Ordinal Priors for Colonoscopy Temporal Segmentation

Seunghyun Jang[1,2*], Dongheon Lee[3,4*], Yoonsung Jung[1], Eue Young Kim[5], Na Mi Shin[5], Byeong Gwan Kim[5], Jong Pil Im[5], Sihyun Kim[5], and Chang Min Park[3†]

[1] Interdisciplinary Program in Bioengineering, Seoul National University Graduate School, Seoul, South Korea
[2] Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, South Korea
[3] Department of Radiology, Seoul National University Hospital, Seoul, South Korea
[4] Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, South Korea
[5] Department of Internal Medicine and Liver Research Institute, Seoul National University Hospital, Seoul, South Korea
morphius@snu.ac.kr

Abstract. Temporal segmentation of colonoscopy withdrawal videos into anatomical regions is a key component for automated post-procedure reporting, enabling coverage assessment and structured documentation. While recent temporal segmentation models show promising accuracy, they typically treat colon segments as unordered categories, ignoring the inherent anatomical ordinality of the colon. This often yields temporally inconsistent predictions with frequent flickering and anatomically implausible jumps between non-adjacent regions. We propose an ordinal-aware framework that injects anatomical priors into both training and inference. During training, we add an auxiliary ordinal regression objective that penalizes errors in proportion to anatomical distance. At inference, we apply an ordinal-constrained decoder that enforces anatomically plausible transitions. On the REAL-Colon dataset, our method achieves 0.7621 weighted F1 and 0.6286 weighted Jaccard while substantially reducing label transitions to 15.84 per video. On the CAS-Colon dataset, it attains 0.7300 weighted F1 and 0.5935 weighted Jaccard without target-domain fine-tuning, indicating improved cross-dataset generalization. Overall, incorporating anatomical ordinality improves temporal consistency and supports clinically applicable, anatomically grounded segmentation. Code is available at https://github.com/jangbi1/OColT.

Keywords: Colonoscopy Temporal Segmentation · Colonoscopy Video Analysis · Ordinal Regression · Temporal Consistency.

---

* Equal contributions.
† Corresponding author.

1 Introduction

Colonoscopy remains the gold standard for colorectal cancer screening, and the accurate identification of anatomical colon segments during the procedure is essential for multiple clinical tasks [22]. Anatomical localization underpins automated post-procedure reporting, supporting the documentation of polyp location [17], the assessment of cecal and ileal intubation rates [6], and the computation of bowel preparation quality scores [15, 19]. Furthermore, accurate segment identification also informs treatment and surveillance decisions, as in ulcerative colitis where disease extent is classified by how far inflammation spreads proximally from the rectum [14, 21]. Despite its clinical importance, automated anatomical localization remains challenging due to the limited field of view, absence of reliable landmarks, and inter-patient variability in colon morphology [4, 11].

Deep learning for surgical workflow analysis[9], including surgical phase recognition [7] and action segmentation [26, 27], has made substantial progress, and similar efforts are beginning to emerge in colonoscopy video analysis [8, 28]. Early approaches applied frame-level classification to identify anatomical regions from individual images [20], while subsequent work explored clustering-based methods to segment colonoscopy videos into anatomical sections [13]. More recently, transformer-based models have been introduced to leverage temporal context for anatomical segments in endoscopy videos [18], and the release of large-scale annotated colonoscopy datasets [1] has enabled TCN-based frameworks for full-procedure video temporal segmentation [2].

However, all existing approaches treat anatomical colon segments as an unordered set of independent categories. This formulation ignores the inherent anatomical ordinality of the colon, the fact that segments follow a fixed proximal-to-distal sequence, leading to temporally inconsistent predictions with transitions between non-adjacent regions that are anatomically implausible [5].

In this work, we propose an ordinal-aware framework that incorporates anatomical ordinal priors into both training and inference via distance-weighted ordinal regression losses and a transition-constrained decoder. We evaluate our framework on the REAL-Colon dataset [1] and assess cross-dataset generalization on the CAS-Colon dataset [23] without any target-domain fine-tuning. The main contributions of this work are as follows:

1. We propose distance-weighted ordinal regression losses and a transition-constrained decoder that jointly enforce anatomical ordinality during training and inference.
2. Experiments on the REAL-Colon benchmark demonstrate improved performance and temporal consistency, with cross-dataset generalization validated on CAS-Colon without target-domain fine-tuning.

