FracSegmentator: Fracture Instance Segmentation with Trauma-Prior-Guided Contrastive Learning

Yanzhen Liu[1], **Sutuke Yibulayimu**[1], **Yang Zhou**[1], **Yudi Sang**[2*], **Yu Wang**[1*]

[1]Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing, China

[2]Beijing Rossum Robot Technology Co., Ltd., Beijing, China

{yanzhenliu, sutuk, zhouyangbme, wangyu}@buaa.edu.cn, sangyudi@rossumrobot.cn

Abstract

Fracture injuries often lead to complex bone fragmentations, posing significant challenges for accurate segmentation in surgical planning and trauma assessment. Manual annotation of each fragment is time-consuming and inconsistent, while existing automated methods often fail to separate individual fragments due to the wide variation in fracture types, irregular fracture surface, and close inter-fragment contact. To address these challenges, we introduce FracSegmentator, a deep learning approach for bone fragment instance segmentation. The model takes extracted bone regions in CT as input and isolates individual fragments by identifying fracture surfaces and separating closely contacting structures. Central to our approach is a Trauma-Prior-Guided Contrastive Learning module, which incorporates clinical knowledge through memory-based attention to better distinguish fractured surfaces from healthy regions. We evaluate FracSegmentator on four datasets that cover a range of anatomical sites and fracture patterns. The method achieves state-of-the-art results across all datasets and demonstrates strong generalization capabilities. By delivering accurate and efficient fragment-level segmentation, FracSegmentator supports critical downstream tasks such as automated fracture diagnosis, surgical planning, and preoperative reduction simulation.

Code & Dataset —
https://github.com/YzzLiu/FracSegmentator

Introduction

Bone fracture is a severe clinical problem in trauma care, often requiring surgical intervention to realign and fix the broken bone pieces. In complex injuries, such as comminuted or high energy fractures, multiple bone fragments must be anatomically reduced before definitive fixation can proceed. Modern orthopedic practice relies on 3D imaging, e.g., CT scanning, to visualize fracture anatomy and plan the reduction procedure. Precise identification and delineation of each bone fragment in these images is essential for guiding surgery and optimizing outcomes. However, manual slice-by-slice segmentation of fracture fragments by experts is extremely labor-intensive, subjective, and error-prone. This motivates automated fragment segmentation methods that can provide fast, objective, and reproducible interpretations to assist clinical decision-making and reduction planning.

Fragment segmentation remains a challenging task due to variations in fragment size, shape, count, as well as the irregular and complex contact fracture surfaces resulting from fragment collisions. Various traditional methods have been explored to separate fracture fragments from CT scans, including fixed or adaptive thresholding (Tomazevic et al. 2010), watershed algorithms (Neubauer et al. 2005), non-rigid registration (Pettersson, Knutsson, and Borga 2006), graph cut approaches (Han et al. 2021), and region growing (Bittner-Frank et al. 2024). These techniques often rely on boundary gradient intensity similarity and continuity for segmentation. However, their implementation is often hindered by the difficulty in accurately selecting appropriate thresholds and the inability to precisely identify and separate contacting fragments.

The challenges in instance segmentation of fractured bones arise from several factors: (1) The wide variation in fracture types, shapes, and fragment quantities makes extracting reliable structural information difficult. (2) The fracture surface can exhibit various forms, including large gaps (where fragments are displaced and isolated), small gaps (where fragments are isolated but stationary), creases (where fragments remain partially connected), and compression (where fragments collide). These diverse fracture patterns pose significant challenges in accurately capturing the surface features. (3) The intricate anatomical context and frequent inter-fragment contact across different skeletal regions introduce ambiguous boundaries and overlapping features, making accurate instance separation particularly challenging.

In this study, we present **FracSegmentator**, a generalizable framework for instance-level bone fracture segmentation. Our major contributions can be summarized in three key aspects:

(1) We propose a two-stage segmentation pipeline that first extracts bone structures and then segments each fracture fragment as an individual instance, addressing the challenge of arbitrary fragment counts and subtle fracture boundaries.

