Cyclope

Presented by:
Imane El Missaoui
Student in second year Health Data Science/apprentice at LIP6.

Date of the presentation:
02/02/2024

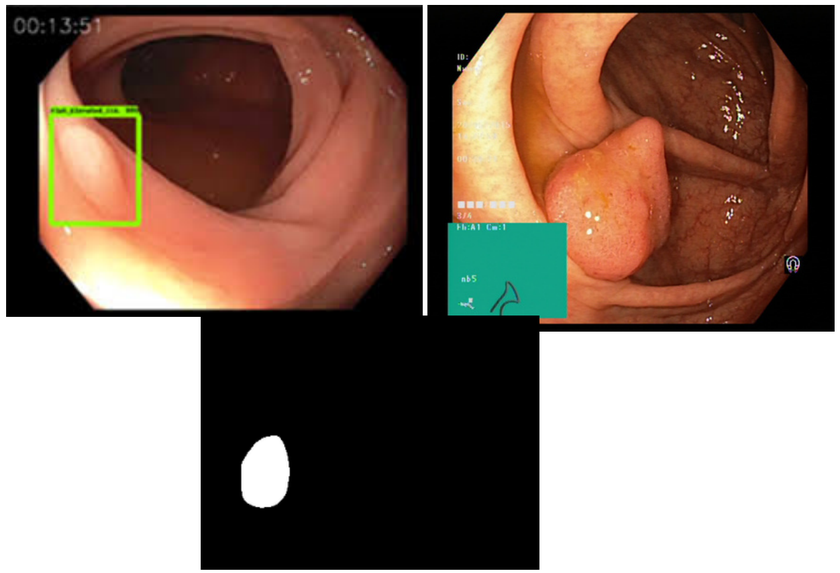
Main Objectives

- **Organization and Structuring of the Endoscopic Capsule Database**
- **Study and Development of Protocols for Medical Data Collection**
- **Study and Design of a Machine Learning Inference Model for polyps classification**

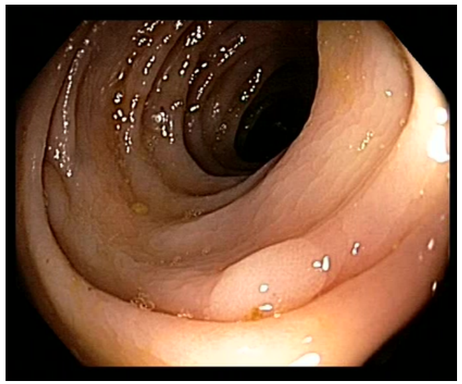
Data Collection

Research

- Polyp Detection and Localization
- Polyp Classification
- Simultaneous Polyp Detection and Classification



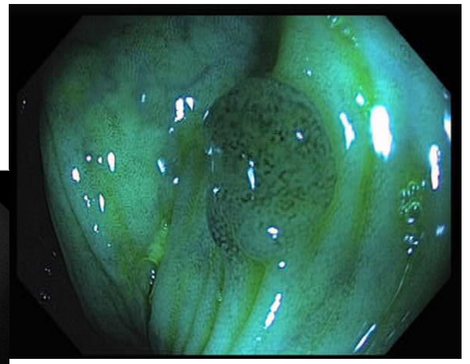
Annotation



White Light Frame

NBI Frame

Gray scale



Blue light

Different resolutions and color temperature + Different ways of Annotations

Public datasets :

	Format	Color	Resolution (w x h)	Annotation	Data	Endoscopy type	Link
Kvasir-Capsule Pua H and al 2021	jpg	RGB	336 × 336	Polyp locations (bounding box)Polyp classification : 14 different classes representing the labelled images	47,238 images with labels and bounding box masks, 43 corresponding labelled videos and 74 unlabelled videos	WCE	https://osf.io/dv2ag/
KUMC datasetLi K. et al. 2021	jpg	RGB	Various resolutions	Polyp locations (bounding box)Polyp classification: Adenoma vs. Hyperplastic	80 colonoscopy video sequences. It also aggregates the CVC-ColonDB, ASU-Mayo Clinic Colonoscopy Video, and Colonoscopic Dataset datasets.	Conventional	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FCBUORPolypsSet.zip
PICCOLO Sánchez-Peralta et al. 2020	tif	RGB	854 × 480, 1920 × 1080	Polyp locations (binary mask) Polyp classification, including: Paris and NICE classifications, Adenocarcinoma vs. Adenoma vs. Hyperplastic, and histological stratification	3 433 images (2 131 WL and 1 302 NBI) from 76 lesions from 40 patients.	Conventional	https://www.biobancovasco.org/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html Access requested* (accepted +needs documents)

