The Need for Medically Aware Video Compression in Gastroenterology

Joel Shor*
Verily Life Sciences
joelshor@verily.com

Nick Johnston*
Google Research
nickj@google.com

Abstract

Compression is essential to storing and transmitting medical videos, but the effect of compression on downstream medical tasks is often ignored. Furthermore, systems in practice rely on standard video codecs, which naively allocate bits between medically relevant frames or parts of frames. In this work, we present an empirical study of some deficiencies of classical codecs on gastroenterology videos, and motivate our ongoing work to train a learned compression model for colonoscopy videos. We show that **two of the most common classical codecs, H264 and HEVC, compress medically relevant frames statistically significantly worse than medically nonrelevant ones**, and that polyp detector performance degrades rapidly as compression increases. We explain how a learned compressor could allocate bits to important regions and allow detection performance to degrade more gracefully. Many of our proposed techniques generalize to medical video domains beyond gastroenterology.

1 Introduction

Colorectal cancer is the third most common cancer diagnosed in the US (1) and worldwide (2). Colonoscopies, a video-based diagnostic procedure, are one of the most common screening tools. Over 15 million colonoscopies are performed in the US each year (3), leading to an enormous amount of video transmission and storage for tasks like medical records, physician education, report generation, and training medical models. Data-driven machine learning models that perform polyp detection (4) and coverage detection (5) need these videos to train and evaluate.

H264 (6) and High Efficiency Video Coding (HEVC) (7) are two of the most common classical (non-machine learning based) video codecs. The quality versus size tradeoff is controlled using the "Quantization Parameter" (QP), which ranges from 0 (lossless) to 51 (most compressed). QP and other parameters can be tuned to match target quality requirements, but the transform-based compression algorithms are unable to take advantage of video domain specific properties, such as the texture and camera motion that are specific to colonoscopies.

Previous attempts to achieve diagnostically-lossless compression, or to reduce the degradation of medically relevant regions of interest, include modifying classical codecs to take advantage of medical properties (8; 9), leveraging superresolution (10), and applying different quality compression algorithms on specifically identified regions of interest (11). Researchers have also explored tuning classical codecs to specific medical domains, such as ultrasound (12). However, fully data-driven, modern image and video compression algorithms (13; 14) have yet to be applied to the medical domain.

*Authors contributed equally.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

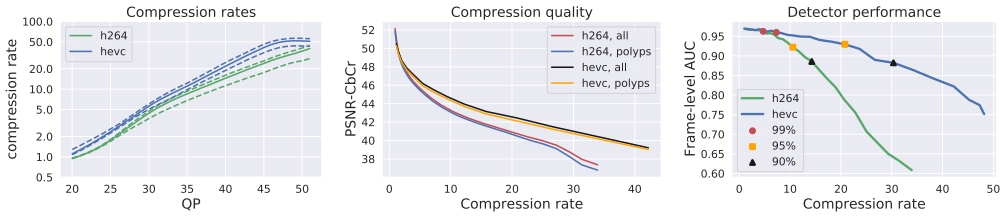

Figure 1: Left) 25th, 50th, and 75th percentile compression rates across 80 colonoscopy videos for difference QP values of H264 and HEVC. **Middle)** Frame compression frame quality, as measured by PSNR-CbCr. See Appendix for other metrics. **Right)** Polyp detector performance as a function of compression factor.

2 Experiments

Data: As in (4), our colonoscopy video dataset was collected from screening colonoscopy procedures performed in three hospitals. We used a detector trained on a 16K video subset of the dataset, and we computed our classical codec analyses on a different 80 video subset (2.2M frames, 15K polyp frames). All videos and metadata were deidentified according to the Health Insurance Portability and Accountability Act Safe Harbor. Ground truth polyp labeling was provided by the gastroenterologist annotators described in (4). The annotators were paid on an hourly basis, and pay was not based on the results they provided. The videos were compressed using H264 QP20 when transmitted from the hospitals. To justify our analyzing already-compressed videos, we used a small number of lossless videos to investigate the impact of re-compressing compressed videos to QPN (so called “generation loss"). We found that the impact was negligible (see Appendix for details).