(2) We design a novel Trauma Prior-Guided Contrastive Learning module that leverages domain-specific priors via a memory attention mechanism, enabling the network to bet-

*Co-corresponding authors

ter distinguish between fracture surfaces and intact bone regions.

(3) We reformulate fragment instance segmentation as a multi-class problem through the introduction of a Boundary-Interior-Contact Mask (BICM) representation, which enhances structural supervision and improves separation of touching or adjacent fragments.

We evaluate FracSegmentator on multiple CT datasets and demonstrate state-of-the-art accuracy and robustness.

Related Work

Bone Fracture Instance Segmentation

Early work on fracture segmentation was based on intensity thresholding, region growing, watershed transforms, and graph-cut methods (Neubauer et al. 2005; Bittner-Frank et al. 2024; Tomazevic et al. 2010; Han et al. 2021). These techniques require delicate parameter tuning and frequently break down when fragments are tightly apposed or the fracture surface has low contrast, leading to either under- or over-segmentation. With the rise of deep learning, 3D U-Net variants, Vision Transformers, and nnUNet consistently excel at anatomical segmentation but they do not separate each fragment (Çiçek et al. 2016; He et al. 2023; Isensee et al. 2021). A common practice is to perform connected component analysis on the binary bone mask to get fragment instances, but this fails when fragments touch or overlap. Several studies have introduced auxiliary labels to distinguish primary and secondary fragments, followed by post-processing to separate the secondary pieces. Distance-aware networks, such as FDM-UNet, and dual-stream designs, such as FDD-Net, show improved performance for specific patterns, for example fractures of the anterior and posterior pelvic ring (Liu et al. 2023; Zeng et al. 2024). These methods falter, though, when the primary fragment is ambiguous or when multiple secondary fragments coexist in close proximity. Similarly, Two-stage pipelines that first detect fractures and then segment them encounter similar limitations in regions containing several closely spaced breaks (Liu et al. 2024). The scale of the challenge is highlighted by the PENGWIN 2024 benchmark for pelvic fracture segmentation, where the top algorithm achieved 93% mean IoU (Sang et al. 2025). Notably, even state-of-the-art models can struggle with fragments that are non-displaced or overlapping.

Contrastive Learning in Medical Segmentation

Contrastive learning has emerged as a powerful tool to learn discriminative features in medical imaging, both for self-supervised pretraining and as auxiliary loss in segmentation models. Self-supervised frameworks such as MoCo, SimCLR, and Voco have been adapted to medical images for organ or lesion segmentation, often delivering stronger boundary delineation after pre-training (He et al. 2020; Chen et al. 2020; Wu, Zhuang, and Chen 2024). These strategies, however, define positives via generic augmentations or coarse class labels and therefore overlook pathology-specific cues. In the context of trauma imaging, some research utilized a structure-aware contrastive strategy to exploit the bilateral symmetry for fracture detection (Zeng et al. 2023), proving that incorporating domain-specific cues into contrastive learning can improve model sensitivity to abnormalities. In fracture CT, the visual differences between intact cortex and subtle cracks are subtle and highly localized, making off-the-shelf contrastive schemes less effective.

Explicit Priors in Segmentation

Integrating prior knowledge about anatomy or object shape is a well-established strategy to improve segmentation robustness. Traditional medical segmentation methods often employed statistical shape models or atlas-based priors to ensure that results conformed to plausible anatomical shapes (Kalinic 2009; Liang et al. 2022). These priors work well for anatomies with consistent geometry but cannot accommodate the drastic variability in fragment number, size, and configuration characteristic of fractures. Some methods embed prior knowledge by adding auxiliary embedding, such as distance maps, boundary feature channels or task-specific loss, so the model can learn additional cues that guide the segmentation toward more reliable results (Zeng et al. 2024; You et al. 2024). Inspired by these strategies, we introduce trauma priors into our framework. Common patterns observed at fracture sites, including the texture of cortical break surfaces and the gaps between displaced pieces, are encoded in a learnable memory module, enabling the network to retrieve and apply this domain knowledge.