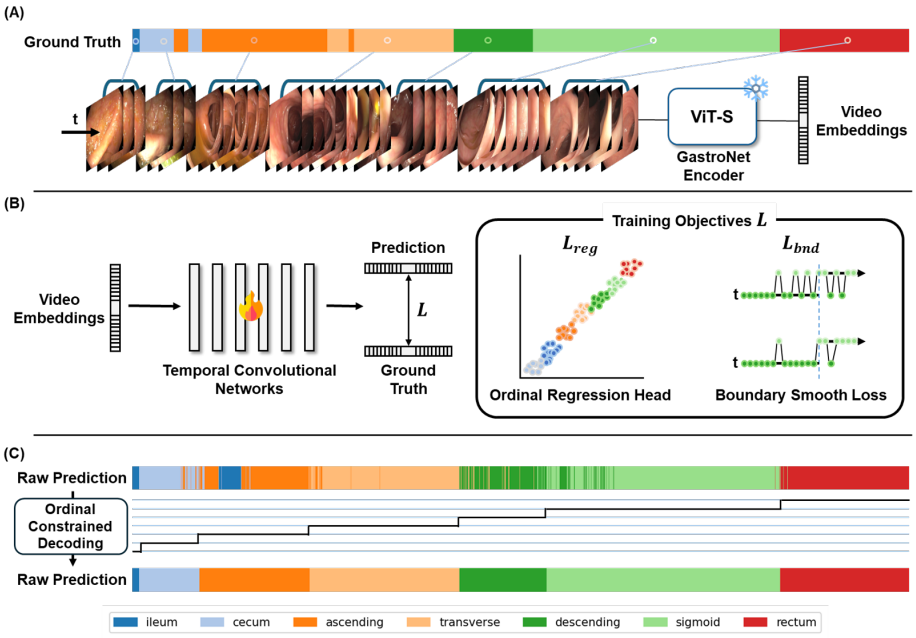

Fig. 1. Overview of the proposed ordinal-aware framework. (A) Colonoscopy frames are encoded using a GastroNet backbone [3] into a sequence of frame-level embeddings. (B) A temporal segmentation model is trained with categorical, boundary smooth loss, and ordinal regression loss. (C) Raw predictions are refined through an ordinal-constrained decoding to produce the final result.

2 Methods

2.1 System Overview

We address the temporal segmentation of colonoscopy withdrawal videos, where each frame is assigned to one of seven anatomical segments traversed during withdrawal procedure: ileum, cecum, ascending, transverse, descending, sigmoid, and rectum. A pretrained vision encoder extracts per-frame embeddings $\mathbf{X} \in R^{T\times D}$, and a temporal model f_θ produces frame-wise predictions $\hat{y}_t \in \{0, \ldots, 6\}$.

A distinctive property of this task is that the label space carries a strict anatomical ordering. During withdrawal, the endoscope retraces a fixed proximal-to-distal path, yielding a near-monotonic sequence ileum(0) $\rightarrow \cdots \rightarrow$ rectum(6) where backward transitions are rare and segment skipping is anatomically implausible. Standard temporal segmentation methods treat all classes as unordered, penalizing a misclassification from cecum to ascending identically to one from cecum to rectum, and impose no constraint on the direction of label transitions. We build upon ColonTCN [2], which processes the embeddings through stacked dilated temporal convolutions. ColonTCN establishes a strong baseline but does not exploit the ordinal structure of the label space. In the

following sections, we introduce components at both training and inference time that explicitly encode this anatomical ordering. Fig. 1 summarizes the overall pipeline, from GastroNet-based frame embeddings to ordinal-aware training objectives and the ordinal-constrained decoding stage.

2.2 Dataset

We evaluate on two colonoscopy video datasets. The **REAL-Colon** dataset [1] comprises 60 full-procedure colonoscopy videos, with frame-level annotations for seven anatomical withdrawal-phase segments (ileum, cecum, ascending, transverse, descending, sigmoid, and rectum). Following ColonTCN, videos are subsampled to 5 fps, and frames with insertion phases are excluded. Frames annotated as uncertain ($<$0.2% of all frames) are excluded from loss computation. Frame embeddings are extracted using a ViT backbone with GastroNet-5M DINOv1 weights [3], yielding 384-dimensional embeddings. For training, each video is augmented with a single, video-consistent spatial transform comprising random horizontal flip, vertical flip, and HSV color jitter.

