Category	Full Name	Related Disease(s)	Imaging Modality	Data Scale	Annotation / Attributes	Size	Institution	Data Availability	URL
Polyp Segmentation & Detection	Kvasir-SEG	Colorectal Polyps	Colonoscopy (WLI)	1,000 polyp images (332x487 to 1920x1072)	Pixel-level masks, bounding boxes (JSON)	44.1 MB	Simula Research Laboratory, Vestre Viken Health Trust, Norway	Publicly available (research & education)	https://datasets.simula.no/kvasir-seg/
	Kvasir-sessile	Flat, sessile & diminutive polyps (<10 mm)	Colonoscopy (WLI)	196 polyp images (subset of Kvasir-SEG)	Pixel-level masks, bounding boxes	13.7 MB	Simula Research Laboratory, NTNU, Vestre Viken Health	Publicly available (research & education)	https://datasets.simula.no/kvasir-seg/
	CVC-ClinicDB	Colorectal Polyps	Colonoscopy (WLI)	612 frames from 31 sequences (23 patients; 384x288)	Pixel-level binary polyp masks	50.23 MB (327.2 MB PraNet Test)	Computer Vision Center (CVC, UAB) & Hospital Clinic Barcelona, Spain	Publicly available	https://github.com/DengPingFan/PraNet (Google Drive: https://drive.google.com/file/d/1Y2z7FD5p5y31vkZwQQomXFRB0HutHyao/view)
	CVC-ColonDB	Colorectal Polyps	Colonoscopy (WLI)	300 frames from 13 sequences (13 patients; 574x500)	Pixel-level binary polyp masks, lumen	327.2 MB (PraNet Test bundle)	Computer Vision Center (CVC), Universitat Autonoma de Barcelona	Publicly available	https://github.com/DengPingFan/PraNet (Google Drive: https://drive.google.com/file/d/1Y2z7FD5p5y31vkZwQQomXFRB0HutHyao/view)
	ETIS-LaribPolypDB	Colorectal Polyps (small/subtle polyps)	Colonoscopy (HD WLI)	196 frames from 34 sequences (44 polyps; 1225x966)	Pixel-level binary polyp masks	327.2 MB (PraNet Test bundle)	ETIS Lab & Lariboisiere Hospital, Paris, France	Publicly available (legacy site may require mirror)	https://polyp.grand-challenge.org/ETISLarib/
	BKAI-IGH NeoPolyp-Small	Neoplastic & non-neoplastic colorectal polyps	Colonoscopy (WLI, LCI, BLI, FICE)	1,200 polyp images (1,000 train, 200 test)	Pixel-level neoplastic/non-neoplastic masks, bounding boxes	313.72 MB	BKAI (HUST) & Institute of Gastroenterology and Hepatology, Vietnam	Publicly available (Kaggle)	https://www.kaggle.com/c/bkai-igh-neopolyp/
	PolypGen (2023)	Colorectal Polyps	Colonoscopy (WLI)	8,037 frames (3,762 positive, 4,275 negative) across >300 patients	Pixel-level masks, bounding boxes, center/sequence tags	1.57 GB	Multi-center (6 hospitals: UK, Norway, Egypt, Italy, Sweden, France)	Publicly available (Synapse account)	https://www.synapse.org/Synapse:syn26376615/wiki/613312
	PolypDB (2024)	Colorectal Polyps	Colonoscopy (BLI, FICE, LCI, NBI, WLI)	3,934 polyp images across 5 imaging modalities & 3 centers	Pixel-level masks, bounding boxes, modality & center labels	~1.5 GB	Norway, Sweden & Vietnam (3 medical centers)	Publicly available (Kaggle)	https://www.kaggle.com/datasets/debeshjha1/polypdb
	FAR-POLYP-SEG (2026)	Colorectal Polyps (with normal mucosa)	Colonoscopy (WLI)	455 patients (incl. 7,749 normal-mucosa frames)	Pixel-level masks, patient age (17-85) & sex, 5-fold split	~1.2 GB	Farhikhtegan Hospital, Islamic Azad Univ., Tehran, Iran	Publicly available (CC BY 4.0, Zenodo)	https://zenodo.org/records/20284781
