IMPROVING STOOL SEGMENTATION IN ABDOMINAL X-RAYS USING GAS MASK GUIDANCE WITH U-NET

NAOYA TAKASHIMA[1], TSUYOSHI SANUKI[2], YOSHIKAZU KINOSHITA[2], SYOJI KOBASHI[1]

[1] University of Hyogo, Graduate School of Engineering, Hyogo, Japan
[2] Hyogo Prefectural Harima-Himeji General Medical Center, Japan
E-MAIL: kobashi@eng.u-hyogo.ac.jp

Abstract:

Segmenting targets with ambiguous features in medical images poses a persistent challenge for quantitative analysis. This is particularly true for stool segmentation in abdominal X-rays—vital for the Stool Volume Score (SVS) in constipation management—where conventional single-channel deep learning methods often struggle. This study investigates if incorporating expert-annotated intestinal gas masks as an explicit, additional input channel to a U-Net model enhances stool segmentation accuracy and SVS reliability. An ablation study compared a dual-channel U-Net (X-ray + gas mask) against an identical single-channel baseline (X-ray only), using the same architectures and training protocols. Performance was assessed using the Dice coefficient, Precision, Recall, and SVS analysis (Pearson correlation, Bland-Altman agreement). The dual-channel model showed an improved Dice coefficient (0.726 vs. 0.710) and Recall (0.706 vs. 0.675) compared to the single-channel baseline, while Precision was slightly higher for the baseline (0.748 vs. 0.747). Crucially, the enhanced segmentation by the dual-channel model yielded a stronger correlation between predicted and ground-truth SVS (Pearson r=0.852 vs. 0.818) and excellent SVS agreement (Bland-Altman bias: 0.01, limits of agreement: -0.06 to 0.07). Explicit gas mask guidance thus improves overall stool segmentation, primarily by achieving a more comprehensive capture of stool regions, which in turn enables more reliable SVS quantification. This dual-channel strategy offers a promising approach for objective fecal load assessment in clinical practice.

Keywords:

Ambiguous segmentation; Stool segmentation; Abdominal X-ray; Deep learning; U-Net; Medical image segmentation

1. Introduction

Accurate segmentation of anatomical structures and pathological regions in medical images is indispensable for quantitative assessment, diagnostic support, and treatment planning [1]. This importance is particularly pronounced in clinical scenarios demanding objective metrics for evaluating disease status and guiding management decisions. However, in many medical images, "ambiguous segmentation tasks" arise due to factors such as low contrast of the target region, indistinct boundaries, or obscuration by other structures [2]. These ambiguities present significant challenges for conventional segmentation algorithms, hindering accuracy improvements. This issue is a common and universal challenge across various modalities and application fields.

Deep learning, particularly convolutional neural networks (CNNs) like U-Net, has achieved remarkable success in various medical image segmentation tasks [1]. However, in the ambiguous cases mentioned above, standard approaches relying solely on the primary image input often struggle to accurately identify the target regions [3]. Local image features alone are often insufficient to resolve these ambiguities.

One promising strategy to improve segmentation accuracy in such ambiguous regions is to incorporate supplementary information or contextual priors into deep learning models. Indeed, studies have shown improvements by providing networks with global binary masks outlining the target organ's approximate location, which notably boosts accuracy in low-data regimes [4]. Others have successfully utilized patient-specific prior segmentations from earlier scans as an input channel to guide delineation in subsequent images, significantly enhancing consistency and accuracy, particularly in radiotherapy planning [5]. Furthermore, incorporating masks of adjacent anatomical structures [6] or structural priors derived from other modalities or analyses [7] has been explored. These approaches collectively suggest that guiding the network with explicit spatial constraints or prior knowledge, often encoded as binary or probability masks, can effectively address challenges posed by ambiguous boundaries or low-contrast targets by focusing the model's attention and leveraging structural relationships.

Abdominal radiography (X-ray) is a widely used modality due to its accessibility and non-invasive nature. However, it presents significant challenges for automated image analysis, particularly for segmenting stool regions [8]. Stool often appears with low contrast against surrounding tissues and has indistinct boundaries, making its accurate delineation an ambiguous segmentation task. To address this, this study introduces and investigates a deep learning approach where expert-annotated intestinal gas masks are utilized as an explicit, additional input channel to a U-Net model. This dual-channel architecture is designed to leverage the clearer structural information from gas patterns to improve the segmentation of the more ambiguous stool regions within the same X-ray image.