Method

Overview

We aim to automatically segment and isolate each bone fragment in fracture CT. As shown in figure 1, the proposed workflow (FracSegmentator) consists of two major stages. In the initial bone extraction stage, we utilize a 3D UNet to segment anatomical structures, following established methods (Liu et al. 2021, 2025). To improve the model's generalizability, we fine-tune the TotalSegmentator-pretrained weights using fracture datasets (Wasserthal et al. 2023). In the fragment-instance stage, we design a trauma-prior-guided contrastive network (TPC-Net) that predicts a Boundary-Interior-Contact Mask (BICM). The network outputs are integrated to isolate and label each bone fragment.

Bone Extraction

In the first stage, we extract bone regions using a cascaded 3D UNet architecture. The initial UNet operates on downsampled CT images to capture global anatomical context, producing coarse segmentation maps. These are then refined by a second UNet working at full resolution, which takes the original CT along with the coarse predictions as input. To improve generalizability, we initialize the networks with weights pretrained on the TotalSegmentator dataset, which contains 1230 CT scans annotated with comprehensive anatomical labels. The pretrained model is further fine-tuned on fracture datasets to better adapt to pathological variations.

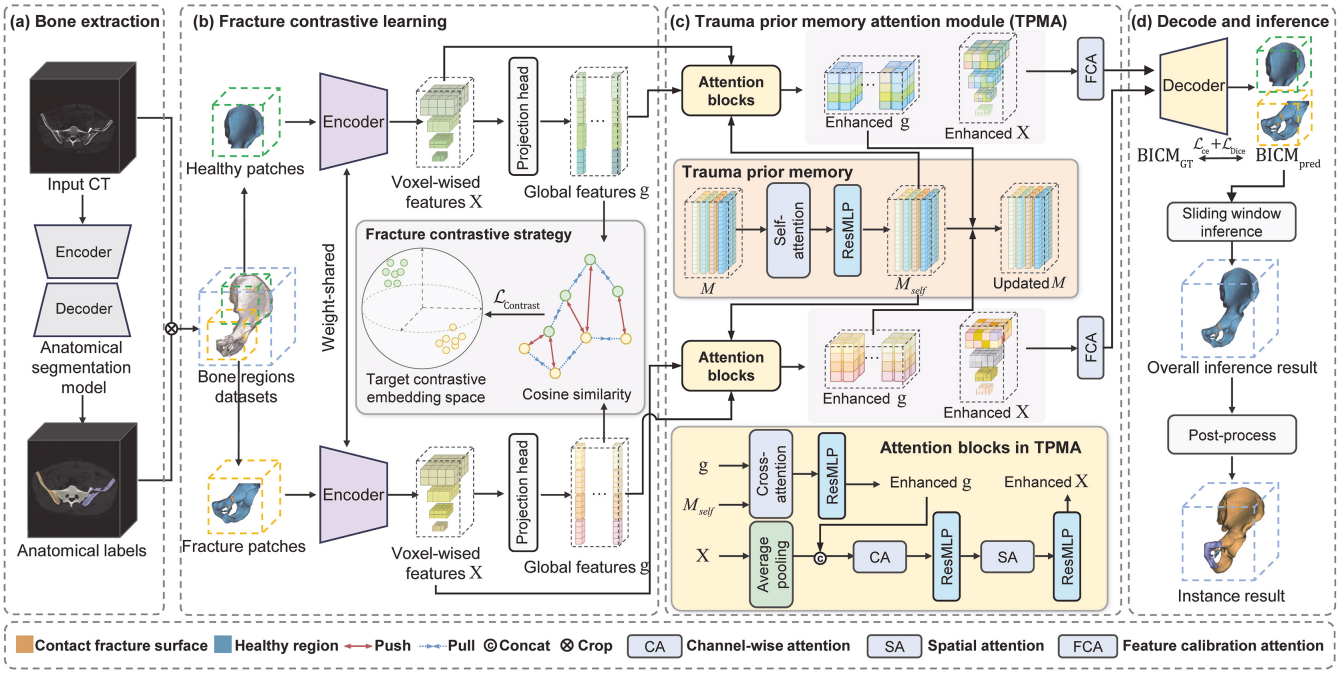

Figure 1: Overview of the proposed workflow, FracSegmentator, consisting of two major stages: (a) the initial bone extraction stage; (b-d) the fragment segmentation stage, which employs the proposed trauma-prior-guided contrastive network (TPC-Net) to predict Boundary-Interior-Contact Masks (BICMs).

