Overview of ImageCLEFmedical 2024 – Medical Visual Question Answering for Gastrointestinal Tract

Steven Hicks[1,*], Andrea Storås[1,2], Pål Halvorsen[1,2], Michael Riegler[1] and Vajira Thambawita[1]

[1]*SimulaMet - Simula Metropolitan Center for Digital Engineering, Oslo, Norway*
[2]*OsloMet - Oslo Metropolitan University, Oslo, Norway*

Abstract

This paper provides details on the second edition of the Medical Visual Question Answering for the Gastrointestinal Tract (MedVQA-GI) challenge, which took place during ImageCLEF 2024. This year, we changed the task from visual question answering to the application of text-to-image models for the creation of synthetic medical images. There were two sub-tasks in this challenge. The first sub-task involved using prompts to generate realistic looking images from the gastrointestinal tract. The second sub-task focused on the technical aspects involved in the implementation of these models, and optimizing the prompts to generate realistic-looking images using a low number of tokens. Despite considerable interest in the task, the rate of submissions remained low, suggesting that participants may have encountered barriers or found the task too complex to complete.

Keywords

Machine learning, medical ai, endoscopy

1. Introduction

The second edition of the Medical Visual Question Answering for the Gastrointestinal Tract (MedVQA-GI) challenge at ImageCLEF [1] introduces a new goal that focuses on the use of generative models of text-to-image in medical diagnosis. This combines natural language processing and image generation to potentially improve diagnostic processes in healthcare by providing more comprehensive datasets that can be used for training machine learning models. In contrast to last year's focus on a Visual Question Answering (VQA) task that required retrieving images or masks from user questions, this year's overall goal was to use generative models to create synthetic medical images from textual inputs. Participants were tasked with generating synthetic images using existing generative models developed using a dataset derived from last year's MedVQA-GI challenge [2].

Machine learning has been a common method used to identify lesions in gastrointestinal (GI) images [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Traditionally, the emphasis in GI analysis has been on disease detection from images or videos, focusing mostly on polyp detection [14, 15, 16, 17, 18, 19, 20]. Several challenges have demonstrated consistent advancements in this field, including some challenges we have organized in the past [21, 22, 23, 24, 25]. However, there has been a growing interest in extending the capabilities of image analysis GI through the generation of synthetic images [26, 27]. This new focus aims to develop models that generate realistic GI images that can be used in-place of real data. Such synthetic images can be used to train medical professionals, refine diagnostic algorithms without the privacy concerns of real patient data, and improve the interpretability and reliability of AI systems in clinical settings. To this end, this year's MedVQA-GI focuses on synthetic GI image generation. The dataset and the scripts used to verify and evaluate submissions are available in our public GitHub repository[1].

The remainder of this paper is organized as follows. First, we start with an explanation of the creation of the dataset, looking at how the data was collected and organized. Then, we discuss the specific sub-tasks involved in the MedVQA-GI challenge and the evaluation methods used. Finally, we present statistics on the participants and the results of the submitted runs.

---

*Corresponding author.

✉ steven@simula.no (S. Hicks); andrea@simula.no (A. Storås); paalh@simula.no (P. Halvorsen); michael@simula.no (M. Riegler); vajira@simula.no (V. Thambawita)

[1]https://github.com/simula/imageCLEFmed-MEDVQA-GI-2024

Generate an image containing a polyp.

Generate an image from a colonoscopy procedure.

Generate an image containing text.

Generate an image containing oesophagitis.

Generate an image from a gastroscopy procedure.

Generate an image containing metal clip.

Generate an image containing biopsy forceps.

Generate an image containing the z-line.

Generate an image with 2 findings.

Generate an image containing tube.

Generate an image with a polyp of size < 5mm.