For cross-dataset generalization, we evaluate on **CAS-Colon** [23], which provides timestamp-based annotations across ten colon regions. Hepatic flexure and splenic flexure are merged into transverse colon, and anal canal into rectum, yielding the same seven-segment label space. We randomly sample 20 videos as the test split, with no target-domain fine-tuning performed.

2.3 Experimental Setup

Baseline Models We evaluate against ColonTCN [2], ASFormer [26], and TeCNO [7], representing task-specific TCN, transformer-based, and multi-stage convolutional approaches to temporal segmentation. To ensure fairness, we retrain all baselines using the same split, embeddings, and evaluation protocol.

Implementation Details Models are trained on the REAL-Colon dataset split into 44 training, 4 validation, and 12 test videos using AdamW (lr = 5e-4) for 100 epochs with early stopping on validation weighted Jaccard. All baseline models are retrained under the same data pipeline and evaluation protocol to ensure a fair comparison. Experiments are conducted on a single NVIDIA A6000 GPU (Python 3.12, CUDA 12.2). Loss weights: λ_{CE}=0.5, λ_{TMSE}=0.15 (τ=4), λ_{bnd}=0.5 (r=5), λ_{reg}=0.3. OCD: α=10, β=5, free start, w=3.

Metrics We report weighted F1 (wF1) and weighted Jaccard (wJacc) averaged by class support, alongside weighted mean absolute percentage error (WMAPE) measuring proportional deviation between predicted and ground-truth segment durations. To quantify prediction stability, we additionally report the average of label transitions.

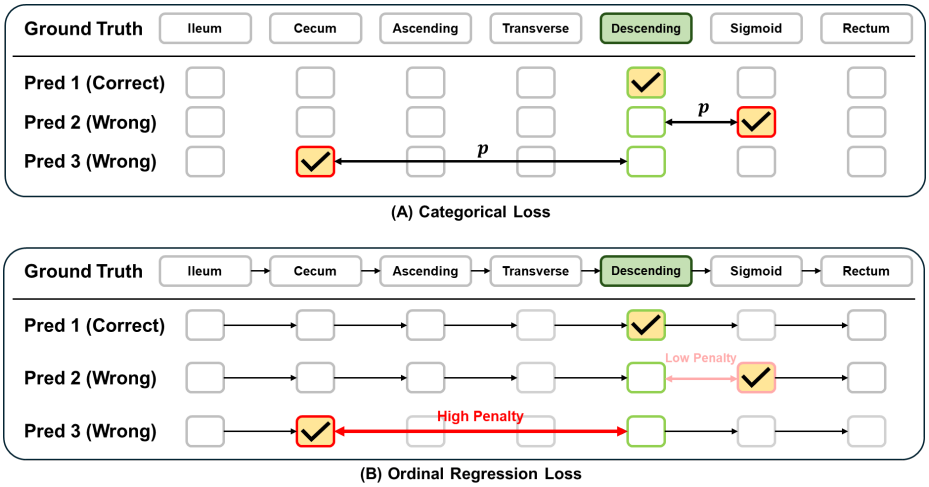

Fig. 2. Illustration of how anatomical ordinality is incorporated into the loss function. (A) Categorical loss penalizes all incorrect predictions equally regardless of anatomical proximity. (B) The proposed ordinal regression loss penalizes predictions in proportion to their anatomical distance from the ground truth segment, assigning a small penalty to adjacent errors (Pred 2) and a large penalty to distant errors (Pred 3).

2.4 Ordinal-Aware Objectives

We augment the standard training objective with two auxiliary losses that encode the ordinal structure of the label space and promote temporal stability of predictions. As illustrated in Fig. 2, our auxiliary ordinal regression head penalizes misclassifications in proportion to anatomical distance, unlike categorical loss which treats all errors equally.

Ordinal Regression Head (ORH) To further enforce ordinal awareness, we attach a lightweight auxiliary head h_ϕ that maps the temporal model's logits to a scalar anatomical position $[5, 25]$:

$$\hat{r}_t = \sigma\big(h_\phi(f_\theta(\mathbf{x}_t))\big) \cdot (C-1) \tag{1}$$

where σ is the sigmoid function, constraining $\hat{r}_t \in [0, C-1]$. The head is trained with mean absolute error against the integer class label: $\mathcal{L}_{\text{reg}} = \frac{1}{|\mathcal{V}|}\sum_{t\in\mathcal{V}} |\hat{r}_t - y_t|$, where $\mathcal{V}$ denotes valid (non-ignored) frames. This scalar formulation provides a graded supervision signal aligned with anatomy, directly reflecting the distance-sensitive behavior depicted in Fig. 2.