Video Polyp Detection & Segmentation	SUN-SEG (2022)	Colorectal Polyps (6 pathology types)	Colonoscopy Video (WLI)	158,690 frames (49,136 polyp, 109,554 normal) from 1,106 clips	Masks, boundaries, scribbles, polygons, 10 visual attributes	~65.5 GB (21.2 GB pos +44.3	Showa University, Nagoya University & IIAI	Publicly available (research/non-commercial)	https://github.com/GewelsJI/VPS
	REAL-Colon (2024)	Colorectal Polyps (real-world screening)	Colonoscopy Video (Full-Res WLI)	2.7M frames from 60 full-resolution colonoscopy videos	350k bounding boxes, age, sex, endoscope brand, BBPS, histology	880.78 GB in total , but support partial download	Multi-center (Japan, Austria, Italy, USA)	Publicly available (Figshare)	https://plus.figshare.com/articles/media/REAL-colon_dataset/22202866

Multi-Class GI, Pediatric & Polyp Sizing	HyperKvasir	GI tract findings (23 classes, polyps, colitis, etc.)	Upper & Lower GI Endoscopy (WLI)	110,079 images (10,662 labeled; 1,000 polyp masks) & 374 videos	Class labels, pixel-level polyp masks, bounding boxes	58.6 GB (46 MB seg)	Simula Research Laboratory, Baerum Hospital, Norway	Publicly available	https://datasets.simula.no/hyper-kvasir/
	CP-CHILD (2020)	Pediatric Colorectal Polyps	Colonoscopy (WLI)	9,500 images (1,400 polyp, 8,100 normal; 1,600 pediatric pts)	File-level binary polyp/non-polyp labels	410 MB	China (Wang et al., 2020)	Publicly available (Kaggle)	https://www.kaggle.com/datasets/mahdieh002/colonoscopy-cp-child?resource=download
	GastroVision (2023)	27 GI classes (landmarks, polyps, cancer, etc.)	Upper & Lower GI Endoscopy	8,000 images across 27 anatomical & pathological classes	Multi-class expert diagnostic labels	~1.8 GB	Baerum Hospital (Norway) & Karolinska Univ. Hospital (Sweden)	Publicly available	https://github.com/DebeshJha/GastroVision
	Polyp-Size (2025)	Colorectal Polyps (size estimation)	Colonoscopy Video (WLI)	42 videos (1000x1170, 30 fps, 410s total; 3,858 sampled frames)	0.01 mm post-resection caliper size, Paris class, histology, age/sex	345.74 MB	Renji Hospital & Shanghai Jiao Tong University, China	Publicly available (Figshare)	https://doi.org/10.6084/m9.figshare.28030115
Multimodal VQA & Instruction Tuning	Kvasir-VQA (2024)	GI findings (Normal, Polyps, Esophagitis, UC, Tools)	GI Endoscopy (WLI)	6,500 images with 58,849 question-answer pairs	VQA pairs, captions, weak localization & counting QA	~1.5 GB (images)	SimulaMet & Oslo Metropolitan University, Norway	Publicly available	https://datasets.simula.no/kvasir-vqa/
	ColonINST (2024)	62 colonoscopy categories	Colonoscopy (Multimodal)	300k+ images, 128k+ captions, 450k+ instruction dialogues	Classification, detection, segmentation & VQA instructions	311 MB (JSON)	AI4Colonoscopy / IntelliScope Team	Publicly available (Hugging Face)	https://huggingface.co/datasets/ai4colonoscopy/ColonINST-v1
	SAGE (2026)	18 GI classes (landmarks, polyps, abnormalities)	Upper & Lower GI Endoscopy	1,300 images with 14,276 VQA pairs (South Asian cohort)	Captions, hallucination span tags, 18 multi-label tags, VQA	~450 MB	NAAMII & Dhulikhel Hospital, Nepal (Open Data MICCAI 2026)	Publicly available (CC BY-SA 4.0, Synapse)	https://www.synapse.org/Synapse:syn75397327/files/

More related information see:https://github.com/cloneiq/Awesome-GI-Endoscopy-Datasets/tree/main/datasets