Investigation into the Categorization of Crohn's Segments from Capsule Endoscopy Videos:

Introducing A Thick Data Categorization Framework

Jinan Fiaidhi
Department of Computer Science
Lakehead University
Thunder Bay, Canada
jfiaidhi@lakeheadu.ca

Sabah Mohammed
Department of Computer Science
Lakehead University
Thunder Bay, Canada
mohammed@lakeheadu.ca

Petros Zezos
Northern Ontario School of Medicine
NOSM University
Thunder Bay, Canada
pzezos@nosm.ca

2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) | 979-8-3503-9836-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IRASET57153.2023.10153022

***Abstract*—Crohn's disease (CD) is a complex disorder that causes patchy inflammation and ulcerative effects in the gut from mouth to anus, which requires diagnosis for proper treatment. Capsule endoscopy (CE) procedure emerged as important investigation for screening the small bowel and the entire colon. However, examining the CE videos by gastroenterologists is a tedious and time consuming task as the capsule camera takes more than eighty thousand frames, with a frequency rate of 2 frames per second, for one segment of the gastrointestinal tract. However, having an automatic way to categorize the patchy frames that hold the markers of Crohn's disease will help the examiners to provide more reliable diagnosis and treatment. In this article, we are describing a thick data analysis framework that can help categorizing Crohn's disease sequences through the use of triplet-loss Siamese neural network that can learn from few shots provided by the expert to detect frames with markers of the Crohn's disease as well as to eliminate those frames having reduced mucosal view. Additionally, the framework uses a fuzzy filler to produce intermediate sets of frames having similar markers. These sets can be annotated and returned back to their position at the original CE video. The use of our framework showed promising results in re-orienting the original CE video into marked sequences of Crohn's disease frames. We have trained and tested our framework using two notable CE datasets (KVASIR Capsule and CrohnIPI) that contain the expert annotation on related markers of Crohn's disease.**

Keywords—Thick Data Analytics; Siamase Neural Network, Triplet-Loss Functions, Crohn's Disease Categorization

I. INTRODUCTION

Video capsule endoscopy (CE) since its introduction 25 years ago, has revolutionized our way to examine the gastrointestinal tract and offers gastroenterologists a new noninvasive modality in assessing the gastrointestinal tract for a variety of gastrointestinal disorders including gastrointestinal cancer, Crohn's disease, celiac disease, unexplained bleeding [1]. Unlike the traditional wire-based video endoscopy which requires pushing wide cables, the flexible endoscopes, into the gastrointestinal tract to examine far areas, the capsule endoscopy has been proven to be an important examination method for far-reaching areas like the small intestine. However, examining the CE videos by gastroenterologists is a tedious and time consuming task as the capsule camera takes more than eighty thousand frames, with a frequency rate of 2 frames per second, for one segment of the gastrointestinal tract. Thus, there is an urgent need to develop an automatic categorization of anomalies in CE videos [2]. However, detecting video segments in stored video and breaking it up into smaller clips for better indexing is a normal task used in the cinema and media industry [3]. In the movie industry, they use automatic detection software like the Amazon Rekognition[1] software to segment videos into scenes and frames as well as to filter some video contents. Figure 1 illustrates the categorization process used in the movie industry. The purpose of the automatic categorization for movie industry and video surveillance applications is providing a high-quality viewing experience and better monetizing content categorizing. They use techniques like [4]:

- Finding where the opening and end credits are in a piece of content
- Choosing the right spots to insert advertisements, such as in silent black frame sequences
- Breaking up videos into smaller clips for better indexing

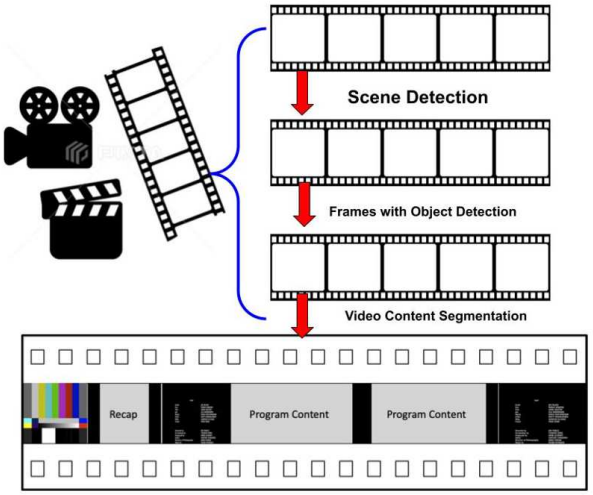


Fig. 1: Video Categorization used in Movie Industry.