The primary objective of this research is to quantitatively evaluate the effectiveness of this dual-channel U-Net model (X-ray + gas mask) in enhancing stool segmentation accuracy. Specifically, we conduct a rigorous ablation study to compare its performance against an identical single-channel U-Net baseline (X-ray only) which does not receive gas mask guidance. Both models utilize the same network architecture and training protocols to ensure a fair comparison. Furthermore, we assess the impact of improved segmentation on the reliability of derived quantitative metrics, such as the Stool Volume Score (SVS) [8], for objective fecal load assessment.

2. Materials and methods

2.1. Dataset and annotations

This study utilized a curated dataset of abdominal radiographs designed for developing and validating automated segmentation models for intestinal gas and stool. The dataset encompasses images from a total of 285 unique patients (Dataset A: 95 patients, 48 male, 47 female, from the former Steel Memorial Hirohata Hospital; Dataset B: 190 patients from the Hyogo Prefectural Harima-Himeji General Medical Hospital), carefully selected to represent a diverse range of non-pathological abdominal appearance. To ensure the model's ability to generalize across varied patient conditions, individuals were included randomly from clinical archives, without selection bias concerning constipation symptoms or the visually estimated amount of gas or stool present; images confirmed to show significant lesions were excluded. Stringent anonymization procedures were implemented at both sites, with gender and age data also masked for the Dataset B cohort to maximize patient privacy.

Image acquisition followed a standardized protocol: all radiographs were obtained with the patient in an upright stance, positioned 1,000 mm from the X-ray intensifier, ensuring consistency. Following acquisition and anonymization, each radiograph was manually annotated in detail by expert gastroenterologists using the 3D Slicer software environment (https://www.slicer.org). Specialists meticulously delineated the boundaries of all visible gas and stool regions, aided by interactive tools for brightness/contrast adjustments and zooming, particularly for ambiguous areas such as stool near the liver or in the upper pelvic region. Annotators relied on subtle textural cues, anatomical continuity, edge characteristics, and collaborative peer review for consensus in challenging cases, followed by a final verification step to ensure dataset-wide consistency. Gas and stool were annotated independently as distinct layers to preserve full spatial extent information, as they can occupy overlapping spaces on a 2D image. This annotated dataset forms the foundation for training and evaluating the segmentation models presented in this work. Figure 1 shows example images from the dataset.

(a) Dataset A **(b)** Dataset B

FIGURE 1. Example abdominal X-ray images from Dataset A and B.

2.2. Proposed Dual-Channel Guided U-Net

Figure 2 illustrates the overall architecture of the network proposed in this study for stool region segmentation. The core of our approach is a U-Net model, renowned for its efficacy in biomedical image segmentation, featuring a symmetrical encoder-decoder structure. For efficient feature extraction, the U-Net employs an EfficientNet encoder with weights pre-trained on the ImageNet dataset. The decoder upsamples the feature maps to generate a pixel-level probability map for the stool class.

The key innovation of our proposed model is the utilization of a dual-channel input strategy designed to explicitly provide the network with spatial information about gas distribution. Input abdominal X-ray images are first resized to 512×512 pixels and normalized to the range [0,1]. The network then accepts a two-channel input:

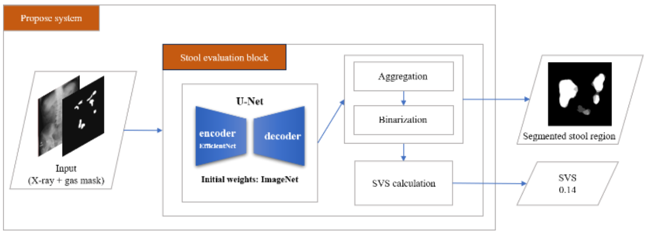


FIGURE 2. Overall architecture of the proposed network architecture. It accepts a two-channel input comprising the abdominal X-ray image and a pre-extracted gas region mask. The central "Stool evaluation block" utilizes a U-Net with an ImageNet pre-trained EfficientNet encoder. The U-Net output undergoes post-processing (Aggregation and Binarization) to generate the final stool region segmentation mask, from which the Stool Volume Score (SVS) is calculated.

- Channel 1: The preprocessed abdominal X-ray image.
- Channel 2: A corresponding pre-extracted binary gas region mask, derived from expert annotations.

Our hypothesis is that providing explicit, pixel-level information about gas pocket locations offers strong contextual cues, enabling the model to better infer the extent of nearby or partially obscured stool regions, which often have complementary spatial distributions. The probability map output from the U-Net undergoes post-processing, including an aggregation step and subsequent binarization using a threshold of 0.5, to obtain the final segmented stool region mask image. From this mask, the SVS is calculated.