Private datasets :

	Format	Color	Resolution (w x h)	Annotation	Data	Endoscopy type	Link
Cheng Tao Pu et al. 2020	NA	NA	NA	Polyp classification: [MS I: 3, MS II: 5, MS IIo: 2, MS IIIa: 7, MS IIIb: 3]	20 images	Conventional	https://www.sciencedirect.com/science/article/pii/S0016510720302182?via%3Dihub
Tian Y. et al. 2019	NA	NA	NA	Polyp Detection(bounding box) Polyp classification: 5 classes [MS I: 102, MS II: 346, MS IIo: 281, MS IIIa: 79, MS IIIb: 63]	871 images, 218 patients	Conventional	https://ieeexplore.ieee.org/document/8759521
Zhang R. et al. 2017, Zheng Y. et al. 2018	NA	NA	NA	Polyp classification (hyperplastic vs. adenomatous)	1930 Without polyps: 1 104 Hyperplastic /263 Adenomatous: 563,215 unique polyps (65 hyperplastic and 150 adenomatous) Images taken under either WL or NBI endoscopy.	Conventional	https://ieeexplore.ieee.org/document/8513337

Database design

Images and videos storage :

Binary Large Object (BLOB) Storage	File System Storage	Base64 Encoding
Images can be directly stored as binary data **Performance issues** 	The image or video is stored on the file system, and the database contains a reference to the location of the file **More flexible and fast** 	The image or video is encoded as a Base64 string and stored in a text column in the database. **Increases the size of the data stored in the database.**

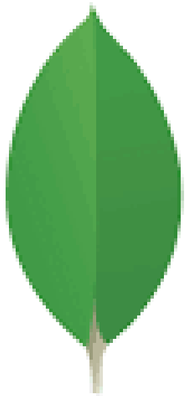
➡ Recommended : to store the file in a File System or cloud storage service and include the path and metadata in the database.

Database Management Systems

SQL

NoSQL

DBMS Choice



mongoDB

Pros:

- Performance Levels
- High Speed and Availability
- JSON-Like Documents
- Simple Request Syntax
- Ease of Use

Cons:

- Transactions (Limitations)
- Joins are not supported
- Indexing can be problematic
- High Memory Usage (RAM)

MongoDB Compass - localhost:27017/Cyclope.WCEcuratedcolon_dataset

Connect Edit View Collection Help

localhost:27017

My Queries

Performance

Databases

Search

Cyclope

PolypsSet_dataset

WCEcuratedcolon_d...

admin

config

local

My Queries | Cyclope | WCEcuratedcolon_dataset

Cyclope.WCEcuratedcolon_dataset

6.0k DOCUMENTS

1 INDEXES

Documents | Aggregations | Schema | Indexes | Validation

Filter | Type a query: { field: 'value' } or Generate query | Explain | Reset | Find | Options