Experiments utilizing similar techniques used at the movie industry applied to medical videos like CE revealed great limitations in identifying anomalies like Crohn's disease segments in capsule endoscopy videos [5]. The best published experiment in this direction was reported by [6] in which a convolutional neural network was trained to classify capsule images into either normal mucosa or mucosal ulcers trained with large dataset that include 17,640 CE images from 49 patients: 7391 images with mucosal ulcers and 10,249 images of normal mucosa. The results of categorizing images of three

[1] https://aws.amazon.com/rekognition/

979-8-3503-9836-6/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.

patients CE videos is provided in figure 2. However, using quantitative scoring techniques like the Lewis score12 [7] and the Capsule Endoscopy Crohn's Disease Activity Index [8] failed to accurately identify Crohn's disease segments from the identified ulcerative frames [9] due to the complexity of these scoring techniques and the inter-observer variability of the scoring results. Actually, similar scoring systems were used on other CE to detect anomalies like celiac disease, polyps, and hookworm infection reported high accuracy detection with impressive specificity and sensitivity [10].

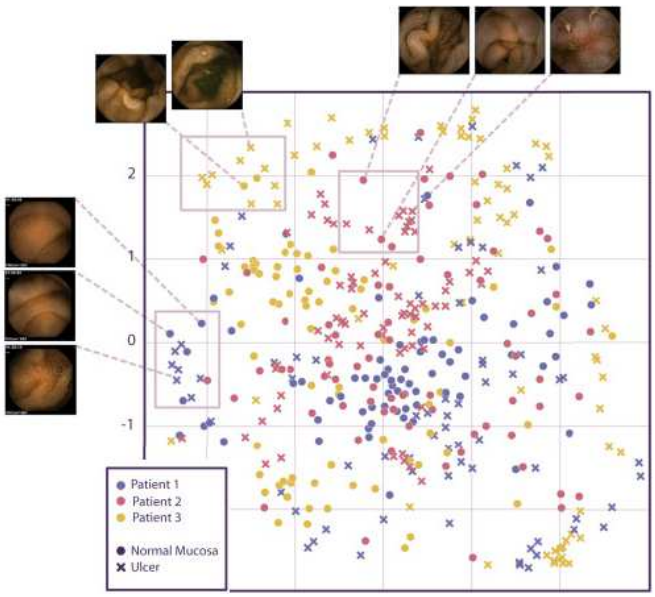


Fig. 2: Capsule Video Categorization based on Deep Learning Techniques.

As illustrated, the use of machine learning and content video content detection borrowed from movie industry cannot accurately identify Crohn's disease segments due to their nature of being scattered throughout the small bowel and/or the colon and the variability of these patterns (see Figure 3). What we actually need is a method that combines several heuristics and learning mechanisms including those borrowed from image augmentation, physician experience in sensing Crohn's patterns as well as the ability to learn from few similar shots provided by the expert. We are calling this approach, a thick data analytics [11, 12].

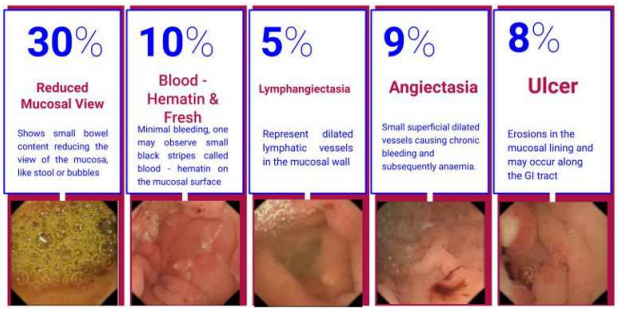


Fig. 3: Percentage of Common Frames Found in any CE Video Screening.

In this article we are performing an investigation on how to categorize Crohn's disease segments from CE videos reliably based on thick data analytics that can learn from small dataset samples.