Fragment Segmentation

To isolate individual fragments, we propose TPC-Net that predict a BICM directly from bone regions. As shown in figure 1, TPC-Net uses a shared encoder with five convolutional layers to extract multi-scale features from CT patches of healthy and fractured regions. The encoder's ability to identify fractures is enhanced by comparing the differences between the fractured and healthy image patches. In addition, we introduce a trauma prior memory attention module (TPMA) to enhance the discriminative ability of sample features. The prediction is generated by a decoder that processes the feature representation.

Boundary-Interior-Contact Mask. We construct BICM to reformulate the task of fracture instance segmentation. Instead of directly assigning a unique fragment ID to each voxel, which is challenging due to the variable number of fragments across cases, we convert the problem into a three-class semantic segmentation task. Specifically, each foreground voxel is classified into one of the three categories: **(1) Boundary (B):** Voxels on the outermost surface of each individual fragment; **(2) Interior (I):** Voxels in the interior of a fragment, away from any fracture boundary; and **(3) Contact fracture surface (C):** Voxels that belong to a fracture surface in contact with an adjacent fragment, which typically occur in comminuted fractures where pieces are touching.

We derive the training BICM labels from fragment instance labels. To identify contact fracture surface (CFS), we analyze each voxel's local neighborhood within a $7 \times 7 \times 7$ window. If a neighboring voxel belongs to a different fragment, the center voxel is labeled as part of C. To differentiate boundaries from interiors, we apply morphological erosion with an adaptive kernel size guided by local structural thickness. Eroded voxels are labeled as B, while the remaining voxels are assigned to I. This reformulation guides the network to focus on the morphological composition of fractures by distinguishing core anatomical regions, outer boundaries, and inter-fragment interactions. Figure 2 illustrates an exam-

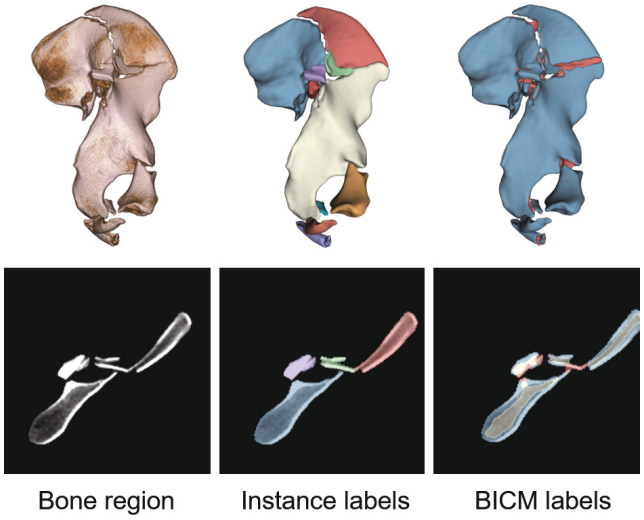

Figure 2: Example of instance and BICM label strategy.

ple of this representation.

Fracture Contrastive Encoder. To help the model distinguish between fractured and healthy bone regions, we design a contrastive encoder that learns discriminative representations at multiple scales.

We begin by randomly sampling an equal number of fractured and healthy 3D patches from CT volumes. Fractured patches contain at least one CFS whereas healthy patches do not. These patches are passed through a shared 3D encoder to extract multi-scale feature maps $\{f_\ell\}_{\ell=1}^{L}$.

At each scale ℓ, the feature map is first processed by a 3D linear convolution to unify the channel dimension. The output is then passed through a projection head $P_\ell(x)$, implemented as a lightweight multi-layer perceptron (MLP), yielding a global embedding vector $\mathbf{z}_\ell^i \in \mathbb{R}^{256}$ for each sample i. All embeddings are ℓ_2-normalized to maintain numerical stability during contrastive learning.