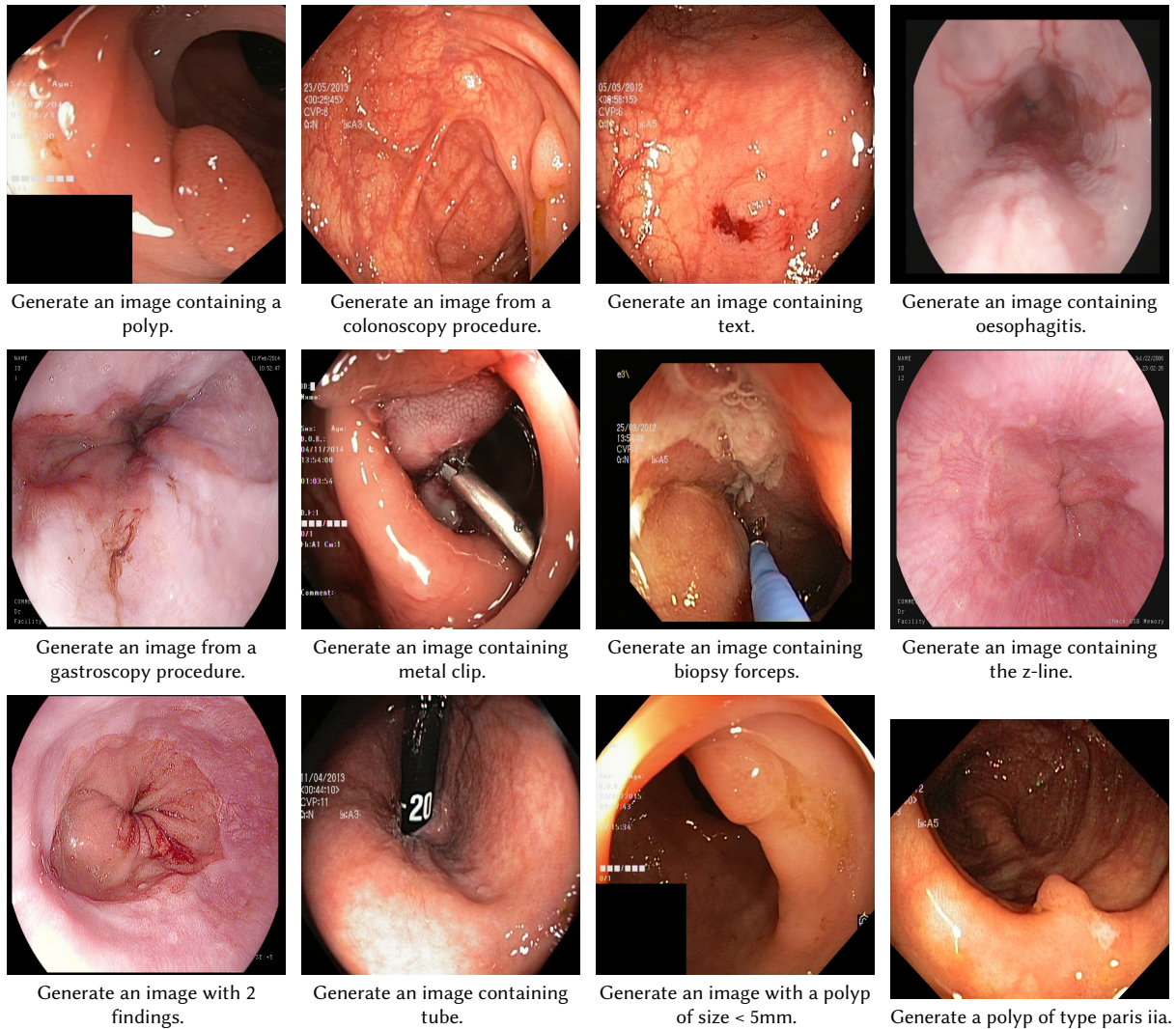Generate a polyp of type paris iia.

Figure 1: Examples from the development dataset that was provided by the challenge organizers. The samples represent different types of images contained within the dataset. Under each image is a prompt that is associated with an image. Note that there can be multiple prompts that match to the same image.

2. Dataset

The dataset used for this challenge is based on data developed for last year's challenge, which was based on the HyperKvasir dataset [28] and the Kvasir-Instrument dataset [29] datasets. Participants were provided with a dataset consisting of $2,000$ image and text pairs, which was organized in a directory containing the images and CSV files with prompts and connections to the image filenames. Example images and corresponding prompts can be seen in Figure 1.

