Glossary :

*WL : White light

*NBI : Narrow Band Imaging, imaging technique that enhances visualization by using specific narrow bands of light to highlight certain features of tissues and blood vessels

*resolution = total number of pixels in an image and = width(w)*height(h)

*RCNN : Region-based Convolutional Neural Network

Database +reference	format of images/videos	Number & color	Ground truth	Resolution (w x h)	Data base Architecture	Source of data	link
CVC-ClinicDB Bernal et al. 2015	tif	612 images + noir et blanc	Polyp locations (binary mask)	384 × 288	1) Original images folder: original/frame_number.tiff 2) ground truth folder : Polyp mask: ground truth/frame_number.tiff	612 sequential WL images with polyps extracted from 31 sequences (23 patients) with 31 different polyps.	https://polyp.grand-challenge.org/CVCClinicDB/
CVC-ColonDB Bernal et al. 2012 Vázquez et al. 2017	png	380 images + couleur	Polyp locations (binary mask)	574 × 500	2 folders /images and /masks	300 sequential WL images with polyps extracted from 13 sequences (13 patients).	https://figshare.com/articles/figure/Polyp_DataSet_zip/21221579 a part of the daatset.zip
CVC-EndoScene Still dataset Vázquez et al. 2017	png	couleur	Locations for polyp, background, lumen and specular lights (binary mask)	574 × 500, 384 × 288	Train, Test, val folders	912 WL images with polyps extracted from 44 videos (CVC-ClinicDB + CVC-ColonDB).	https://drive.google.com/file/d/1MuO2SbGgOL_jdBu3ffSf92feBtj8pbnw/view
CVC-PolypHD Bernal et al. 2012 Vázquez et al. 2017 Bernal et al. 2021			Polyp locations (binary mask)	1920 × 1080		56 WL images.	https://giana.grand-challenge.org/ *can't find

ETIS-Larib Silva et al. 2014	tif	couleur	Polyp locations (binary mask)	1225 × 966	2 folders : /ETIS-LaribPolypDB contains images and /Ground Truth contains masks	196 WL images with polyps extracted from 34 sequences with 44 different polyps	https://polyp.grand-challenge.org/ETISLarib/
Kvasir-SEG / HyperKvasir Pogorelov et al. 2017 Jha et al. 2020 Borgli et al. 2020	jpg/ jpeg	couleur/	Polyp locations (binary mask and bounding box)	Various resolutions	2 folders : /images and /masks —--- 4 folders /Labeled image data, /unlabeled image data, /segmented image data, and /annotated video data.	1 000 polyp images	https://datasets.simula.no/kvasir-seg https://datasets.simula.no/hyper-kvasir/
ASU-Mayo Clinic Colonoscopy Video Tajbakhsh et al. 2016	video		Polyp locations (binary mask)	688 × 550		38 small SD and HD video sequences: 20 training videos annotated with ground truth and 18 testing videos without ground truth annotations. WL and NBI.	https://polyp.grand-challenge.org/AsuMayo/ *not free access, need to contact Prof. Jianming Liang at Arizona State University
CVC-ClinicVideo DB Angermann et al. 2017 Bernal et al. 2018 Bernal et al. 2021	video		Polyp locations (binary mask)	768 × 576		38 short and long sequences: 18 SD videos for training.	https://giana.grand-challenge.org *can't find
Colonoscopic Dataset Mesejo et al. 2016	video	RGB	Polyp classification (Hyperplastic vs. adenoma vs. serrated)	768 × 576	data is not organized	76 short videos (both NBI and WL).	http://www.depeca.uah.es/colonoscopy_dataset/ not able to download the dataset at once, but the data is available

PICCOLO Sánchez-Peralta et al. 2020	image tif		Polyp locations (binary mask) Polyp classification, including: Paris and NICE classifications, Adenocarcinoma vs. Adenoma vs. Hyperplastic, and histological stratification	854 × 480, 1920 × 1080		3 433 images (2 131 WL and 1 302 NBI) from 76 lesions from 40 patients.	https://www.biobancovasco.org/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html it is necessary to fill out this form: https://labur.eus/EzJUN request sent
LDPolypVideo Ma Y. et al. 2021	video		Polyp locations (bounding box)	768 x 576 (videos), 560 × 480 (images)		160 videos (40 187 frames: 33 876 polyp images and 6 311 non-polyp images) with 200 labeled polyps. 103 videos (861 400 frames: 371 400 polyp images and 490 000 non-polyp images) without full annotations.	https://github.com/dashishi/LDPolypVideo-Benchmark *available but test and train folders are in rar format, cant open it
KUMC dataset Li K. et al. 2021	image jpg	color	Polyp locations (bounding box) Polyp classification: Adenoma vs. Hyperplastic	Various resolutions	PolypsSet/train2019/Image/.jpg PolypsSet/train2019/annotation/.xml train2019 or test2019 or val2019	80 colonoscopy video sequences. It also aggregates the CVC-ColonDB, ASU-Mayo Clinic Colonoscopy Video, and Colonoscopic Dataset datasets.	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FCBUOR PolypsSet.zip
CP-CHILD-A, CP-CHILD-B Wang W. et al. 2020	image jpg	RGB	Polyp detection: polyp vs. non-polyp annotations	256 × 256	CP-CHILD-A/Test/Polyp / jpg or nonPolyp/jpg then Train/ …. same for CP-CHILD-B	CP-CHILD-A contains 1 000 polyp images and 7 000 non-polyp images. CP-CHILD-B contains 400 polyp images and 1 100 normal or other pathological images.	https://figshare.com/articles/dataset/CP-CHILD_zip/12554042
SUN Misawa et al.			Polyp locations (bounding box)	N/A		49 136 images with polyps from	http://amed8k.sundatabase.or