Boundary Smooth Loss Temporal mean squared error penalize differences between adjacent frames uniformly, which suppresses chattering but also blurs genuine segment boundaries [10]. We introduce a boundary-aware variant that

applies smoothing only within stable regions. Specifically, we identify GT boundary frames and dilate them by a radius r, then compute the symmetrized KL divergence only over stable frame pairs:

$$\mathcal{L}_{\text{bnd}} = \frac{1}{|\mathcal{S}|} \sum_{(t,t-1)\in\mathcal{S}} \frac{1}{2}\big[D_{\text{KL}}(p_t \| p_{t-1}) + D_{\text{KL}}(p_{t-1} \| p_t)\big] \tag{2}$$

where $\mathcal{S}$ is the set of consecutive frame pairs in which neither frame falls within radius r of a GT boundary. This preserves sharp transitions at true segment boundaries while suppressing mid-segment fluctuations. The total loss combines the above with the standard multi-stage cross-entropy:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{bnd}}\mathcal{L}_{\text{bnd}} \tag{3}$$

Ordinal-Constrained Decoding (OCD) While training encourages ordinal consistency, per-frame argmax predictions can still exhibit anatomically implausible transitions such as backward jumps or segment skipping. We address this at inference with a penalized Viterbi decoding that finds the globally optimal path under an ordinal transition prior [16, 24].

Given predicted probabilities $\mathbf{p}_t \in [0,1]^C$, we apply temporal average pooling with window size w to suppress transient fluctuations, then convert to log-probabilities $\boldsymbol{\ell}_t \in R^C$. A transition score matrix $\mathbf{A} \in R^{C\times C}$ encodes the anatomical plausibility of moving from state j at time $t-1$ to state i at time t:

$$A_{i,j} = \begin{cases} 0 & \text{if } i = j \text{ (stay)} \\ 0 & \text{if } i = j+1 \text{ (forward-1)} \\ -\alpha \cdot (j-i) & \text{if } i < j \text{ (backward)} \\ -\beta \cdot (i-j-1) & \text{if } i > j+1 \text{ (skip)} \end{cases} \tag{4}$$

where α and β are the backward and skip penalty coefficients. The optimal path is obtained via dynamic programming:

$$V_t(i) = \ell_t(i) + \max_j \big[V_{t-1}(j) + A_{i,j}\big] \tag{5}$$

with backtracking to recover $\hat{\mathbf{y}}^* = \arg\max V_T$. Forward progression and staying incur no cost, while backward transitions and non-adjacent forward jumps are penalized proportionally to their magnitude.

3 Results

3.1 Comparative Performances

Table 1 summarizes the segmentation performance on the REAL-Colon internal test set [1]. The proposed method achieves the highest performance across all segmentation metrics, yielding a wF1 of 0.7621 and a wJacc of 0.6286. Furthermore, it records the lowest WMAPE (0.1400) and reduces the average number

Table 1. Comparison of temporal segmentation models on the REAL-Colon dataset [1]. Evaluation metrics include weighted F1 (wF1), weighted Jaccard (wJacc), weighted mean absolute percentage error (WMAPE), and the average number of label transitions (Transitions). Best results are highlighted in **bold**.

Method	wF1 (↑)	wJacc (↑)	WMAPE (↓)	Transitions (↓)
TeCNO [7]	0.6242	0.5147	0.5203	92.34
ASFormer [26]	0.6884	0.5503	0.2050	**13.75**
ColonTCN [2]	0.7120	0.5748	0.2566	30.25
Ours	**0.7621**	**0.6286**	**0.1400**	15.84

of label transitions to 15.84 per video, demonstrating improved temporal consistency compared to the baseline models.