3. Task Description and Evaluation

This year, participants could participate in two sub-tasks: Image Synthesis and Optimal Prompt Generation. Participants could submit to either sub-task and were not limited to the number of submissions.

3.1. Sub-task 1: Image Synthesis

The first sub-task, Image Synthesis, involves using text-to-image generative models to construct a comprehensive dataset of medical images from textual descriptions. This sub-task requires participants

to create accurate visual representations of various medical conditions described solely in text. For example, with a description such as "An early-stage colorectal polyp," participants must generate an image that precisely reflects the given text. Participants could use the development dataset, described in Section 2, to develop their models.

For the submission, each participant received a list of $5,000$ prompts. They were required to create synthetic images based on these prompts and submit them to the organizers by email. Each submitted image file was named according to the prompt's index number from the list. The quality of the synthetic images was assessed using two metrics: the Inception Score (IS) [30] and the Fréchet Inception Distance (FID) [31]. These metrics evaluated how the synthetic images were compared with three distinct testing datasets. The first data set consisted of images from the previous year's MedVQA-GI challenge. The second dataset was GastroVision [32], which is a newly released open-source collection that includes $8,000$ images obtained from various medical centers. The final data set used for the evaluation was a combination of the first two datasets.

3.2. Sub-task 2: Optimal Prompt Generation

The second sub-task, Optimal Prompt Generation, focuses on participants creating their own prompts to generate images that meet specific medical imaging requirements. This sub-task asked participants to tailor their prompt generation skills to produce images that accurately match a set of predefined categories. These categories are designed to test the model's ability to generate precise and clinically relevant images based on the prompts. Participants had to devise prompts for:

- A prompt that generates an image containing n polyps.
- A prompt that generates a polyp in a specific region of the image.
- A prompt that generates a polyp of a specific type and size.
- A prompt that generates an image containing no findings from either the esophagus or large bowel.
- A prompt that generates an image containing one of the following instruments: biopsy forceps, metal clip, and tube.
- A prompt that generates an image containing one of the following anatomical landmarks: Z-line, Pylorus, Cecum.

For evaluation, the effectiveness of each prompt was evaluated not only on the accuracy of the image it produced but also on the conciseness of the prompt itself. Shorter and more precise prompts were preferred, as they are more beneficial in clinical settings where clarity and efficiency are necessary. Additionally, the generated images were subjected to the same quantitative evaluation metrics as in sub-task 1, IS and FID, to ensure consistency in assessing the quality of images across different tasks. This dual approach, which combined qualitative assessment of prompt effectiveness with quantitative image quality metrics, provided a comprehensive assessment of participants' proficiency in generating both relevant prompts and high-quality synthetic images.

4. Participation and Results

This section provides an overview of participation in the challenge and discusses the results submitted by those who completed it.

4.1. Participation

In total, 22 teams signed up for the task, 2 teams submitted runs, and 2 teams submitted working notes papers [33, 34]. Table 1 shows an overview of the participants and the number of submissions to each sub-task alongside the number of participants from last year's challenge. This year experienced a noticeable decline in participation compared to last year's challenge. One reason for this may be the

complexity of the task and the hardware and model requirements. Furthermore, future editions could benefit from enhanced outreach and support mechanisms, such as tutorials or "getting started" scripts, to broaden participant engagement and lower the entry barriers.

4.2. Results

A total of six runs were submitted to sub-task 1, and no runs were submitted to sub-task 2. This section gives an overview of the results of each run and briefly discusses the approach submitted by each participant. The results can be seen in Table 2.