II. Crohn's Disease Thick Data Analytics

Crohn's disease belongs to a group of conditions known as inflammatory bowel diseases (IBD). It is named after Dr. Burrill B. Crohn, who first described the disease in 1932. It can affect any part of the gastrointestinal tract from the mouth to the anus, but most commonly affects the end of the small bowel (ileum) and the beginning of the colon. It can affect the entire thickness of the bowel wall and this inflammation of the intestine can "skip," or leave normal areas in between patches of diseased intestine (see figure 4).

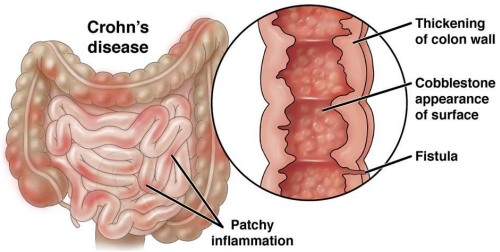


Fig.4: Illustration of the Patchy Nature and the Variability of Crohn's disease.

Due to this variability and the patchy nature, the use of deep learning techniques alone to categorize Crohn's disease is becoming a challenge as the neural net hidden black boxes are unable to learning the non-linear distribution nature and variability of the Crohn's disease patchy patterns types even the availability of large training samples [13]. This challenge pushes for the use of new combined venues of AI and machine learning that can accommodate technologies utilizing both qualitative and quantitative learning. This means incorporating the human and expert intelligence to boost the machine learning or what we call the Thick Data analytics. Thick data techniques may use variety of mechanisms including object detection, image augmentation techniques, domain heuristics, few shots learning techniques to learn from few samples and explainable AI (xAI) techniques to associate white boxes with the neural nets black boxes for full comprehension of the detection nature. Figure 5 illustrates the techniques involved in thick data analytics.

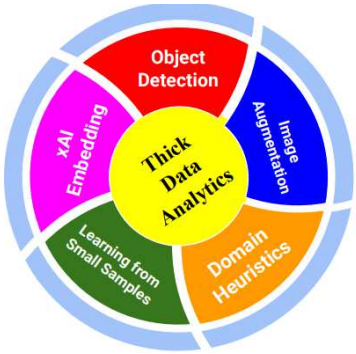


Fig. 5: Major Participants of Thick Data Analytics.

In this article we are proposing a four-stage investigation to use some of the thick data analytics approaches for categorizing anomalies related to Crohn's disease in CD videos:

- **Detection of Reduced Mucosal View Frames**: Detecting and identifying these frames will eliminate the noise provided to the machine learning algorithms. We are proposing to use the triplet-loss Siamese neural net

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.

to learn from few shots of reduced mucosal view frames to identify the other similar frames [11].

- **Detection of Crohn's Disease Markers:** Identifying endoscopic features that may relate to the presence of Crohn's disease like ulcerative areas, cobblestones, fistula areas and other markers in any of the CE frames, will play important role in categorizing Crohn's disease segments. We are proposing to use the triplet-loss Siamese neural net to learn from the few anchor images that contain these markers and provided by the expert [11].
- **Categorization of Crohn's Disease Frame Sets:** By using temporal modeling techniques that use a fuzzy filter to categorize frames into three appropriate category sets (Highly related, Mid related, Low related) for each Crohn's Disease marker [1].
- **Memorization of the Frame Sets Sequence**: Memorization algorithm refers to remembering the original frames locations from the categorized frame sets and returning the remembered sequence. It is like a cache method that return the original location of every frame from the categorized three sets (Highly related, Mid related, Low related) of every marker to construct the original sequence of frames associated with their categories. For this purpose we used the Fast-Backward Replay algorithm [14]. Figure 6 illustrates the proposed approach.

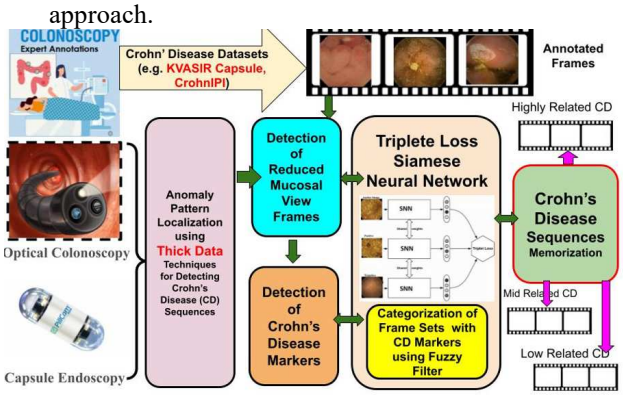


Fig. 6: The proposed 4 Stage Thick Data Categorization Approach.