2.3. Experimental design and protocols

2.3.1. Ablation study setup

To evaluate the effectiveness of the gas mask guidance, we conducted an ablation study. We compare our proposed dual-channel U-Net model (X-ray + gas mask) with a baseline single-channel U-Net model that segments stool using only the X-ray image as input. Both models share the same U-Net architecture with an EfficientNet backbone and employ identical data preprocessing, ensuring a fair comparison focused solely on the contribution of the gas mask input channel.

2.3.2. Implementation and training details

To train the network, we utilized the Dice Loss function, which is well-suited for segmentation tasks, particularly with potentially imbalanced classes. The Dice Loss (L_{DICE}) is defined as:

$$L_{Dice} = 1 - \frac{1}{N}\sum_{i=1}^{N} \frac{2TP_i}{2TP_i + FP_i + FN_i} \qquad (1)$$

where N represents the number of validation images, and TP_i denotes the number of true positive pixels for the ith image. FP_i and FN_i represent the number of false positive and false negative pixels for the ith image, respectively. Network training was conducted over 200 epochs, utilizing a batch size of 16. The initial learning rate was set to 10^{-3} and adjusted throughout training using a cosine annealing decay schedule. To enhance robustness and prevent overfitting, data augmentation techniques, including random rotations, flipping, scaling, and translations, were applied during the training phase.

2.3.3. Cross-validation strategy

A 5-fold cross-validation approach was adopted for robust evaluation. In each fold, the data was partitioned into 60% for training, 20% for validation, and 20% for testing. Stratified sampling, based on annotated stool volumes, was implemented during data partitioning to mitigate potential sample imbalance bias.

2.4. Evaluation metrics

2.4.1. Segmentation performance metrics

Stool segmentation performance was evaluated using the Dice coefficient, Recall, and Precision. These metrics were calculated based on aggregated true positive (TP), false positive (FP), and false negative (FN) pixel counts from the entire test set for each fold to account for variations in object size and distribution. The definitions are as follows:

$$Dice = \frac{2\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(2TP_i + FP_i + FN_i)} \qquad (2)$$

$$Recall = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \qquad (3)$$

$$Precision = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FP_i)} \qquad (4)$$

2.4.2. Stool volume score and agreement analysis

Building upon our previous work, the SVS was used to quantify the relative proportion of stool in the abdominal region. The SVS is defined as:

$$SVS = \frac{N_{stool}}{N_{abdominal}}, \qquad (5)$$

where N_{stool} denotes the number of pixels labeled as stool and $N_{abdominal}$ is the total number of pixels in the abdominal region. To assess the reliability of the SVS derived from the model's segmentations, we analyzed its

TABLE 1. Quantitative comparison of stool segmentation performance. Bold indicates the highest values.

Method	Dice	Recall	Precision
Single-channel (Baseline)	0.710	0.675	**0.748**
Dual-channel (Proposed)	**0.726**	**0.706**	0.747

agreement with the SVS calculated from ground truth annotations. This involved calculating the Pearson correlation coefficient (r) and performing Bland-Altman analysis to evaluate bias and limits of agreement (LoA).

3. Experimental results and discussion

3.1. Quantitative segmentation performance

The quantitative results comparing the stool segmentation performance of the baseline single-channel model against the proposed dual-channel model with gas mask guidance are presented in Table 1. As shown, the proposed dual-channel method achieved a Dice coefficient of 0.726, compared to 0.710 for the baseline model. The proposed model also demonstrated higher Recall (0.706 vs. 0.675). The baseline model showed slightly higher Precision (0.748 vs. 0.747 for the proposed model).

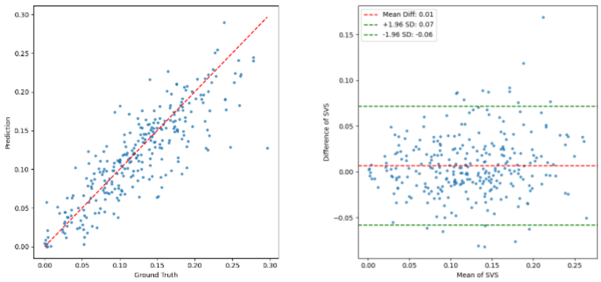
3.2. Stool volume score analysis

The clinical relevance of the segmentation improvements was further assessed by analyzing the derived SVS. The Pearson correlation coefficient between the SVS predicted by the proposed dual-channel model and the ground truth SVS reached 0.852. This marked an improvement over the baseline model, which achieved a correlation coefficient of 0.818. Figure 3(a) provides a scatter plot illustrating this relationship for the proposed method.