4.2.1. MMCP Team

Team MMCP's approach was based on two methods: fine-tuning Kandinsky models and implementing a Medical Synthesis with Diffusion Model (MSDM). They fine-tune pre-trained Kandinsky models to generate images from text prompts. In addition, they experimented with MSDM, showing improved results over Kandinsky-based models. Example images of each sumbission can be seen in Figure 2. For more information on their approach, please read their working notes paper [33].
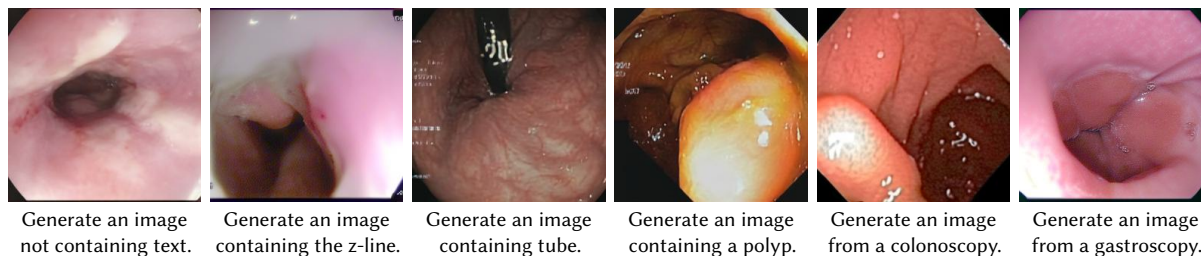

| Generate an image not containing text. | Generate an image containing the z-line. | Generate an image containing tube. | Generate an image containing a polyp. | Generate an image from a colonoscopy. | Generate an image from a gastroscopy. |

Figure 2: Team 2 submission examples. Please note that these images have been cherry picked, please see the participant paper for more details [33].

4.2.2. Team 2

Team 2 used a different approach for each submission. The first approach did not generate synthetic images, rather it retrieved images that closely related to the input prompt. To do this, they used a Connecting text and images (CLIP) model. The second submission used a fine-tuned stable diffusion model that generated synthetic images. The third submission used a fine-tuned Low-Rank Adaptation of Large Language Models (LoRA) model to generate images. This method uses LoRA to modify pre-existing stable diffusion model to enable the production of high-quality images that closely align with the input specifications. Example images of each submission can be seen in Figure 3.

Table 1
An overview of the submissions to each task availalbe at MedVQA-GI.

	MedVQA 2023	MedVQA 2024	Difference
# Registrations	26	22	-4
# Teams that submitted	8	2	-6
# Submissions to Task 1	10	6	-4
# Submissions to Task 2	4	0	-4
# Submissions to Task 3	2	-	-
# Paper Submissions	6	2	-4

Table 2
Results for Task 1. Each submission is evaluated using the FID and the Inception Score (IS). The FID scores is calculated against the MedVQA testing datasert (Single), GastroVision (Multi), and a combination of the two (Both). The IS socre is calculated on a 10-way split of the synthetic images, where we display the mean (avg), standard deviation (sd), and median (med).

Team	Submission	FID (Single)	FID (Multi)	FID (Both)	IS (avg)	IS (std)	IS (med)
MMCP Team	submission1	0.125	0.121	0.119	1.773	0.023	1.775
	submission2	0.120	0.117	0.115	1.791	0.028	1.792
	submission3	0.086	0.064	0.066	1.624	0.031	1.633
team2	submission1*	0.114	0.128	0.124	1.568	0.025	1.560
	submission2	0.099	0.064	0.067	2.327	0.065	2.339
	submission3	0.110	0.073	0.076	2.362	0.050	2.359

Generate an image from a colonoscopy procedure.

Generate an image with 1 polyp.

Generate an image from a gastroscopy procedure.

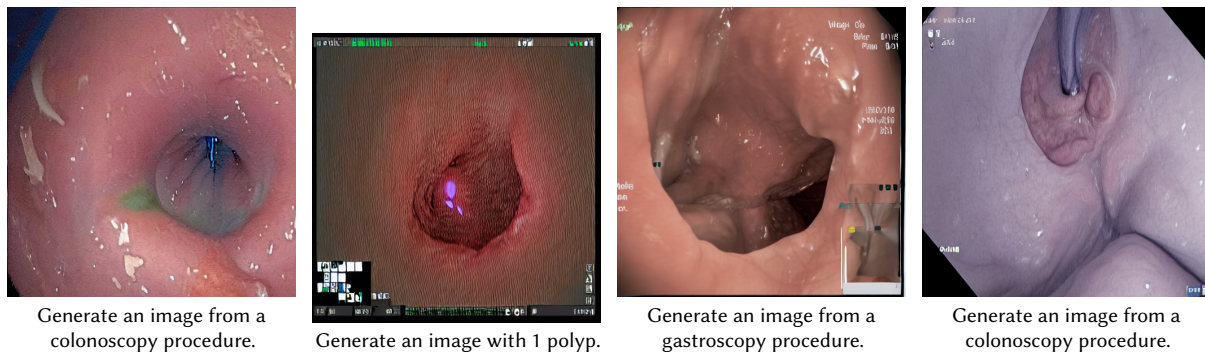Generate an image from a colonoscopy procedure.