The annotated dataset that contains the expert annotation on capsule video frames indicating the presence of Crohn's disease and other markers are provided by two notable datasets (KVASIR Capsule [15] and CrohnIPI [16]).

A. Detecting Reduced Mucosal Views and Crohn's Disease Markers through Triplet Loss Siamase Net

Availability of extremely large datasets to build pre-trained models is a challenge in the medical area as most of the annotated samples are small. The use of deep learning in this case will produce models that cannot predict with acceptable accuracy any anomaly. Solving this challenge requires the search for models that can learn from the expert based on few shots to detect anomalies and predict with high accuracy future inputs. This kind of neural model is called Siamese Neural Network (SNN) [17]. Using the Siamese networks, the algorithm take an input frame of Crohn's disease marker and find out the encodings of that image, then, we take the same network without performing any updates on weights or biases and input from the provided test images and again predict it's encodings. The algorithm then compares these two encodings to check for their similarity. However, the triple loss Siamese neural network is utilizing advanced models that employ three Siamese networks and requires the expert to provide anchor image (a) for every Crohn's disease marker as well as a positive near (p) match image and a negative match that have no related relevance to the anchor image (n). It uses a triple loss function that can be mathematically represented as follows:

$$\textbf{Loss=max(distance(a,p)}-\textbf{distance(a,n)+margin,0).}$$

The learning algorithm will use the triplete loss model so minimize it by finding the distance(a,p) to be near 0 and to maximize the distance(a,n) to be greater than distance(a,p)+margin. This triple loss model will detect positives the positive frames will be closer to the anchor frames while the negative frames will be far from it. Figure 7 illustrate the structure of the triple loss Siamese neural network model that can learn based on few shots the presence of Crohn's disease markers or the presence of other markers like the reduced mucosal view frames.

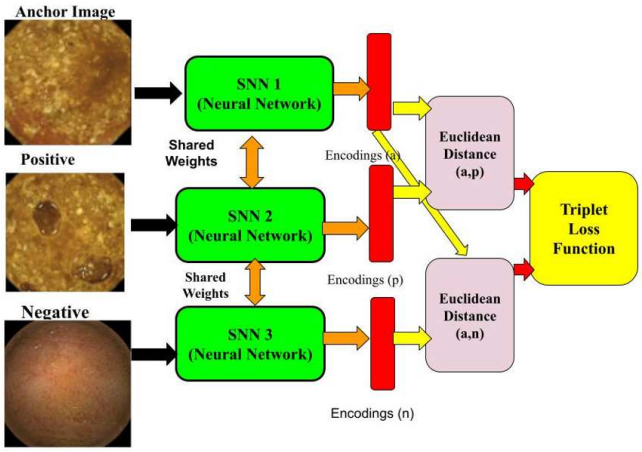
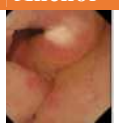
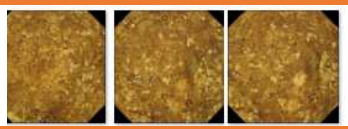
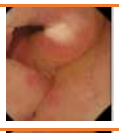
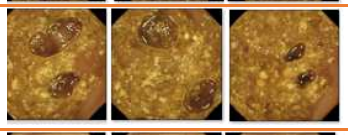
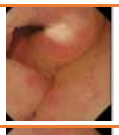
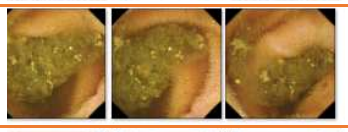
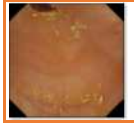
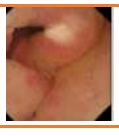
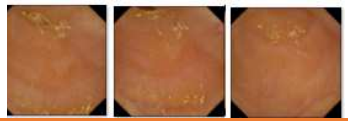


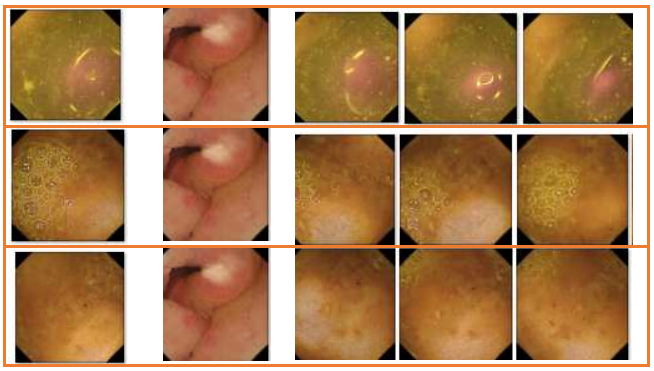
Fig. 7: The Structure of the Triplet Loss Siamese Neural Network.