Compression metrics: We evaluate frame quality using two standard image metrics: PSNR-CbCr and PSNR-Y. PSNR is the standard "Peak Signal to Noise Ratio" derived from the mean squared error between pixels in the original frame and the compressed frame. The value of the pixels depends on the type of PSNR computation:“CbCr" corresponds to PSNR between chroma of the frames, and “Y" corresponds to PSNR between luminance of the frames.

Polyp detector metrics: We evaluate polyp detector performance using the same methodology as (4). We report the AUC value of sensitivity versus false positive rate for various thresholds of the detector. The threshold determines how "confident" the polyp detector must be to register a detection, so the AUC curve captures the detector’s performance across a range of sensitivity scores.

Polyp detector model: The polyp detector is a production-grade polyp detector, as described in (4). The architecture is RetinaNet (15) with training and hyperparameters described in (4). The polyp detector demonstrates state-of-the-art performance on colonoscopy procedure videos as well as diagnostically challenging polyps.

3 Results

Compression rates: Figure 1 (left) shows the distribution of compression rates on colonoscopy videos for H264 and HEVC by QP value. At QP 51 (the highest compression rate for both classical codecs), HEVC achieved significantly more compression: the 25th, 50th, and 75th percentile compression rates were (43.6, 51.8, 56.5) for HEVC and (28.3, 39.8, 44.5) for H264.

Compression quality: Figure 1 (middle) shows the compression rate versus frame quality distribution for H264 and HEVC. Importantly, we see that **H264 and HEVC compress the most medically relvant frames statistically significantly worse**: treating each QP value separately, a two-sided Kolmogorov-Smirnov test between distribution of PSNR-CbCr shows that the frame quality is lower for polyp frames than for all frames. For each QP value, $N_1 = 2189948$, $N_2 = 15457$, H264 (HEVC) maximum p-value over all tests is $1.4 * 10^{-118}$ ($1.4 * 10^{-30}$), mean test statistic 0.13 (0.11). For the same test with PSNR-Y, see the Appendix. Figure 2 top two rows show the lowest quality compressed frames inside the body according to PSNR-CbCr, with and without polyps (for the absolute worst quality compressed frames, see the Appendix).

Detector performance: Figure 1 (right) shows the polyp detector performance as a function of compression rate. Videos can be compressed by factors of 4.7x and 7.3x before dropping below 99%

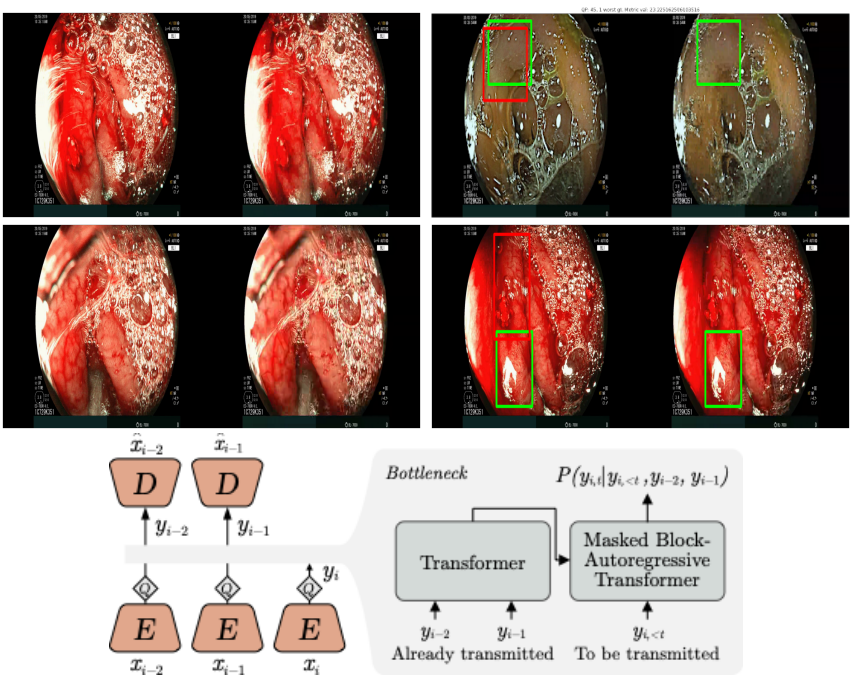

Figure 2: Qualities measured by PSNR-CbCr. Concatenated images are original frame on the left, compressed on the right. **Left upper (lower))** Worst compressed frames by H264 (HEVC) QP40. **Right upper (lower))** Worst compressed frames by H264 (HEVC) QP40 that have a polyp. **3rd row)** A high level description of learned neural video compression (diagram from (16)).

