the golden standard, with inconsistence resolved by discussing with a senior colonoscopists.

**Appendix C**

The CADe was built based on the You Only Look Once (Yolo) v2 deep learning framework. The system was a cascade of 2 modules, which were developed and evaluated separately. Firstly, a classification module was developed to distinguish white light colonoscopy from other enhanced optic modes, such as narrow band imaging (NBI). Secondly, a detection module

was developed to predict the boundaries that tightly enclose an identified polyp. While previous studies that evaluate polyp detection or polyp localization have operated under the assumption that only one polyp is present per image, our detection module was designed to accurately localize an arbitrary number of polyps per image.

Both modules were developed based on CNNs, details of which will be introduced in the following subsections. To mitigate the gap of our relatively small data size compared to natural image datasets, transfer learning was applied for all the modules with ImageNet pretrained convolution layers [1]. Both training and testing were done within a TensorFlow framework [2], except for the detection module which was trained within a Darknet framework [3]. All the training used a stochastic gradient descent with a Nesterov momentum optimizer (momentum of 0.9). For detection/classification, we used a batch size of 512/60, weight decay of 0.0005/0.004, a starting learning rate of 0.008/0.001, and dropout rate of 0.2 (only for classification). The detection module was built based on You Only Look Once (Yolo) v2[3, 4]. Compared to other start-of-the-art detection algorithms, such as region-convolution neural network (RCNN) and its variants, Yolo unifies the whole detection pipeline into a single neural network to extract features and predict bounding boxes and class probabilities concurrently from the entire image. Therefore, Yolo runs extremely fast, which is suitable for our aim of real-time detection. Additionally, Yolo reasons on a larger context and thus is less likely to mistake backgrounds for objects [3, 4].

To construct the detection dataset, the location of each polyp was first annotated by colonoscopists with a polygon, which was then converted into a bounding box that tightly enclosed the polygon. Image was resized such as its shorter side equals to 576 pixels during testing, or randomly drawn from 320 to 608 pixels during training. Data augmentation was used, including random crops and color shifting, to enhance robustness against variation during testing.

For each training, a model with the lowest validation error was selected to evaluate on testing set. For evaluating the detection module, intersection over union (IoU) was measured between ground truth bounding box and predicted bounding box, which was defined as

IoU= (Area of Overlap)/(Area of Union)

We defined a true positive if IoU > 0.3, which was clinically helpful for colonoscopists to identify potentially missed polyps. Images with false-negatives and false-positives were read and analyzed by 2 experienced colonoscopists (Yu Bai, Xia Yang) to classify potential causes for these errors. All experiments were conducted on the graphics processing units of NVIDIA Tesla M40 with a 24GB RAM. Data augmentation, weight decay, early stopping and dropout were all utilized during model training to mitigate the risk of overfitting.

**Appendix D**

The study adopted several standards to define the identification and classification of colonic polyps, which was modified from previous reports[5, 6]: 1) when the system identified and confirmed any polyp on an image without a polyp or cancer, the result was judged to be false positives (FP); 2) when the system confirmed and correctly localized a polyp on an image, the result was judged to be true positives (TP); 3) when the system did not confirm or correctly localize a polyp, the result was judged to be a false negative (FN); 4) when the system confirmed no polyps on an image without polyps or cancer, the result was judged to be true negatives (TN). The diagnostic performance in the test were calculated as follows: accuracy = (TP+TN)/(TP+TN+FP+FN); sensitivity = TP/(TP+FN); specificity = TN/(TN+FP); PPV = TP/(TP+FP); NPV = TN/(TN+FN). For the video test, adenomas located proximal to the splenic flexure were defined as proximal adenomas, whereas those located more distally were defined as distal adenomas; the videos were carefully read by an experienced colonoscopists (Bai Yu) with the visibility of polyps determined.

**References for the supplementary materials**

1. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the Inception Architecture for Computer Vision. Computer Vision and Pattern Recognition, pp 2818-2826
2. Rampasek L, Goldenberg A (2016) TensorFlow: Biology's Gateway to Deep Learning? Cell systems 2:12-14
3. Redmon J, Divvala S, Girshick R, Farhadi A (2015) You Only Look Once: Unified, Real-Time Object Detection.
4. Redmon J, Farhadi A (2017) YOLO9000: Better, Faster, Stronger
5. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T (2018) Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric cancer: official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association 21:653-660
6. Horie Y, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, Hirasawa T, Tuchida T, Ozawa T, Ishihara S, Kumagai Y, Fujishiro M, Maetani I, Fujisaki J, Tada T (2018) The diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointestinal endoscopy