Table 1 illustrate our experimentation with using the triple loss Siamese neural network to learn the presence of reduced mucosal view frames in CE videos.

Table 1: Training Triplet-Loss Siamese Neural Network to Detect Reduced Mucosal Frames

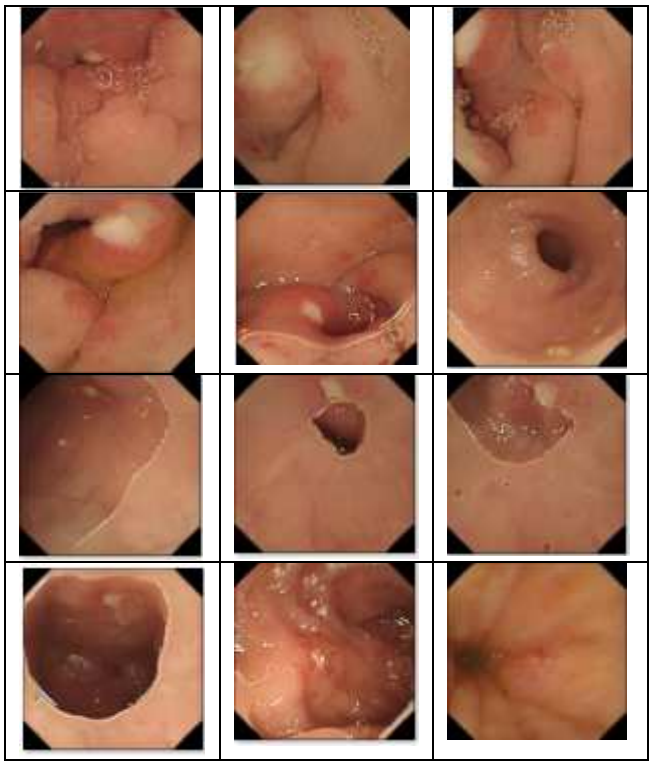
Positive Anchor	Negative Anchor	Similar Images >= 90%
		
		
		
		

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.



However, for training the triplet-loss Siamese neural network to detect Crohn's disease markers like frames with ulcers we use the following anchor images provided by the two annotated datasets (see Table 2).

Table 2: The Expert Anchor Images for Ulcerative Frames.



The python code snippet for the triplet-loss Siamese neural network is described as follows:

```
class MC_DistanceLayer(layers.Layer):
    def __init__(self, **kwargs):
        super().__init__(**kwargs)
    def call(self, anchor, positive, negative):
        ap_distance = tf.reduce_sum(tf.square(anchor - positive), -1)
        an_distance = tf.reduce_sum(tf.square(anchor - negative), -1)
        return (ap_distance, an_distance)

anchor_input = layers.Input(name="anchor", shape=target_shape + (3,))
positive_input = layers.Input(name="positive", shape=target_shape + (3,))
negative_input = layers.Input(name="negative", shape=target_shape + (3,))

distances = MC_DistanceLayer()(
    embedding(resnet.preprocess_input(anchor_input)),
    embedding(resnet.preprocess_input(positive_input)),
    embedding(resnet.preprocess_input(negative_input)),
)

siamese_network = Model(
    inputs=[anchor_input, positive_input, negative_input], outputs=distances
)
```

We experimented with training our triplet loss Siamese neural network by providing ten set of images (anchor, positive and negative) and the training and testing accuracies reached an average of 60% (See figure 8). Certainly, by providing more expert triplets we are expected the accuracy to reach higher than our present result.