the base performance for H264 and HEVC respectively, 10.6x and 20.8x for 95%, and 14.3x and 30.4x for 90%. In addition to getting better frame quality and higher compression rates, **the detector performs 0.057 AUC better and a 29% relative improvement[2] on HEVC videos compared to H264, as the same compression rate**. The same holds in the "practical" regime of compression rates that achieve at least 95% the AUC of the original model: AUC is on average 0.26 AUC improved with a 23% relative improvement in AUC.

4 Future work

Ongoing work involves addressing the deficiencies of classical codecs in preserving medically relevant information. Our primary proposed solution is to leverage data-driven or "learned" compression using the recent Video Compression Transformer model (16). This model shows especially useful properties, such as the ability to capture domain-specific synthetic camera motion and domain-specific texture characteristics. We also plan to explore the complex relationship between training on videos compressed using one algorithm (e.g. H264 QP20), but run on a different type of compression during inference (e.g. lossless).

There are a number of ways to inject colonoscopy-specific information into a learned compressor. First, we can simply train the compression model on colonoscopy data, and it will learn to recreate videos with colonoscopy texture and video motion. Second, we can use colonoscopy data as a validation set to pick model hyperparameters (see (17) for an example of the potential impact of just using data to select model hyperparameters). Third, we can explore oversampling from polyp frames during training. Fourth, we can add a weight factor to the training loss on polyp frames during training. Fifth, we can add a weight factor on the training loss for polyp subregions of frames. Sixth, we can add polyp detectors directly to the compressor loss function (as well as other differentiable medical models, such as polyp type classification). Seventh, we can co-train the compression model with the detection model to maximize detector performance on the compressed videos.

[2]relative improvement defined as $(AUC_{hevc} - AUC_{h264})/(1.0 - AUC_{h264})$

Acknowledgments and Disclosure of Funding

5 Acknowledgements

This work was funded by Verily LLC. We'd like to thank Roman Goldberg and Ehud Rivlen for their technical guidance. We'd like to thank Joe Shao, Stephen Lanham, Bimba Rao, Josh Widen, Bryce Evans, and Brijesh Patel for their suggestions and feedback.

References

[1] "Key statistics for colorectal cancer," https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html, accessed: 2022-09-04.

[2] "Colorectal cancer statistics," https://www.wcrf.org/cancer-trends/colorectal-cancer-statistics, accessed: 2022-09-04.

[3] "Let's get screened, part 1: Colon cancer screening," https://www.pennmedicine.org/cancer/about/focus-on-cancer/2019/december/lets-get-screened-colonoscopy, accessed: 2022-09-04.

[4] D. M. Livovsky, D. Veikherman, T. Golany, A. Aides, V. Dashinsky, N. Rabani, D. Ben Shimol, Y. Blau, L. Katzir, I. Shimshoni, Y. Liu, O. Segol, E. Goldin, G. Corrado, J. Lachter, Y. Matias, E. Rivlin, and D. Freedman, "Detection of elusive polyps using a large-scale artificial intelligence system (with videos)," *Gastrointest Endosc*, vol. 94, no. 6, pp. 1099–1109, Dec 2021.

[5] Y. Blau, D. Freedman, V. Dashinsky, R. Goldenberg, and E. Rivlin, "Unsupervised 3d shape coverage estimation with applications to colonoscopy," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 3364–3374.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.

[7] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[8] Y.-J. Chang, P.-H. Tsai, and C.-L. Lin, "Novel medical video compression methods over lossless hevc coder," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 1687–1691.

[9] H. Yu, Z. Lin, and F. Pan, "Applications and improvement of h.264 in medical video compression," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, pp. 2707 – 2716, 01 2006.

