Leveraging Zero-Shot Medical Segmentation Models for Data Annotation and Model Training in Lumbar Vertebrae Segmentation

Veysel Kocaman[*1] VEYSEL@GESUND.AI
Sumir Patel[*1] SUMIR@GESUND.AI
Enes Hosgor[*1] ENES@GESUND.AI
[1] *47 Thorndike St, Cambridge, MA 02141, US*

Editors: Under Review for MIDL 2025

Abstract

This study investigates the potential of zero-shot medical segmentation models, specifically TotalSegmentator and MedSAM-2, for automated data annotation and subsequent training of UNet models in lumbar vertebrae segmentation tasks. We evaluated the performance of these models on both binary and multi-label segmentation tasks using a ground truth test set. Three UNet models (two binary and one multi-label) were trained using masks generated by TotalSegmentator. The performance of MedSAM-2 , TotalSegmentator, and the trained UNet models was assessed on a ground truth test set. Additionally, we evaluated the binary UNet model on a synthetic test set generated by TotalSegmentator. Results show that MedSAM-2 achieved the highest Dice scores in binary tasks (0.829 and 0.739), while TotalSegmentator outperformed in the multi-label scenario (Dice: 0.760). The UNet models trained on TotalSegmentator-generated masks demonstrated competitive performance, particularly in binary tasks (Dice: 0.802 and 0.741). Notably, the binary UNet model achieved a high Dice score of 0.898 on the synthetic test set, indicating strong consistency with TotalSegmentator's annotations. However, the performance gap between synthetic and ground truth evaluations suggests potential domain adaptation challenges. These findings indicate that while zero-shot models can significantly reduce annotation burdens, their synthetic labels may require refinement for optimal downstream model training, especially in complex multi-class settings. This study highlights the context-dependent utility of zero-shot models in medical image segmentation and underscores the importance of robust validation strategies when leveraging synthetic annotations for model training.

Keywords: Zero-shot medical segmentation, TotalSegmentator, MedSAM-2, automated data annotation, UNet models, lumbar vertebrae segmentation

1. Introduction

The field of medical image segmentation has seen significant advancements with the introduction of deep learning techniques. However, the scarcity of annotated medical imaging data remains a significant challenge, particularly in specialized domains such as lumbar vertebrae segmentation. This study explores the potential of zero-shot medical segmentation models, specifically TotalSegmentator (Wasserthal et al., 2023) and MedSAM-2 (Zhu et al., 2024), to address this challenge by automating data annotation and facilitating the training of downstream models. Zero-shot learning models have gained attention for their ability to

[*] Contributed equally

perform tasks without prior training on specific datasets. In the context of medical imaging, these models offer the promise of reducing the labor-intensive process of manual annotation while potentially improving the scalability and generalizability of segmentation tasks. This research investigates the efficacy of such models in the domain of lumbar vertebrae segmentation, a critical task in diagnosing and treating spinal conditions. The primary objectives of this study are:

- To evaluate the performance of zero-shot models (TotalSegmentator and MedSAM-2) in lumbar vertebrae segmentation tasks.
- To assess the viability of using synthetically generated annotations from these models to train downstream UNet models.
- To compare the performance of trained UNet models against zero-shot models on both ground truth and synthetic test sets.

By addressing these objectives, this study aims to provide insights into the potential and limitations of leveraging zero-shot models for medical image segmentation tasks, particularly in the context of lumbar vertebrae segmentation.

2. Relevant Work

Medical image segmentation plays a crucial role in diagnosis, treatment planning, and disease monitoring. However, the scarcity of annotated data in the medical domain poses significant challenges for training deep learning models (Shaharabany and Wolf, 2024). Zero-shot learning (ZSL) has emerged as a promising approach to address this limitation by enabling models to recognize and classify unseen classes without requiring additional labeled data. In the context of lumbar vertebrae segmentation, traditional methods have faced difficulties due to the complex anatomy and variability in image quality. Recent advancements in deep learning techniques have shown potential in automating this task, but the need for large annotated datasets remains a bottleneck (Lu et al., 2023).

Shaharabany and Wolf (2024) proposed a novel approach for minimally-guided zero-shot segmentation of medical images using the Segment Anything Model (SAM). Their method leverages SAM's ability to segment arbitrary objects in natural scenes and adapts it to the medical domain without the need for labeled medical data, except for a few foreground and background points on the test image itself (Shaharabany and Wolf, 2024).