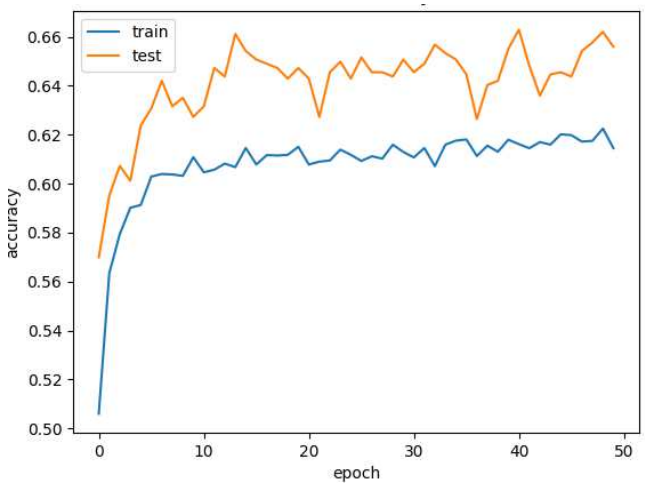


Fig. 8: Accuracy of the Triplet-Loss for Detecting Crohn's Disease Markers.

B. Localization of Crohn's Disease Frame Sequence:

Although the SNN filtered out segments that have similarity to the expert anchor images, this similarity may have different ranges according to the trained sample. For this purpose, we are proposing one more layer to be added after the SNN triplet-Loss Function is to add a Fuzzy filter that can gauge these similarity ranges into clusters of high, mid and low similarity. The fuzzy filter can be modeled using fuzzy triangular functions as described in figure 9.

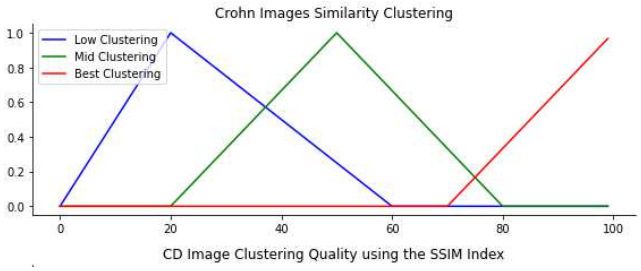


Fig. 9: Fuzzy Clustering Filter to Group Highly Similar Frames with Crohn's Disease Markers.

The python code snippet of this fuzzy filter is illustrated below:

```
import numpy as np
import skfuzzy as fuzz
import matplotlib.pyplot as plt
x_qual = np.arange(0, 100, 1)
x_serv = np.arange(0, 60, 1)
x_tip  = np.arange(0, 26, 1)
# Generate fuzzy membership functions
qual_lo = fuzz.trimf(x_qual, [0, 20, 60])
qual_md = fuzz.trimf(x_qual, [20, 50, 80])
qual_hi = fuzz.trimf(x_qual, [70, 100, 100])
fig, (ax0, ax1, ax2) = plt.subplots(nrows=3, figsize=(8, 9))
ax0.plot(x_qual, qual_lo, 'b', linewidth=1.5, label='Low Clustering')
ax0.plot(x_qual, qual_md, 'g', linewidth=1.5, label='Mid Clustering')
ax0.plot(x_qual, qual_hi, 'r', linewidth=1.5, label='Best Clustering')
ax0.set_title('Crohn Images Similarity Clustering')
ax0.legend()
ax0.spines['top'].set_visible(False)
ax0.spines['right'].set_visible(False)
ax0.get_xaxis().tick_bottom()
ax0.get_yaxis().tick_left()
plt.tight_layout()
```

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.

C. *Memorization of the Frames Sequence:*

Storing and retrieving temporal sequences of marked frames is a crucial to task for cognitive system like the thick data approach that we are proposing. To accomplish this task, two core components on the frame sequences need to be stored: the marked frames and their temporal order. For this purpose, we have adopted the Fast-Backward Replay algorithm [14, 18, 19]. During the algorithm *encoding period* involving detecting and localization of CD frames, the frames are labeled with the following CD markers:

- Erosion appearance (A)
- Ulceration appearance (E)
- Stenosis appearance (B)
- Edema appearance (D)
- Cobblestone appearance (C)