[10] D. Bonanno and C. J. Debono, "A medical video coding scheme with preserved diagnostic quality," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[11] Z. Zuo, X. Lan, L. Deng, S. Yao, and X. Wang, "An improved medical image compression technique with lossless region of interest," *Optik - International Journal for Light and Electron Optics*, vol. 126, 07 2015.

[12] M. Razaak and M. G. Martini, "Rate-distortion and rate-quality performance analysis of hevc compression of medical ultrasound videos," *Procedia Computer Science*, vol. 40, pp. 230–236, 2014, fourth International Conference on Selected Topics in Mobile Wireless Networking (MoWNet'2014). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050914016032

[13] E. Agustsson, D. Minnen, N. Johnston, J. Ballé, S. J. Hwang, and G. Toderici, "Scale-space flow for end-to-end optimized video compression," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8500–8509.

[14] G. Toderici, D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor, and M. Covell, "Full resolution image compression with recurrent neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 5306–5314.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017. [Online]. Available: https://arxiv.org/abs/1708.02002

[16] F. Mentzer, G. Toderici, D. Minnen, S.-J. Hwang, S. Caelles, M. Lucic, and E. Agustsson, "Vct: A video compression transformer," 2022. [Online]. Available: https://arxiv.org/abs/2206.07307

[17] J. Shor, A. Jansen, W. Han, D. Park, and Y. Zhang, "Universal paralinguistic speech representations using self-supervised conformers," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3169–3173.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [N/A] Code not released.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Experiments and Results section for justification.
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] No theoretical results.
   (b) Did you include complete proofs of all theoretical results? [N/A] No proofs.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A] No new models trained.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We show error bars in all of our analyses.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] No new models trained.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We cite the polyp detection paper for their models.
   (b) Did you mention the license of the assets? [Yes] Data is not publically available.
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A] No new assets.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] All patient data was provided after written consent.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] Data has been anonymized according to HIPAA.
5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No] Participants were paid an hourly wage with no contingency on performance. The wage information for these participants is not publicly available.

6 Potential negative societal impact

Since this research deals with medical data and has clinical applications, the usual potential negative societal impacts are applicable. Incorrect model training or deidentification can lead to accidentally leaked and/or memorized patient information. A model with biased performance can lead to negative clinical outcomes for protected subgroups. Finally, good performance on a test set doesn't mean the model is necessarily ready for immediate clinical application, and having a performant model doesn't necessarily mean it can replace a medical professional.

A Appendix

A.1 Generation loss

Since our dataset was compressed to be transferred from hospitals, we first investigated the impact of multiple stages of compression ("generation loss"). We used lossless video (24 seconds, 1.4GB) collected from an endoscope viewing dyed, non-human tissue. This gave the video the motion and texture characteristics of a colonoscopy. We then compared video quality between two compression schemes: compressing using H264 QP N, where $20 <= N <= 51$, and compressing to H264 QP20, then to H264 QPN (when referring explicitly to this comparison, we will concisely refer to the latter as 'QP'). We then compared the average frame quality between these two schemes using two quality metrics (see 'Metrics' section). The quality differences between these two compression schemes was minor in terms of PSNR-CbCr, which justifies our working with video data already compressed by H264 QP20.

A.2 Compression quality

See Figure 4 for compression quality as a function of QP value instead of compression rate, as well as quality measured by PSNR-Y. Furthermore, the Kolmogorov-Smirnov tests on PSNR-Y show the same behavior: treating each QP value separately, a two-sided between distributions shows that the frame quality is lower for polyp frames than for all frames (for each QP value, $N_1 = 2189948$, $N_2 = 15457$, H264 PSNR-Y max p-value is $9.5 * 10^{-77}$, mean test statistic 0.10, HEVC PSNR-Y maximum p-value over all tests is $2.4 * 10^{-13}$, mean test statistic 0.13). This result holds for each QP value between 20 and 51.

A.3 Lowest quality compressed frames

Figure 2 shows the worst compressed colonoscopy frames inside the body, and with polyps. Interestingly, the worst quality compressed frames were actually outside the body. We include them here for completeness.