To encourage class-specific feature separation, we apply a contrastive loss at each scale. For a given sample i with label y^i, we select only one same-class sample $p(i)$ as the positive, and treat all samples j such that $y^j \neq y^i$ as negatives. Pairwise cosine similarity is computed as:

$$s_\ell^{ij} = \mathrm{CosSim}(\mathbf{z}_\ell^i, \mathbf{z}_\ell^j) = \frac{(\mathbf{z}_\ell^i)^\top \mathbf{z}_\ell^j}{\|\mathbf{z}_\ell^i\| \cdot \|\mathbf{z}_\ell^j\|}. \tag{1}$$

We adopt the InfoNCE loss to maximize similarity with the positive and minimize similarity with the negatives:

$$\mathcal{L}_\ell^i = -\log \frac{\exp(s_\ell^{i\,p(i)})}{\exp(s_\ell^{i\,p(i)}) + \sum_{y^j \neq y^i} \exp(s_\ell^{ij})}, \tag{2}$$

where y^i and $y^j \in \{0, 1\}$ denote binary labels for healthy and fractured patches, respectively (Parulekar et al. 2023). To capture consistent contrastive signals across scales, we aggregate the losses from all levels:

$$\mathcal{L}_{\text{contrast}} = \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathcal{L}_\ell^i, \tag{3}$$

where $\mathcal{V}$ denotes the set of samples with at least one valid positive.

This multi-scale contrastive objective encourages the encoder to focus on fracture-specific visual cues, such as discontinuities, missing cortical bone, and irregular textures near CFS. As a result, the encoder learns representations that are more sensitive to fracture morphology and better suited for downstream segmentation tasks.

Trauma Prior Memory Attention Module. To incorporate trauma-specific knowledge into the feature encoding process, we introduce the TPMA module. This component enhances feature representations by leveraging a learned memory of fracture patterns, enabling the model to better distinguish true fracture cues from normal anatomical variations.

The TPMA module maintains a learnable memory bank $\mathbf{M} \in \mathbb{R}^{K \times L \times F}$, where K denotes the number of representative entries per scale, L is the number of scales, and F is the feature dimension. Each memory vector captures a prototypical visual pattern observed at fracture interfaces, such as cortical discontinuities or gap-like textures.

To refine the memory embeddings, we apply a multi-head self-attention mechanism followed by a residual MLP update:

$$\mathbf{M}_{self} \leftarrow \mathbf{M} + \mathrm{MLP}\big(\mathrm{LN}(\mathrm{SelfAttn}(\mathbf{M}))\big), \tag{4}$$

where $\mathrm{SelfAttn}(\cdot)$ denotes multi-head attention, and $\mathrm{LN}(\cdot)$ is layer normalization.

Given the global features $\mathbf{g}_\ell \in \mathbb{R}^{B \times F}$ from the encoder at scale s, the module retrieves relevant trauma priors using cross-attention:

$$\tilde{\mathbf{g}}_\ell = \mathbf{g}_\ell + \mathrm{MLP}\big(\mathrm{LN}(\mathrm{CrossAttn}(\mathbf{g}_\ell, \mathbf{M}_{self}, \mathbf{M}_{self}))\big), \tag{5}$$

where $\mathbf{g}_\ell$ is used as the query, and $\mathbf{M}$ serves as both key and value. The output $\tilde{\mathbf{g}}_\ell$ represents trauma-aware features, refined based on similarity to stored prior patterns.

To keep the memory bank representative of recent training data, we update it using enhanced features across all scales:

$$\mathbf{M} \leftarrow \mathrm{UpdateMemory}\big(\mathbf{M}_{self}, \{\tilde{\mathbf{g}}_\ell\}_{\ell=1}^{L}\big). \tag{6}$$

When the number of new features exceeds the memory size, the oldest entries are replaced to maintain a fixed capacity.