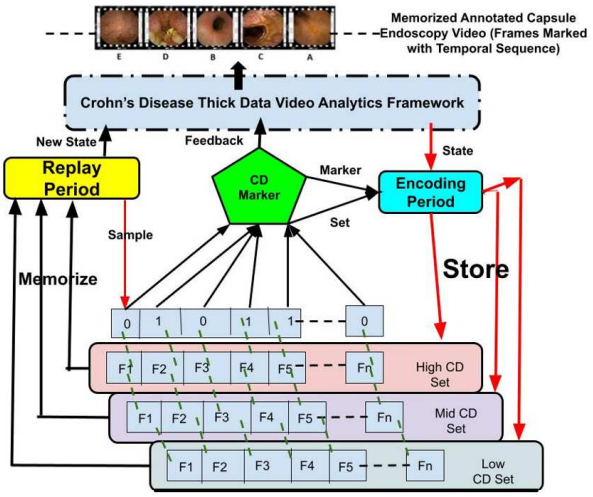
During the encoding period, the temporal timestamp is stored with every frame (e.g. D(t) C(t) A(t) E(t) B(t)). However, during the *recalling or replay period* of the algorithm assign the recorded mark and move the frame to its relative position of the generated video. The memorization algorithm is best described using figures 10(a) and 10 (b).

Algorithm 1 CD Sequence Memorization Algorithm

Require: Access the Stored Sets from the Fuzzy Filter
Ensure: Updating the CD Video with the CD Marked Frames

```
HighCDMatch ← HighSet
MidCDMatch ← MidSet
LowCDMatch ← LowSet
while iteration do
    while timestamp do
        if StorePeriod is DONE then
            Replay ← Mark × Time
            NewSTATE ← Replay(Update)
        else if StorePeriod is ACTIVE then
            DiscardReducedMucosalView
            ActivateFuzzyFilter ← Frame
            DetectCDmark ← Mark
            StoreFrame ← Mark × TimeStamp
            UpdateSEQUENCES ← High × Mid × Low
        end if
    end while
end while
```

(a)



(b)

Fig. 10: Crohn's Disease Memorization Algorithm.

III. Conclusions

Crohn's disease (CD) is a complex inflammatory bowel disorder which exhibits significant clinical and endoscopic heterogeneity due to variation of the markers captured by the lens of the capsule endoscopy camera. Existing categorization and classification methods used for CD are unreliable due to the high dependency of neural nets to determine this classification. In this article, we are describing an investigation into the use thick data analytics approach to enhance the reliability of such CD categorization and classification by adding the expert heuristics even with the availability of small data. Our described framework utilizes the triplet-loss Siamese Neural Net that can learn from few anchor images provided by the expert to eliminate noisy frames like those with reduced mucosal view as well as to detect important markers contributing to CD categorization (e.g. Erosion, Ulceration, Stenosis, Edema and Cobblestone severe ulceration). The framework uses a fuzzy classifier to stream frames of each marker into three sequences (Highly related, Mid related and Low related). The memorization component can associate these sets with their location at the original capsule video with an index of their marker as index. The expert judgment and the annotated capsule videos used in our investigation was from two notable datasets (KVASIR Capsule and the CrohnIPI). We are continuing our research to test the framework on our endoscopic data as soon as the ethical approval has been granted between our collaborating institutions (Lakehead and NOSM).

Acknowledgment

The first two authors acknowledge the financial support to this research project from MTACS Accelerates Grant (IT22305-2020) and the first author NSERC DDG Grant (DDG-2020-00037).

References

[1] Jinan Fiaidhi, Sabah Mohammed and Petros Zezos, An Investigation into Crohn's Disease Lesions Variability Sensing Using Video Colonoscopy and Machine Learning Techniques) to 2022 IEEE International Conference on E-health Networking, Application & Services (IEEE HealthCom2022), Genoa, Italy Oct. 17-19, 2022

[2] Goenka, Mahesh K., Shounak Majumder, and Usha Goenka. "Capsule endoscopy: Present status and future expectation." World Journal of Gastroenterology: WJG 20, no. 29 (2014): 10024.

[3] Ionescu, Bogdan, Patrick Lambert, Didier Coquin, Alexandru Marin, and Constantin Vertan. "Analyzing Animated Movie Contents for Automatic Video Indexing." In Machine Learning Techniques for Adaptive Multimedia Retrieval: Technologies Applications and Perspectives, pp. 228-260. IGI Global, 2011.

[4] Haq, Ijaz Ul, Khan Muhammad, Tanveer Hussain, Soonil Kwon, Maleerat Sodanil, Sung Wook Baik, and Mi Young Lee. "Movie scene segmentation using object detection and set theory." International Journal of Distributed Sensor Networks 15, no. 6 (2019): 1550147719845277.