2021						different 100 polyps. 109 554 non-polyp images from 13 video sequences.	g/ *can't find
Colorectal Polyp Image Cohort (PIBAdb)	Video and image		Polyp locations (bounding box) Polyp classification: Adenoma vs. Hyperplastic vs. Sessile Serrated Adenoma vs. Traditional Serrated Adenoma vs. Non Epithelial Neoplastic vs. Invasive	768 × 576		~31 400 polyp images (~22 600 WL and ~8 800 NBI) from 1 176 different polyps. ~17 300 non-polyp images (including ~2 800 normal-mucosa images and ~500 clean-mucosa images)	https://www.iisgaliciasur.es/home/biobanco/colorectal-polyp-image-cohort-pibadb/?lang=en *needs to fill a form and send it to **investigacion.pibadb@iisgaliciasur.es**
ENDOTEST Fitting et al. 2022	Video and image		Polyp locations (bounding box)	N/A		Validation dataset: 24 polyp and their corresponding non-polyp video sequences (22 856 images: 12 161 with polyps and 10 695 without polyps) Performance dataset: 10 full length colonoscopy videos with 24 different polyps (230 898 images).	
POLAR (POLyp Artificial Recognition) database	Image		Polyp locations (bounding box) Polyp classification: Adenomas vs. Hyperplastic vs. Sessile Serrated Adenoma	N/A		Training dataset: 2 637 non-magnified NBI images from 1 339 unique polyps detected during 555 different colonoscopies. Validation dataset: 730 polyps from 251 patients, prospectively collected by 20 endoscopists from 8 hospitals.	https://clinicaltrials.gov/study/NCT03822390 https://www.amc.nl/web/polar-database.htm *request sent

NBIPolyp-UCdb Figueiredo et al. 2019	Image		Polyp locations (binary mask)	576 × 720 pixels		86 NBI images from 11 colonoscopy videos.	https://www.mat.uc.pt/~isabelf/Polyp-UCdb/NBIPolyp-UCdb.html *link doesn't work
WLPolyp-UCdb Figueiredo et al. 2019 Figueiredo et al. 2020	Image		No ground truth provided	576 × 720 pixels		1 680 polyp images from 42 different polyps (40 images/polyp). 1 360 normal colonic mucosa images.	https://www.mat.uc.pt/~isabelf/Polyp-UCdb/WLPolyp-UCdb.html *link doesn't work
PolypGen Ali et al. 2023	Video and image		Polyp locations (binary mask).	N/A		1 537 polyp images, 2 225 positive video sequences, and 4,275 negative frames.	https://github.com/DebeshJha/PolypGen https://www.synapse.org/#!Synapse:syn26376615/wiki/613312 *need to login to download zip file

*Private datasets

Database+Reference	format of images/videos	Number & color	Ground truth	Resolution (w x h)	Method	Source of data	link
Ribeiro et al. 2016			Polyp classification (neoplastic vs non-neoplastic)			8 datasets by combining: (i) with or without staining mucosa, (ii) 4 acquisition modes (without CVC, i-Scan1, i-Scan2,	

						i-Scan3) 66 to 86 Patients, 85 to 126 images	
Zhang R. et al. 2017, Zheng Y. et al. 2018			Polyp classification (hyperplastic vs. adenomatous)			PWH Database. Images taken under either WL or NBI endoscopy. 1930 Without polyps: 1 104 Hyperplastic: 263 Adenomatous: 563,215 unique polyps (65 hyperplastic and 150 adenomatous)	
Tian Y. et al. 2019			hyperplastic polyp (TypeI), sessile serrated adenomas/polyp (TypeIIo), low grade adenoma/tubular adenoma (TypeII), high grade adenoma/tubulovillous adenoma/superficial cancer (TypeIIIa) and invasive cancer (TypeIIIb). 871 images MS I: 102 MS II: 346 MS IIo: 281 MS IIIa: 79 MS IIIb: 63			colonoscope	
Cheng Tao Pu et al. 2020			20 images MS I: 3 MS II: 5			colonoscope	

			MS IIo: 2 MS IIIa: 7 MS IIIb: 3				

Publicly available datasets were analyzed in this study. The CVC-ClinicDB, CVC-ColonDB, CVC-ClinicVideoDB, and CVC-PolypHD datasets are publicly available here: https://giana.grand-challenge.org. The ETIS-Larib dataset is publicly available here: https://polyp.grand-challenge.org/EtisLarib. The Kvasir-SEG dataset is publicly available here: https://datasets.simula.no/kvasir-seg. The PICCOLO dataset is publicly available here: https://www.biobancovasco.org/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html. The KUMC dataset is publicly available here: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FCBUOR. The SUN dataset is publicly available here: http://amed8k.sundatabase.org/. The LDPolypVideo dataset is publicly available here: https://github.com/dashishi/LDPolypVideo-Benchmark .