The Bland-Altman plot, shown in Figure 3(b), was used to assess the agreement between the predicted and ground truth SVS values for the proposed method. The analysis revealed a mean difference (bias) of 0.01, indicating negligible systematic over- or under-estimation of SVS by the proposed model. The 95% limits of agreement (LoA) were narrow, ranging from -0.06 to 0.07, suggesting good agreement across the range of observed stool volumes.

3.3. Qualitative segmentation examples

Figure 4 provides a visual comparison of segmentation outputs from the baseline and proposed models on a representative challenging case. The proposed dual-channel

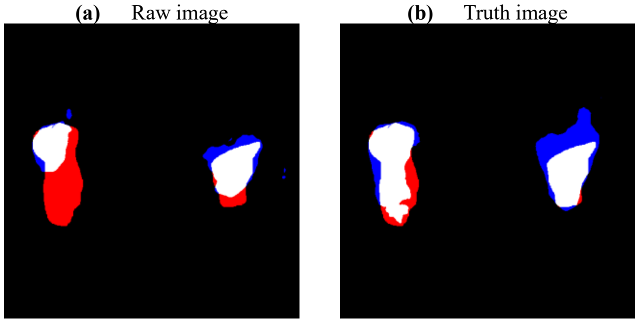


(a) Scatter plot **(b)** Bland-Altman plot

FIGURE 3 Evaluation of SVS predictions.

model (X-ray + gas mask), guided by the gas mask input, demonstrates higher Recall, indicating that it effectively captures more of the true stool regions with fewer missed areas (false negatives) compared to the baseline model. While its Precision is slightly lower than the baseline, suggesting it may include some non-stool areas (false positives), the overall Dice coefficient is improved. The baseline model (X-ray only), conversely, tends to miss more stool regions (lower Recall) despite slightly higher Precision. This visual example illustrates the proposed

(a) Raw image **(b)** Truth image



(c) Single-channel (Baseline) **(d)** Dual-channel (Proposed)

FIGURE 4. Qualitative comparison of stool segmentation results. Visual examples comparing the segmentation output of (c) the baseline model (X-ray only) and (d) the proposed model (X-ray + gas mask). In challenging areas with low contrast, the proposed model (d) effectively identifies more of the true stool regions (higher Recall, thus fewer false negatives indicated by missed red areas if visualizing errors), demonstrating the benefit of gas mask guidance in capturing stool comprehensively. While it may result in slightly more over-extracted areas (false positives, potentially more blue areas if visualizing errors) compared to its slightly higher Precision baseline counterpart (c), the overall balance achieved by the proposed method leads to a better Dice score. The baseline model (c) shows more instances of missed stool regions (lower Recall).

model's strength in achieving comprehensive stool region coverage.

4. Discussion

4.1. Interpretation of segmentation performance

The results of our ablation study demonstrate that incorporating explicit gas mask guidance as an additional input channel to the U-Net model leads to an overall improvement in stool segmentation quality, evidenced by the increased Dice coefficient (0.726 vs 0.710). This improvement is primarily driven by a notable enhancement in Recall (0.706 vs 0.675), indicating that the dual-channel model is more effective at identifying the true extent of stool regions and reducing false negatives. While the Precision of the dual-channel model (0.747) was marginally lower than the baseline (0.748), the substantial gain in Recall contributes to a better overall segmentation performance as reflected by the Dice coefficient. This suggests that the model, guided by gas mask information, becomes more comprehensive in capturing stool, which is particularly valuable in ambiguous, low-contrast regions. The qualitative results in Figure 4, as discussed in Section 3.3, visually support these findings, showing the proposed model's ability to capture more of the true stool regions, even if it occasionally includes some non-stool areas, leading to an overall more accurate segmentation.

4.2. Enhanced reliability of stool volume score

More significantly, the improvements in segmentation accuracy translated directly into a more reliable quantification of fecal load via the SVS. The marked increase in the Pearson correlation coefficient between the predicted SVS (dual-channel model: 0.852) and the ground truth SVS (baseline model: 0.818) underscores this enhancement. This stronger linear relationship indicates that the SVS derived from the gas-guided segmentation more accurately reflects the true stool volume. Furthermore, the Bland-Altman analysis confirmed excellent agreement, with minimal bias (0.01) and narrow limits of agreement (-0.06 to 0.07), reinforcing the conclusion that the proposed method provides a more quantitatively accurate and consistent SVS. This is clinically important, as a reliable SVS can offer an objective measure for diagnosing and managing conditions like chronic constipation.