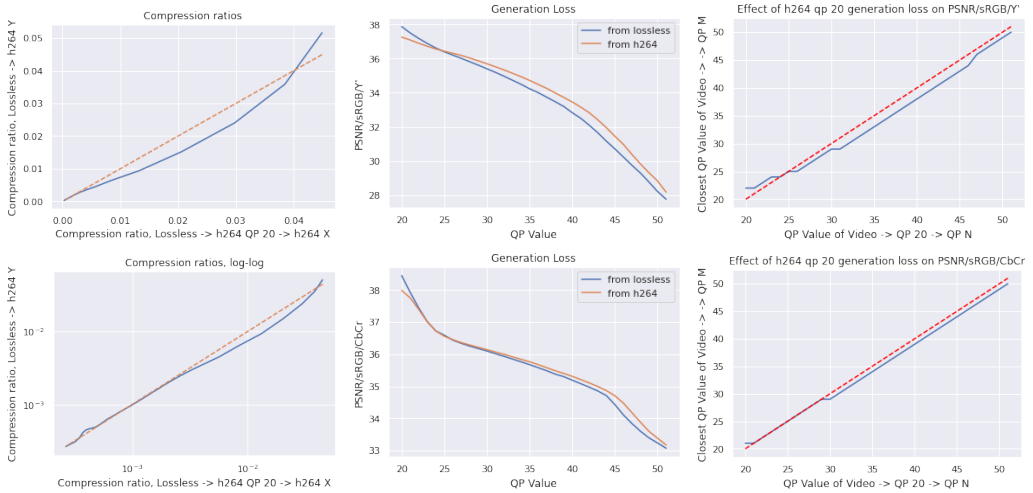

Figure 3: Left upper) Comparison of compression rates for two compression methods using the same QP value N. y-axis is the compression rate of compressing from lossless directly to H264 QPN. x-axis is the compression rate of compressing to H264 QPN through H264 QP20. Dotted line is the $x = y$. Note that for lower compression rates (lower QP value), compressing through H264 QP20 gives a smaller file sizes. **Left lower)** Same as (left upper), but a log-log plot. **Middle upper)** Frame quality metric PSNR-Y, averaged across compressed frames, for the two methods of compression. **Middle upper)** Same as (middle upper) but for PSNR-CbCr. **Right upper)** Correspondence between QP values of the two compression methods, according to closest PSNR-Y values. **Right lower)** Same as Right upper, but for PSNR-CbCr.

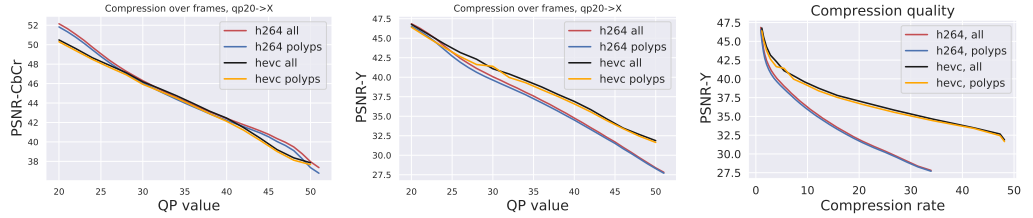

Figure 4: Quality vs QP value and quality vs compression for different quality metrics.

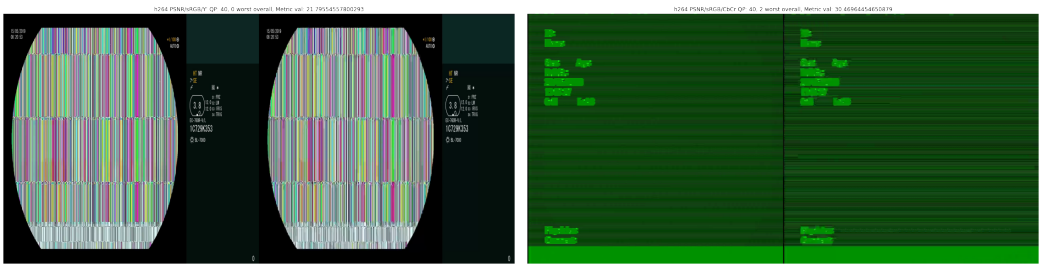

Figure 5: Overall frames with the lowest quality compression for H264 QP40.