[5] Klang, Eyal, Ana Grinman, Shelly Soffer, Reuma Margalit Yehuda, Oranit Barzilay, Michal Marianne Amitai, Eli Konen et al. "Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks." Journal of Crohn's and Colitis 15, no. 5 (2021): 749-756.

[6] Klang, Eyal, Yiftach Barash, Reuma Yehuda Margalit, Shelly Soffer, Orit Shimon, Ahmad Albshesh, Shomron Ben-Horin, Marianne Michal Amitai, Rami Eliakim, and Uri Kopylov. "Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy." Gastrointestinal endoscopy 91, no. 3 (2020): 606-613.

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.

[7] Monteiro, Sara, Pedro Boal Carvalho, Francisca Dias de Castro, Joana Magalhães, Firmino Machado, Maria J. Moreira, Bruno Rosa, and José Cotter. "Capsule endoscopy: diagnostic accuracy of Lewis score in patients with suspected Crohn's disease." Inflammatory bowel diseases 21, no. 10 (2015): 2241-2246.

[8] Niv, Yaron, Eyal Gal, Violeta Gabovitz, Marcela Hershkovitz, Lev Lichtenstein, and Irit Avni. "Capsule endoscopy Crohn's disease activity index (CECDAIic or Niv Score) for the small bowel and colon." Journal of Clinical Gastroenterology 52, no. 1 (2018): 45-49.

[9] Lai, Larry H., Grace LH Wong, Dorothy KL Chow, James YW Lau, Joseph JY Sung, and Wai K. Leung. "Inter-observer variations on interpretation of capsule endoscopies." European journal of gastroenterology & hepatology 18(3), (2006): 283-286.

[10] Hwang, Youngbae, Junseok Park, Yun Jeong Lim, and Hoon Jai Chun. "Application of artificial intelligence in capsule endoscopy: where are we now?." Clinical endoscopy 51, no. 6 (2018): 547-551.

[11] Fiaidhi, Jinan, Darien Sawyer, and Sabah Mohammed. "Thick Data Analytics for Small Training Samples Using Siamese Neural Network and Image Augmentation." In LISS 2021, pp. 57-66. Springer, Singapore, 2022.

[12] Fiaidhi, Jinan, Petros Zezos, and Sabah Mohammed. "Thick Data Analytics for Rating Ulcerative Colitis Severity Using Small Endoscopy Image Sample." In 2021 IEEE International Conference on Big Data (Big Data), pp. 4687-4693. IEEE, 2021.

[13] Ben-Gal, Irad. "Outlier detection." In Data mining and knowledge discovery handbook, pp. 131-146. Springer, Boston, MA, 2005.

[14] Huang, Qiaoli, Jianrong Jia, Qiming Han, and Huan Luo. "Fast-backward replay of sequentially memorized items in humans." Elife 7 (2018), Available Online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6231774/pdf/elife-35164.pdf

[15] Smedsrud, Pia H., Vajira Thambawita, Steven A. Hicks, Henrik Gjestang, Oda Olsen Nedrejord, Espen Næss, Hanna Borgli et al. "Kvasir-Capsule, a video capsule endoscopy dataset." Scientific Data 8, no. 1 (2021): 1-10.

[16] Vallée, Rémi, Astrid De Maissin, Antoine Coutrot, Harold Mouchère, Arnaud Bourreille, and Nicolas Normand. "CrohnIPI: An endoscopic image database for the evaluation of automatic Crohn's disease lesions recognition algorithms." In Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 11317, pp. 440-446. SPIE, 2020.

[17] Melekhov, Iaroslav, Juho Kannala, and Esa Rahtu. "Siamese network features for image matching." In 2016 23rd international conference on pattern recognition (ICPR), pp. 378-383. IEEE, 2016.

[18] Hayes, Tyler L., Giri P. Krishnan, Maxim Bazhenov, Hava T. Siegelmann, Terrence J. Sejnowski, and Christopher Kanan. "Replay in deep learning: Current approaches and missing biological elements." Neural Computation 33, no. 11 (2021): 2908-2950.

[19] Zha, Daochen, Kwei-Herng Lai, Kaixiong Zhou, and Xia Hu. "Experience replay optimization." arXiv preprint arXiv:1906.08387 (2019).

Authorized licensed use limited to: Lakehead University. Downloaded on July 02,2025 at 17:14:38 UTC from IEEE Xplore. Restrictions apply.