To assess generalizability beyond the training distribution, we evaluate the models on held-out CAS-Colon videos without any target-domain fine-tuning [23]. As reported in Table 2, the proposed method maintains its performance advantages on the unseen dataset, achieving a wF1 of 0.7300 and a wJacc of

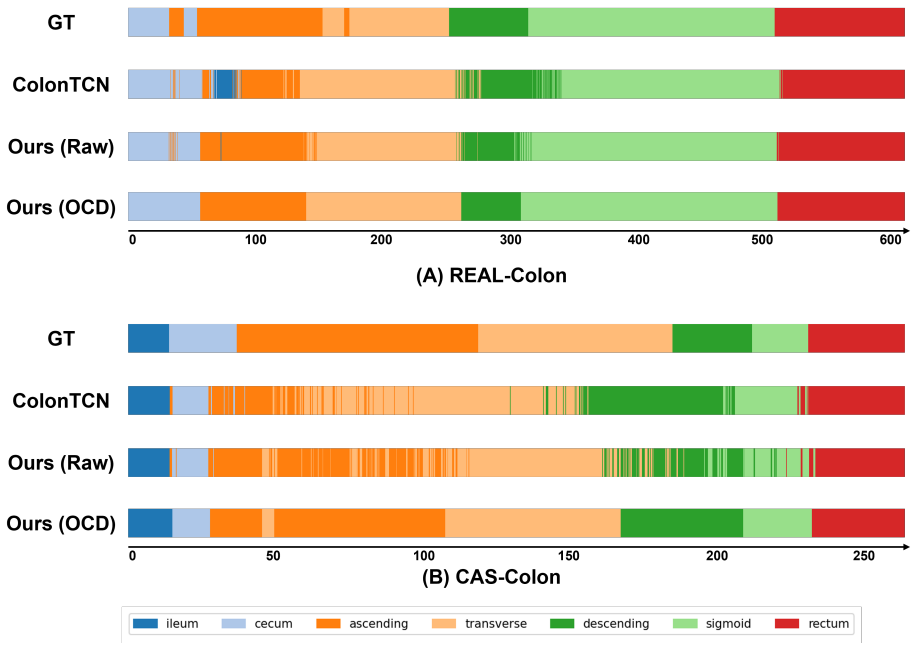

Fig. 3. Qualitative temporal segmentation results on representative videos from the (A) REAL-Colon and (B) CAS-Colon test sets. Each bar visualizes the predicted anatomical segment sequence over the withdrawal phase. Ours (Raw) denotes the model predictions before applying ordinal-constrained decoding (OCD), while 'Ours (OCD)' presents the final output after the OCD process.

Table 2. Cross-dataset generalization results on the CAS-Colon dataset. Models were trained on REAL-Colon and evaluated without any target-domain fine-tuning. Best results are highlighted in **bold**.

Method	wF1 (↑)	wJacc (↑)	WMAPE (↓)	Transitions (↓)
TeCNO [7]	0.6572	0.5355	0.3722	31.25
ASFormer [26]	0.6853	0.5367	0.2646	**6.687**
ColonTCN [2]	0.6412	0.4922	0.3039	79.31
Ours	**0.7300**	**0.5935**	**0.1420**	7.750

Table 3. Ablation study of the proposed components on the REAL-Colon dataset. Evaluation metrics include weighted F1 (wF1), weighted Jaccard (wJacc), weighted mean absolute percentage error (WMAPE), and the average number of label transitions (Transitions). Best results are highlighted in **bold**.

Method	wF1 (↑)	wJacc (↑)	WMAPE (↓)	Transitions (↓)
Baseline (ResNet50 Encoder [12])	0.6850	0.5230	0.4363	46.75
+ Using GastroNet Encoder [3]	0.7255	0.5830	0.2363	31.25
+ Ordinal Regression Head	0.7560	0.6123	0.1852	26.75
+ Ordinal-constrained Decoding	**0.7621**	**0.6286**	**0.1400**	15.84

0.5935. Additionally, the WMAPE remains low at 0.1420, indicating robust segment duration estimation across different clinical centers.

As shown in Fig. 3 (A), ColonTCN [2] produces frequent label flickering at segment boundaries, including erroneous transitions between non-adjacent regions. Our model reduces this effect in its raw output, and the OCD further suppresses residual flickering to yield temporally coherent predictions. On CAS-Colon (Fig. 3 (A)), ColonTCN [2] shows substantial misclassification in the ascending and descending, while our model maintains consistent predictions.

3.2 Ablation Study

We performed an ablation study to evaluate the contribution of each proposed component. Table 3 reports the results on the REAL-Colon test set [1]. The baseline model yields a wF1 of 0.6850 and a WMAPE of 0.4363. Integrating the modified encoder improves the wF1 to 0.7255 and substantially reduces the WMAPE to 0.2363. The subsequent addition of the ORH further increases the wF1 to 0.7560 and decreases the WMAPE to 0.1852. Finally, applying the OCD during inference achieves the best overall performance, peaking at a wF1 of 0.7621 and minimizing the WMAPE to 0.1400.