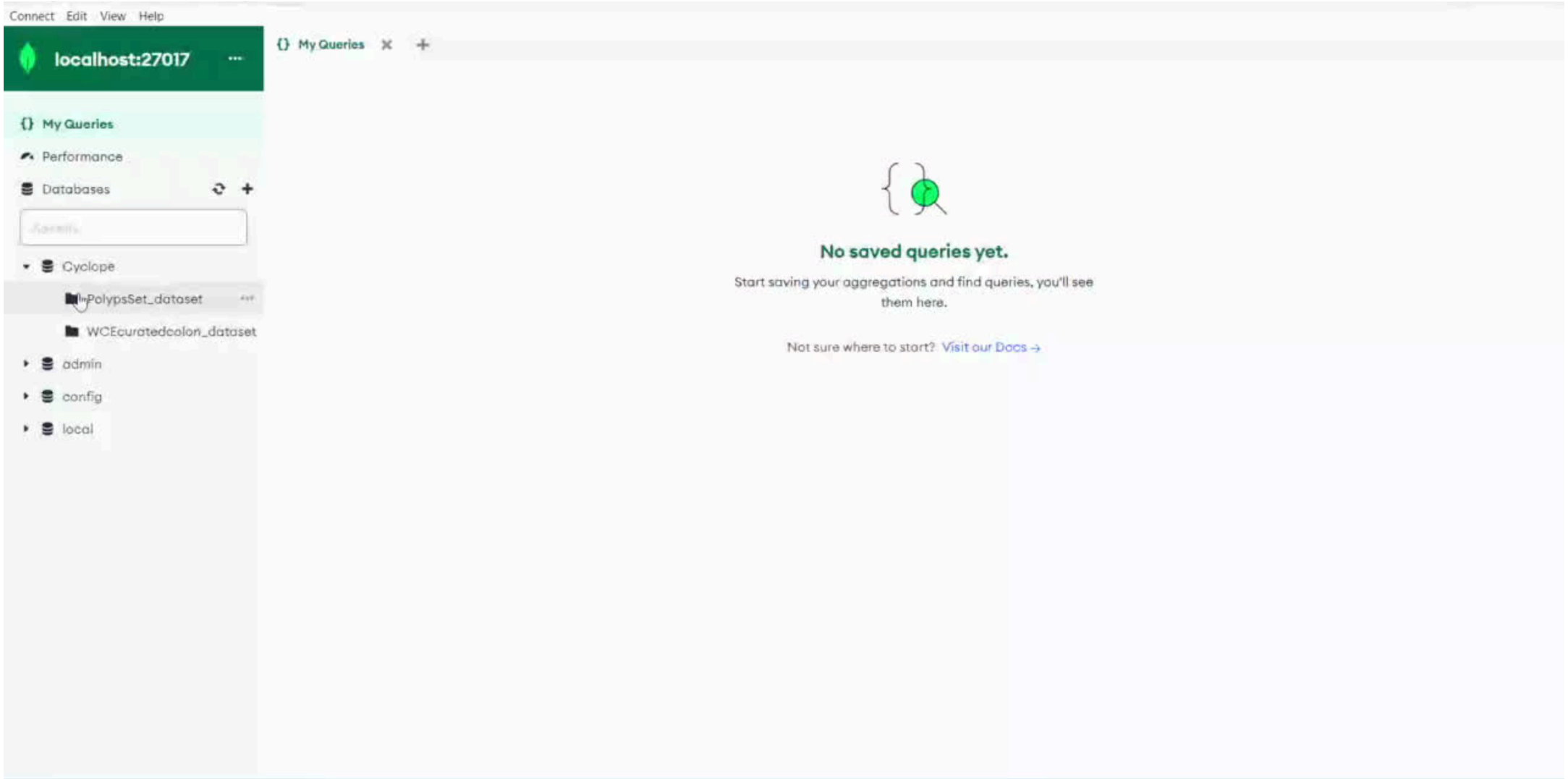
ADD DATA | EXPORT DATA | 1 – 20 of 6000

```
_id: ObjectId('65bbfd152472175d4d531d3b')
filename: "train_normal_ (1).jpg"
format: "JPEG"
dimensions: Array (2)
  0: 615
  1: 531
dataset: "WCEcuratedcolondataset"
path: "train/0_normal/train_normal_ (1).jpg"
label: "normal"
```

```
_id: ObjectId('65bbfd152472175d4d531d3c')
filename: "train_normal_ (10).jpg"
format: "JPEG"
dimensions: Array (2)
  0: 633
  1: 531
dataset: "WCEcuratedcolondataset"
path: "train/0_normal/train_normal_ (10).jpg"
label: "normal"
```

normal

```
_id: ObjectId('65bbfd152472175d4d531d3d')
filename: "train_normal_ (100).jpg"
format: "JPEG"
dimensions: Array (2)
```

>_MONGOSH

Mongodb demo :