4.3. Clinical Significance and Advantages

The findings suggest that leveraging readily available anatomical context, such as intestinal gas patterns, can effectively address the challenges of segmenting ambiguously defined targets in abdominal X-rays. By explicitly guiding the segmentation model with gas mask information, our approach improves the delineation of stool boundaries, primarily by achieving a more comprehensive capture of stool regions, as evidenced by the higher Recall. While maintaining a comparable level of Precision to the baseline, this enhancement in correctly identifying more of the actual stool (higher Recall and Dice) is critical for subsequent quantitative analyses like SVS, where under-segmentation can lead to inaccurate assessments of fecal load. The proposed dual-channel strategy offers a practical means to improve the objectivity and reproducibility of stool assessment from X-rays, potentially aiding clinicians in managing chronic constipation more effectively.

5. Conclusions

This study investigated the efficacy of incorporating a "gas mask" guidance strategy within a U-Net architecture to improve the accuracy of automated stool segmentation from abdominal X-ray images and enhance the reliability of the derived SVS.

Our findings, derived from a controlled ablation study, demonstrate clear advantages of the proposed dual-channel approach (X-ray + gas mask) over the conventional single-channel baseline (X-ray only). The proposed method achieved superior overall segmentation quality, reflected in an improved Dice coefficient (0.726 vs 0.710). This enhancement was primarily attributed to a notable increase in Recall (0.706 vs 0.675), signifying the model's improved ability to capture the full extent of stool regions. While Precision was slightly lower for the proposed method (0.747 vs 0.748 for baseline), the significant improvement in Recall led to a better overall Dice score, indicating a more comprehensive and accurate segmentation.

Crucially, the improvements in segmentation translated directly into more accurate quantitative assessments relevant to clinical practice. The correlation between the SVS predicted by our model and the ground truth SVS showed a clear improvement (Pearson correlation: 0.852 for dual-channel vs. 0.818 for baseline). Additionally, the Bland-Altman analysis confirmed excellent agreement between the predicted and ground truth SVS, characterized by a negligible systematic bias (0.01) and narrow limits of agreement (-0.06 to 0.07).

These results strongly suggest that providing explicit spatial guidance via the gas mask input allows the model to better interpret challenging image features and delineate stool boundaries more comprehensively. The enhanced

Recall, leading to fewer missed stool regions, and the subsequent improvement in SVS reliability underscore the value of this guidance strategy.

In conclusion, the integration of gas mask guidance presents a validated and effective method for improving the accuracy and reliability of automated stool segmentation and SVS calculation from X-ray images. This approach holds significant potential for clinical utility, offering a more robust tool for objective fecal load assessment in the diagnosis and management of conditions such as chronic constipation.

References

[1] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, “Segment anything in medical images”, Nature Communications, Vol. 15, No. 1, p. 654, 2024.

[2] Z. Li, W. Tang, S. Gao, Y. Wang, and S. Wang, “Adapting SAM2 model from natural images for tooth segmentation in dental panoramic X-ray images”, Entropy, Vol. 26, No. 12, p. 1059, 2024.

[3] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, “Structure boundary preserving segmentation for medical image with ambiguous boundary”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4817–4826, 2020.

[4] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, “Medical Image Segmentation Review: The Success of U-Net”, IEEE Trans Pattern Anal Mach Intell, Vol. 46, No. 12, pp. 10076-10095, Dec. 2024.

[5] M. Kazemimoghadam, Z. Yang, M. Chen, L. Ma, W. Lu, and X. Gu, “Leveraging global binary masks for structure segmentation in medical images”, Physics in Medicine & Biology, Vol. 68, No. 18, p. 185023, Sep. 2023.

[6] C. M. Kensen, R. Simões, A. Betgen, L. Wiersema, D. M. J. Lambregts, F. P. Peters, C. A. M. Marijnen, U. A. van der Heide, and T. M. Janssen, “Incorporating patient-specific information for the development of rectal tumor auto-segmentation models for online adaptive magnetic resonance Image-guided radiotherapy”, Physics and Imaging in Radiation Oncology, Vol. 32, p. 100648, 2024.

[7] L. Cai, M. A. Abdelatty, L. Han, D. M. J. Lambregts, J. van Griethuysen, E. Pooch, R. G. H. Beets-Tan, S. Benson, J. Brunekreef, and J. Teuwen, “Improving Rectal Tumor Segmentation with Anomaly Fusion Derived from Anatomical Inpainting: A Multicenter Study”, medRxiv, 2024.

[8] X. Zhu, H. Jiang, and Z. Diao, “CGBO-Net: Cruciform structure guided and boundary-optimized lymphoma segmentation network”, Computers in Biology and Medicine, Vol. 153, p. 106534, 2023.