To inject this global trauma context into scale-specific feature maps $\mathbf{X}_\ell \in \mathbb{R}^{C_\ell \times H_\ell \times W_\ell \times D_\ell}$, each map is spatially averaged into a channel descriptor $\mathbf{z}_\ell$. Concatenating $\mathbf{z}_\ell$ with $\tilde{\mathbf{g}}_\ell$ and projecting the result yields the Channel-wise Attention (CA) query, key, and value vectors $\mathbf{Q}_\ell^{\mathrm{CA}}, \mathbf{K}_\ell^{\mathrm{CA}}, \mathbf{V}_\ell^{\mathrm{CA}}$. Channel-wise attention weights are then computed as:

$$\mathbf{w}_\ell = \mathrm{softmax}\big(\mathrm{LN}\big(\mathbf{Q}_\ell^{\mathrm{CA}}(\mathbf{K}_\ell^{\mathrm{CA}})^\top\big)\big)\mathbf{V}_\ell^{\mathrm{CA}}, \tag{7}$$

where $\mathrm{softmax}(\cdot)$ denotes the soft-max operation. These weights modulate the original feature map in a residual manner:

$$\hat{\mathbf{X}}_\ell = \mathbf{X}_\ell + \mathbf{X}_\ell \odot \sigma\big(\mathbf{w}_\ell\big), \tag{8}$$

with $\odot$ and $\sigma(\cdot)$ indicating channel-wise multiplication and sigmoid function. This step selectively amplifies trauma-relevant channels while suppressing less informative ones.

The refined maps $\hat{\mathbf{X}}_\ell$ are partitioned into non-overlapping 3D patches, embedded as tokens, and concatenated across scales to form the sequence $\mathbf{T}_\ell$. A Spatial Attention (SA) block then captures long-range and cross-scale interactions:

$$O_\ell^{\mathrm{SA}} = \mathrm{softmax}\big(\mathrm{LN}\big(\mathbf{Q}_\ell^{\mathrm{SA}}(\mathbf{K}_\ell^{\mathrm{SA}})^\top\big)\big)\mathbf{V}_\ell^{\mathrm{SA}}, \tag{9}$$

$$\mathbf{T}'_\ell = \mathrm{MLP}\big(\mathrm{LN}(\mathbf{T}_\ell + O_\ell^{\mathrm{SA}})\big) + \mathrm{LN}(\mathbf{T}_\ell + O_\ell^{\mathrm{SA}}), \tag{10}$$

where $\mathbf{Q}_\ell^{\mathrm{SA}} = \mathbf{W}_{\mathrm{SA}}^{q}\mathbf{T}_\ell$, $\mathbf{K}_\ell^{\mathrm{SA}} = \mathbf{W}_{\mathrm{SA}}^{k}\mathbf{T}_\ell$, and $\mathbf{V}_\ell^{\mathrm{SA}} = \mathbf{W}_{\mathrm{SA}}^{v}\mathbf{T}_\ell$. Finally, the updated tokens are reshaped back to the spatial domain and fused with the corresponding feature maps:

$$\tilde{\mathbf{X}}_\ell = \hat{\mathbf{X}}_\ell + \mathcal{P}_\ell^{-1}\big(\mathbf{T}'_\ell\big). \tag{11}$$

This continuous flow of memory retrieval, channel modulation, and spatial attention equips the decoder with sharper, anatomically consistent cues, facilitating precise segmentation of complex fracture patterns.

Feature Calibration Attention Decoder. In the decoding path, to avoid the misalignment caused by rigid fusion, a decoder-guided cross-attention approach is employed to establish a semantic mapping between the encoder and decoder features. This mechanism recalibrates the channel and spatial responses of the encoder features to facilitate the specific task of each branch. The module utilizes decoder features D_ℓ for queries, and $\tilde{\mathbf{X}}_\ell$ for keys and values.

$$M_\ell = \mathrm{softmax}\big(\mathrm{LN}\big((\mathbf{Q}_\ell^{\mathrm{FCA}})^\top \mathbf{K}_\ell^{\mathrm{FCA}}\big)\big) \tag{12}$$

$$\mathbf{O}_\ell^{\mathrm{FCA}} = \mathrm{Linear}\big(\big(M_\ell\, \mathbf{V}_\ell^{\mathrm{FCA}}\big)^\top\big), \tag{13}$$