Figure 3: Team 2 submission examples. Please note that these images have been cherry picked, please see the participant paper for more details [].

4.3. Discussion

The challenge results highlight several important insights and areas for further exploration. Firstly, the performance across the two teams and runs varied. This variability underscores the complexity of creating high-quality medical images. However, we found that the quality of the images did not always correspond to the scores provided by the quantitative metrics, suggesting that we need more robust synthetic image quality metrics specifically for medical images and their applications.

Another notable finding was that there was some confusion surrounding generation of synthetic images. One team submitted a run that retrieved "real" images that corresponded to the submitted prompt. This deviated from the intended goal, as the main point was to generate synthetic images. This highlights the need for clearer communication of the challenge requirements.

Furthermore, reduced participation compared to last year indicates possible entry barriers that may include the complexity of tasks or a lack of foundational resources for newcomers. Addressing these barriers could involve providing more comprehensive datasets, detailed examples of successful implementations, and potentially simplifying the challenge structure to attract a broader range of participants.

5. Conclusion and Future Outlook

This paper discussed the second edition of the MedVQA-GI challenge, which took place at ImageCLEF in 2024. The challenge consisted of two sub-tasks centered on the generation of synthetic images in the gastrointestinal tract. In the future, we plan on making a more robust task with more resources to get started. Furthermore, we also want to merge the tasks from the first year with this year's challenge to keep the task more consistent.

References

[1] B. Ionescu, H. Müller, A.-M. Drăgulinescu, J. Rückert, A. Ben Abacha, A. G. S. de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. Pakull, H. Damm, B. Bracke, C. M. Friedrich, A.-G. Andrei, Y. Prokopchuk, D. Karpenka, A. Radzhabov, V. Kovalev, C. Macaire, D. Schwab, B. Lecouteux, E. Esperança-Rodier, W.-W. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, M. Heinrich, J. Kiesel, M. Potthast, B. Stein, Overview of ImageCLEF 2024: Multimedia retrieval in medical, socialmedia and recommender systems applications, in: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France, 2024.

[2] S. A. Hicks, A. Storås, P. Halvorsen, T. de Lange, M. A. Riegler, V. Thambawita, Overview of imageclefmedical 2023 – medical visual question answering for gastrointestinal tract, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.

[3] C. Hassan, M. Spadaccini, A. Iannone, R. Maselli, M. Jovani, V. T. Chandrasekar, G. Antonelli, H. Yu, M. Areia, M. Dinis-Ribeiro, et al., Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis, Gastrointestinal endoscopy 93 (2021) 77–85.

[4] A. Alammari, A. R. Islam, J. Oh, W. Tavanapong, J. Wong, P. C. De Groen, Classification of ulcerative colitis severity in colonoscopy videos using cnn, in: Proceedings of the ACM International Conference on Information Management and Engineering (ACM ICIME), 2017, pp. 139–144. doi:https://doi.org/10.1145/3149572.3149613.

[5] D. Bychkov, N. Linder, R. Turkki, S. Nordling, P. E. Kovanen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, J. Lundin, Deep learning based tissue analysis predicts outcome in colorectal cancer, Scientific Reports 8 (2018) 3395. URL: http://dx.doi.org/10.1038/s41598-018-21758-3. doi:https://doi.org/10.1038/s41598-018-21758-3.

[6] Y. Mori, S.-e. Kudo, M. Misawa, Y. Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, F. Urushibara, S. Kataoka, Y. Ogawa, Y. Maeda, K. Takeda, H. Nakamura, K. Ichimasa, T. Kudo, T. Hayashi, K. Wakamura, F. Ishida, H. Inoue, H. Itoh, M. Oda, K. Mori, Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study, Annals of Internal Medicine 169 (2018) 357–366. doi:https://doi.org/10.7326/M18-0249.