These studies highlight the growing interest in leveraging zero-shot learning techniques to improve medical image segmentation, particularly in addressing the challenges associated with limited annotated data in specialized domains such as lumbar vertebrae segmentation.

3. Materials and Methods

The study utilized multiple datasets to ensure a comprehensive evaluation:

- Training and Validation Data: The Mendeley Lumbar Spine dataset (Sudirman et al., 2021) was used, comprising 503 training cases and 56 validation cases of T2-weighted sagittal MRI scans.

- Multi-label Test Set: A subset of 96 images extracted from the SPIDER dataset (van der Graaf et al., 2024).

TotalSegmentator was employed to generate two types of annotations: Binary labels focusing on vertebral bodies and multi-label annotations distinguishing each vertebra separately, including posterior arches.

MedSAM-2 inference was prompted using bounding boxes derived from TotalSegmentator outputs, providing an additional layer of annotation. Several image augmentation techniques were applied to increase dataset diversity and model robustness.

The study employed a 3D-UNet model architecture with a ResNet50 backbone. This choice was motivated by the architecture's capability for deep feature extraction and its effectiveness in mitigating the vanishing gradient problem. Three UNet models were trained using masks generated by TotalSegmentator: Two binary segmentation models and One multi-label segmentation model

4. Results and Discussion

MedSAM-2 demonstrated strong performance in binary segmentation, achieving the highest Dice scores (0.829 and 0.739). UNet models trained on TotalSegmentator-generated masks showed competitive results (0.802 and 0.741). For multi-label segmentation, TotalSegmentator outperformed other models with a Dice score of 0.760. Additionally, the binary UNet model achieved a high Dice score of 0.898 on a synthetic test set, indicating strong consistency with synthetic annotations.

5. Conclusion

This study highlights the potential of zero-shot medical segmentation models in automating data annotation and enabling the training of downstream models for lumbar vertebrae segmentation. The strong performance of MedSAM-2 in binary tasks and TotalSegmentator in multi-label scenarios demonstrates the feasibility of using such models to generate high-quality annotations with minimal human intervention. Additionally, the competitive performance of UNet models trained on synthetically generated labels underscores the viability of leveraging zero-shot model outputs for supervised learning in medical image segmentation.

However, the performance gap between synthetic and ground truth evaluations, particularly in multi-label segmentation, suggests that domain adaptation remains a key challenge. The increased complexity of distinguishing individual vertebrae highlights the need for more advanced techniques to enhance segmentation accuracy in multi-class tasks. Addressing these limitations through improved synthetic label refinement and hybrid learning strategies could further bridge the gap between zero-shot and human-annotated datasets.

Overall, while zero-shot models offer a promising avenue for reducing annotation burdens and accelerating medical image analysis, their deployment must be carefully tailored to account for task complexity and domain-specific challenges.

References

Hongjiang Lu, Mingying Li, Kun Yu, Yingjiao Zhang, and Liang Yu. Lumbar spine segmentation method based on deep learning. *Journal of applied clinical medical physics*, 24 (6):e13996, 2023.

Tal Shaharabany and Lior Wolf. Zero-shot medical image segmentation based on sparse prompt using finetuned sam. In *Medical Imaging with Deep Learning*, 2024.

Sud Sudirman, Ala Al Kafri, Friska Natalia, Hira Meidia, Nunik Afriliana, Wasfi Al-Rashdan, Mohammad Bashtawi, and Mohammed Al-Jumaily. Composite dataset of lumbar spine mid-sagittal images with annotations and clinically relevant spinal measurements, 2021.

Jasper W van der Graaf, Miranda L van Hooff, Constantinus FM Buckens, Matthieu Rutten, Job LC van Susante, Robert Jan Kroeze, Marinus de Kleuver, Bram van Ginneken, and Nikolas Lessmann. Lumbar spine segmentation in mr images: a dataset and a public benchmark. *Scientific Data*, 11(1):264, 2024.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023.

Jiayuan Zhu, Abdullah Hamdi, Yunli Qi, Yueming Jin, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2. *arXiv preprint arXiv:2408.00874*, 2024.