where $\mathbf{Q}_\ell^{\mathrm{FCA}} = \mathbf{W}_{\mathrm{FCA}}^q\, \mathrm{Pool}(D_\ell)$, $\mathbf{K}_\ell^{\mathrm{FCA}} = \mathbf{W}_{\mathrm{FCA}}^k\, \tilde{\mathbf{X}}_\ell$, $\mathbf{V}_\ell^{\mathrm{FCA}} = \mathbf{W}_{\mathrm{FCA}}^v\, \tilde{\mathbf{X}}_\ell$. M_ℓ is a similarity matrix capturing channel-level semantic correlations, and $\mathrm{Pool}(\cdot)$ is an adaptive pooling layer with respect to the current patch size. The final output is:

$$D_\ell^{\mathrm{FCA}} = \mathcal{R}(\mathbf{O}_\ell^{\mathrm{FCA}}) \odot \mathcal{M}(D_\ell), \tag{14}$$

where $\mathcal{R}(\cdot)$ reconstructs the aggregated result to the appropriate spatial dimensions, and $\mathcal{M}(D_\ell)$ is a mask obtained by applying convolution, batch normalization, and ReLU activation. The element-wise multiplication of both results achieves feature fusion.

Loss function. The network is trained using a combination of voxel-wise segmentation loss and global contrastive loss. For the segmentation of Boundary, Interior, and Contact regions, we adopt a standard multi-class segmentation objective that combines cross-entropy and Dice loss (Isensee et al. 2021). At the same time, we apply the global contrastive loss to enforce separation between fractured and intact representations in the embedding space. Thus, the total training objective is:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{ce}} + \mathcal{L}_{\mathrm{Dice}} + \mathcal{L}_{\mathrm{contrast}}. \tag{15}$$

This joint supervision strategy guides the encoder to produce feature representations that are both semantically informative for segmentation and structurally discriminative with respect to fracture presence. As a result, the model gains a stronger understanding of both local boundary morphology and global anatomical integrity.

Integrating Topological Information for Instance Segmentation

To produce the final instance-level segmentation of fracture fragments, we incorporate topological structure into a targeted post-processing pipeline. The process begins by excluding CFS regions, which are first expanded using morphological dilation. This step prevents neighboring fragments from being erroneously merged due to shared or touching boundaries. We then isolate voxels labeled as interior and apply connected component analysis to identify distinct fragment candidates. Each connected component is treated as a potential fracture instance. To suppress topological noise, components with a spatial diameter smaller than 5 mm are discarded and reclassified as boundary voxels. Finally, all remaining boundary and CFS voxels are assigned to their nearest interior region by their Euclidean distance.

Experiments

We conduct extensive experiments to evaluate the performance and generalizability of the proposed FracSegmentator across multiple anatomical regions and fracture types. We validate the method using three curated datasets (SacrumFrac, HipFrac, and FemurFrac) and compare against state-of-the-art segmentation baselines. In addition, we perform ablation studies on key components.

Experimental Setup

Data. We utilize four datasets across different anatomical regions:

- **TotalSegmentator** (Wasserthal et al. 2023): A large-scale whole-body CT dataset with over 1,200 subjects and 104 anatomical structure labels. We use its skeleton segmentation subset to pretrain the encoder for anatomical feature extraction.
- **PENGWIN** (Sang et al. 2025) and **CTPelvic1K** (Liu et al. 2021): Two public pelvic CT datasets. We extract sacrum and hip regions in the fracture cases from each dataset and re-organize them into two task-specific benchmarks: **SacrumFrac** (150 cases) and **HipFrac** (300 cases).
- **FemurFrac** (Private): An internal dataset comprising 350 CT scans with radiologist-annotated femoral fractures. The dataset covers a wide spectrum of fracture types, including shaft, neck, and intertrochanteric fractures.

Each segmentation task (SacrumFrac, HipFrac, FemurFrac) is treated as an independent dataset. We train a separate model for each task and randomly split each dataset into 70% training and 30% testing subsets. All annotations are manually verified by two experts in orthopedics for consistency and accuracy. Additional experiments on the full PENGWIN dataset are included in the supplementary material for direct comparison with prior PENGWIN leaderboard methods (Sang et al. 2025).

Evaluation. To evaluate performance, each ground-truth fragment was paired with a predicted fragment based on the highest intersection over union (IoU), ensuring one-to-one correspondence. We then assessed segmentation accuracy using four metrics: the IoU Dice similarity coefficient (DSC), the average symmetric surface distance (ASSD), and the 95th percentile of the Hausdorff distance (HD_{95}).