These results highlight the complementary roles of the two components. The ORH reduces ordinal violations in the raw output, mitigating the risk of OCD propagating erroneous predictions, while OCD eliminates residual boundary flickering that persists without decoding constraints.

4 Discussion

In this study, we proposed an ordinal-aware framework for colonoscopy temporal segmentation incorporating ordinal regression loss and a transition-constrained decoder. It consistently outperformed existing baselines on both the internal test set and the generalization benchmark, with a notably reduced number of label transitions, confirming that ordinal priors provide a generalizable inductive bias for this task. The ablation study confirms that each component contributes to both segmentation accuracy and temporal consistency.

A limitation concerns the decoding strength: overly aggressive α and β may smooth brief but locally valid retrograde events near ambiguous boundaries such as cecum–ascending, so the penalties should be tuned to balance flickering suppression with preservation of genuine short revisits.

Clinically, our framework targets offline post-procedure reporting, documenting examined regions and lesion locations from the completed video. This scope does not cover full-procedure endpoints such as cecal intubation assessment or real-time deployment, which we leave to future work along with integration into automated reporting pipelines.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00454977), and by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No. RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Biffi, C., Antonelli, G., Bernhofer, S., Hassan, C., Hirata, D., Iwatate, M., Maieron, A., Salvagnini, P., Cherubini, A.: Real-colon: A dataset for developing real-world ai applications in colonoscopy. Scientific Data **11**(1), 539 (2024)
2. Biffi, C., Roffo, G., Salvagnini, P., Cherubini, A.: A temporal convolutional network-based approach and a benchmark dataset for colonoscopy video temporal segmentation. Computer Methods and Programs in Biomedicine **270**, 108782 (2025)
3. Boers, T.G., Fockens, K.N., van der Putten, J.A., Jaspers, T.J., Kusters, C.H., Jukema, J.B., Jong, M.R., Struyvenberg, M.R., de Groof, J., Bergman, J.J., et al.: Foundation models in gastrointestinal endoscopic ai: Impact of architecture, pretraining approach and data efficiency. Medical Image Analysis **98**, 103298 (2024)
4. Bourgouin, S., Bège, T., Lalonde, N., Mancini, J., Masson, C., Chaumoitre, K., Brunet, C., Berdah, S.V.: Three-dimensional determination of variability in colon anatomy: applications for numerical modeling of the intestine. journal of surgical research **178**(1), 172–180 (2012)