[7] K. Pogorelov, S. L. Eskeland, T. de Lange, C. Griwodz, K. R. Randel, H. K. Stensland, D.-T. Dang-Nguyen, C. Spampinato, D. Johansen, M. Riegler, P. Halvorsen, A holistic multimedia system for gastrointestinal tract disease detection, in: Proceedings of the ACM on Multimedia Systems Conference (MMSYS), 2017, pp. 112–123. doi:https://doi.org/10.1145/3193740.

[8] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, International Journal of Computer Assisted Radiology and Surgery 9 (2014) 283–293. doi:https://doi.org/10.1007/s11548-013-0926-3.

[9] V. L. Thambawita, D. Jha, H. L. Hammer, H. D. Johansen, D. Johansen, P. Halvorsen, M. Riegler, An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification, ACM Transactions on Computing for Healthcare (2020).

[10] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, H. Johansen, Doubleu-net: A deep convolutional neural network for medical image segmentation, in: Proceeding of the International Symposium on Computer Based Medical Systems (CBMS), 2020.

[11] Q. Angermann, J. Bernal, C. Sánchez-Montes, M. Hammami, G. Fernández-Esparrach, X. Dray, O. Romain, F. J. Sánchez, A. Histace, Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis, in: Proceedings of Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures (CARE CLIP), volume 10550, Springer, 2017, pp. 29–41.

[12] K. Pogorelov, M. Riegler, P. Halvorsen, P. T. Schmidt, C. Griwodz, D. Johansen, S. L. Eskeland, T. de Lange, Gpu-accelerated real-time gastrointestinal diseases detection, in: Proceedings of the International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2016, pp. 185–190.

doi:https://doi.org/10.1109/CBMS.2016.63.

[13] M. Riegler, K. Pogorelov, P. Halvorsen, T. de Lange, C. Griwodz, P. T. Schmidt, S. L. Eskeland, D. Johansen, EIR - efficient computer aided diagnosis framework for gastrointestinal endoscopies, in: Proceedings of the IEEE International Workshop on Content-Based Multimedia Indexing (CBMI), 2016, pp. 1–6. doi:https://doi.org/10.1109/CBMI.2016.7500257.

[14] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, P. C. De Groen, Polyp-alert: Near real-time feedback during colonoscopy, Computer Methods and Programs in Biomedicine 120 (2015) 164–179. doi:https://doi.org/10.1016/j.cmpb.2015.04.002.

[15] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H. D. Johansen, Resunet++: An advanced architecture for medical image segmentation, in: Proceedings of the International Symposium on Multimedia (ISM), 2019, pp. 225–230. doi:https://doi.org/10.1109/ISM46123.2019.00049.

[16] J. Bernal, A. Histace, M. Masana, Q. Angermann, C. Sánchez-Montes, C. Rodriguez, M. Hammami, A. Garcia-Rodriguez, H. Córdova, O. Romain, G. Fernández-Esparrach, X. Dray, J. Sanchez, Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully configurable tool for easy and fast annotation of image databases, in: Proceedings of Computer Assisted Radiology and Surgery (CARS), 2018. doi:https://hal.archives-ouvertes.fr/hal-01846141.

[17] Y. Guo, J. Bernal, B. J Matuszewski, Polyp segmentation with fully convolutional deep neural networks—extended evaluation study, Journal of Imaging 6 (2020) 69.

[18] M. Min, S. Su, W. He, Y. Bi, Z. Ma, Y. Liu, Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology, Scientific reports 9 (2019) 2881. doi:https://doi.org/10.1038/s41598-019-39416-7.

[19] N. M. Ghatwary, X. Ye, M. Zolgharni, Esophageal abnormality detection using densenet based faster r-cnn with gabor features, IEEE Access 7 (2019) 84374–84385. doi:https://doi.org/10.1109/ACCESS.2019.2925585.

[20] S. Shah, N. Park, N. E. H. Chehade, A. Chahine, M. Monachese, A. Tiritilli, Z. Moosvi, R. Ortizo, J. Samarasena, Effect of computer-aided colonoscopy on adenoma miss rates and polyp detection: a systematic review and meta-analysis, Journal of Gastroenterology and Hepatology 38 (2023) 162–176.

[21] S. Hicks, M. Riegler, P. Smedsrud, T. B. Haugen, K. R. Randel, K. Pogorelov, H. K. Stensland, D.-T. Dang-Nguyen, M. Lux, A. Petlund, T. de Lange, P. T. Schmidt, P. Halvorsen, Acm multimedia biomedia 2019 grand challenge overview, in: Proceedings of the ACM International Conference on Multimedia (ACM MM), 2019, pp. 2563–2567. doi:https://doi.org/10.1145/3343031.3356058.

[22] K. Pogorelov, M. Riegler, P. Halvorsen, S. A. Hicks, K. R. Randel, D.-T. Dang-Nguyen, M. Lux, O. Ostroukhova, T. De Lange, Medico multimedia task at mediaeval 2018, in: Proceeding of the MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop (MediaEval), 2018.

[23] M. Riegler, K. Pogorelov, P. Halvorsen, K. Randel, S. Eskeland, D.-T. Dang-Nguyen, M. Lux, C. Griwodz, C. Spampinato, T. de Lange, Multimedia for medicine: the medico task at mediaeval 2017, in: Proceeding of the MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop (MediaEval), 2017.

[24] J. Bernal, H. Aymeric, Miccai endoscopic vision challenge polyp detection and segmentation, https://endovissub2017-giana.grand-challenge.org/home/, 2017. Accessed: 2017-12-11.

[25] S. Hicks, M. Riegler, P. Smedsrud, T. B. Haugen, K. R. Randel, K. Pogorelov, H. K. Stensland, D.-T. Dang-Nguyen, M. Lux, A. Petlund, T. de Lange, P. T. Schmidt, P. Halvorsen, Acm multimedia biomedia 2019 grand challenge overview, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2563–2567. URL: https://doi.org/10.1145/3343031.3356058. doi:10.1145/3343031.3356058.

[26] V. Thambawita, P. Salehi, S. A. Sheshkal, S. A. Hicks, H. L. Hammer, S. Parasa, T. d. Lange, P. Halvorsen, M. A. Riegler, Singan-seg: Synthetic training data generation for medical image segmentation, PLOS ONE 17 (2022) 1–24. URL: https://doi.org/10.1371/journal.pone.0267976. doi:10.1371/journal.pone.0267976.

[27] D. Yoon, H.-J. Kong, B. S. Kim, W. S. Cho, J. C. Lee, M. Cho, M. H. Lim, S. Y. Yang, S. H. Lim, J. Lee, J. H. Song, G. E. Chung, J. M. Choi, H. Y. Kang, J. H. Bae, S. Kim, Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network, Sci Rep 12 (2022) 261.

[28] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen, et al., Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy, Scientific data 7 (2020). doi:10.1038/s41597-020-00622-y.

[29] D. Jha, S. Ali, K. Emanuelsen, S. A. Hicks, V. Thambawita, E. Garcia-Ceja, M. A. Riegler, T. de Lange, P. T. Schmidt, H. D. Johansen, D. Johansen, P. Halvorsen, Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy, in: Proceeedings of the International COnference on MultiMedia Modeling (MMM), 2021, pp. 218–229. doi:10.1007/978-3-030-67835-7_19.

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, Advances in neural information processing systems 29 (2016).

[31] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, Advances in neural information processing systems 30 (2017).

[32] D. Jha, V. Sharma, N. Dasu, N. K. Tomar, S. Hicks, M. Bhuyan, P. K. Das, M. A. Riegler, P. Halvorsen, T. de Lange, U. Bagci, Gastrovision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection, in: ICML Workshop on Machine Learning for Multimodal Healthcare Data (ML4MHD 2023), 2023.

[33] M. Chaychuk, Mmcp team at imageclefmed 2024 task on image synthesis: Diffusion models for text-to-image generation of colonoscopy images, in: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France, 2024.

[34] E.-P. Oluwafemi Ojonugwa, M. Rahman, F. Khalifa, Advancing ai-powered medical image synthesis: Insights from medvqa-gi challenge using clip, fine-tuned stable diffusion, and dream-booth + lora, in: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France, 2024.