5. Cao, W., Mirjalili, V., Raschka, S.: Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognition Letters **140**, 325–331 (2020). https://doi.org/https://doi.org/10.1016/j.patrec.2020.11.008, http://www.sciencedirect.com/science/article/pii/S016786552030413X
6. Chang, Y.Y., Li, P.C., Chang, R.F., Chang, Y.Y., Huang, S.P., Chen, Y.Y., Chang, W.Y., Yen, H.H.: Development and validation of a deep learning-based algorithm for colonoscopy quality assessment. Surgical Endoscopy **36**(9), 6446–6455 (2022)
7. Czempiel, T., Paschali, M., Keicher, M., Simson, W., Feussner, H., Kim, S.T., Navab, N.: Tecno: Surgical phase recognition with multi-stage temporal convolutional networks. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2020 - 23nd International Conference, Shenzhen, China, October 4-8, 2020, Proceedings, Part III. Lecture Notes in Computer Science, vol. 12263, pp. 343–352. Springer (2020)
8. De Carvalho, T., Kader, R., Brandao, P., González-Bueno Puyal, J., Lovat, L.B., Mountney, P., Stoyanov, D.: Automated colonoscopy withdrawal phase duration estimation using cecum detection and surgical tasks classification. Biomedical optics express **14**(6), 2629–2644 (2023)
9. Demir, K.C., Schieber, H., Weise, T., Roth, D., May, M., Maier, A., Yang, S.H.: Deep learning in surgical workflow analysis: a review of phase and step recognition. IEEE Journal of Biomedical and Health Informatics **27**(11), 5405–5417 (2023)
10. Farha, Y.A., Gall, J.: Ms-tcn: Multi-stage temporal convolutional network for action segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3575–3584 (2019)
11. Hanson, M.E., Pickhardt, P.J., Kim, D.H., Pfau, P.R.: Anatomic factors predictive of incomplete colonoscopy based on findings at ct colonography. American journal of roentgenology **189**(4), 774–779 (2007)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
13. Kim, B.S., Cho, M., Chung, G.E., Lee, J., Kang, H.Y., Yoon, D., Cho, W.S., Lee, J.C., Bae, J.H., Kong, H.J., et al.: Density clustering-based automatic anatomical section recognition in colonoscopy video using deep learning. Scientific Reports **14**(1), 872 (2024)
14. Koutroumpakis, E., Katsanos, K.H.: Implementation of the simple endoscopic activity score in crohn's disease. Saudi Journal of Gastroenterology **22**(3), 183–191 (2016)
15. Lai, E.J., Calderwood, A.H., Doros, G., Fix, O.K., Jacobson, B.C.: The boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointestinal endoscopy **69**(3), 620–625 (2009)
16. Li, J., Todorovic, S.: Anchor-constrained viterbi for set-supervised action segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9806–9815 (2021)
17. Lieberman, D., Nadel, M., Smith, R.A., Atkin, W., Duggirala, S.B., Fletcher, R., Glick, S.N., Johnson, C.D., Levin, T.R., Pope, J.B., et al.: Standardized colonoscopy reporting and data system: report of the quality assurance task group of the national colorectal cancer roundtable. Gastrointestinal endoscopy **65**(6), 757–766 (2007)
18. Mobadersany, P., Parmar, C., Damasceno, P.F., Fadnavis, S., Chaitanya, K., Li, S., Schwab, E., Xiao, J., Surace, L., Mansi, T., et al.: Harnessing temporal information for precise frame-level predictions in endoscopy videos. In: International Conference

on Medical Image Computing and Computer-Assisted Intervention. pp. 295–305. Springer (2024)
19. Rees, C.J., Gibson, S.T., Rutter, M.D., Baragwanath, P., Pullan, R., Feeney, M., Haslam, N.: Uk key performance indicators and quality assurance standards for colonoscopy. Gut **65**(12), 1923–1929 (2016)
20. Saito, H., Tanimoto, T., Ozawa, T., Ishihara, S., Fujishiro, M., Shichijo, S., Hirasawa, D., Matsuda, T., Endo, Y., Tada, T.: Automatic anatomical classification of colonoscopic images using deep convolutional neural networks. Gastroenterology report **9**(3), 226–233 (2021)
21. Satsangi, J., Silverberg, M., Vermeire, S., Colombel, J.: The montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut **55**(6), 749–753 (2006)
22. Siegel, R.L., Wagle, N.S., Cercek, A., Smith, R.A., Jemal, A.: Colorectal cancer statistics, 2023. CA: a cancer journal for clinicians **73**(3), 233–254 (2023)
23. Song, Y., Zhang, Z., Wang, R., Zhong, L., Cai, C., Chen, J., Zhou, Y., Wang, X., Li, Z., Yang, L., et al.: Cas-colon: A comprehensive colonoscopy anatomical segmentation dataset for artificial intelligence development. Scientific Data **12**(1), 1382 (2025)
24. Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE transactions on Information Theory **13**(2), 260–269 (2003)
25. Wienholt, P., Hermans, A., Khader, F., Puladi, B., Leibe, B., Kuhl, C., Nebelung, S., Truhn, D.: An ordinal regression framework for a deep learning based severity assessment for chest radiographs. arXiv preprint arXiv:2402.05685 (2024)
26. Yi, F., Wen, H., Jiang, T.: Asformer: Transformer for action segmentation. In: The British Machine Vision Conference (BMVC) (2021)
27. Zhang, B., Goel, B., Sarhan, M.H., Goel, V.K., Abukhalil, R., Kalesan, B., Stottler, N., Petculescu, S.: Surgical workflow recognition with temporal convolution and transformer for action segmentation. International Journal of Computer Assisted Radiology and Surgery **18**(4), 785–794 (2023)
28. Zhou, W., Yao, L., Wu, H., Zheng, B., Hu, S., Zhang, L., Li, X., He, C., Wang, Z., Li, Y., et al.: Multi-step validation of a deep learning-based system for the quantification of bowel preparation: a prospective, observational study. The Lancet Digital Health **3**(11), e697–e706 (2021)