In cases where a fragment was missing in the prediction, HD_{95} and ASSD were set to the diameter and radius, respectively, of the ground-truth fragment's bounding sphere (Sang et al. 2025).

Implementation. Experiments were conducted on an Intel Xeon 16-core CPU, an H100 GPU, and 64 GB of RAM. All CT volumes were resampled with third-order B-spline interpolation to an average spacing and z-score normalized. For the bone-extraction stage, entire scans were processed, whereas for fragment segmentation each scan was cropped to the bone bounding-box. Patch size was set to $128 \times 128 \times 128$.

Figure 3: Example instance segmentation results. (a) Fragments are displaced. (b) Fragments are isolated but not displaced. (c) Fragments are partially connected. (d) Fragments are collided. (e) An example case rendered in 3D view.

Method	SacrumFrac				HipFrac				FemurFrac			
	IoU↑	DSC↑	HD_{95}↓	ASSD↓	IoU↑	DSC↑	HD_{95}↓	ASSD↓	IoU↑	DSC↑	HD_{95}↓	ASSD↓
3DUNet	83.29	87.12	13.76	3.31	81.72	83.97	25.69	6.52	88.48	92.56	13.33	2.56
ResUNet	85.38	87.78	9.75	2.87	83.34	86.08	25.26	5.92	90.68	93.58	14.24	2.76
TransUNet	84.46	88.01	12.92	3.18	84.18	87.06	22.95	5.31	90.76	93.69	14.56	2.79
SwinUNETR	87.58	88.82	9.70	2.99	85.19	87.38	21.89	5.26	90.77	93.70	14.57	2.79
nnUNet	90.84	92.17	7.76	2.45	89.60	91.18	7.83	2.76	91.70	94.46	11.76	2.22
MedNext	92.36	94.59	6.36	1.85	93.56	94.87	9.64	2.17	95.78	97.03	7.00	1.60
U-Mamba	91.04	93.03	7.80	2.21	90.74	92.41	13.38	3.36	94.54	96.33	5.72	1.20
FDMUNet	90.05	91.71	5.97	2.01	87.23	88.66	10.42	3.04	91.06	93.93	13.30	2.42
FDDNet	87.61	89.04	8.27	2.60	86.45	87.99	11.04	3.15	90.86	93.75	13.81	2.51
Ours	**96.56**	**97.59**	**3.22**	**1.30**	**95.58**	**96.78**	**5.06**	**1.50**	**96.58**	**97.74**	**4.92**	**1.04**

Table 1: Quantitative results. HD_{95} and ASSD values are reported in millimeter. Quantitative results. DSC and IoU are reported as percentages (%). HD_{95} and ASSD are measured in millimeters (mm). Best results are bold, and second best are underlined.

In the bone extraction stage, each CT volume was augmented to produce four variants. These images were created by applying random elastic distortions within a range of 80% to 120%, along with random translations and rotations within the ranges of -20 to 20 mm and -30 to 30 ° for each axis, respectively. The model was trained using the Adam optimizer with an initial learning rate of 0.0001 and a batch size of 2. Training followed a five-fold cross-validation protocol, and the loss function combined Dice loss with cross-entropy loss.

In the fragment segmentation stage, TPC-Net received eight augmented patches per bone crop-mirror flips on all three axes plus the same geometric and photometric transforms listed above. Each mini-batch contained two fractured and two healthy patches, enabling balanced contrastive sampling. The model was trained for 1,000 epochs using the Adam optimizer with a fixed learning rate of 0.0001. The InfoNCE temperature parameter was set to 0.07. A lightweight projection head was used to map features into 256-dimensional embedding vectors. The TPMA module employed four attention heads and a channel embedding dimension of 128. For each feature scale, the memory bank contained 1,000 trauma-specific memory keys. The SA operated over 3D windows of size $4 \times 4 \times 4$ voxels. During inference, voxels identified as part of the CFS were dilated by 5 mm to promote structural continuity.

Comparison to State-of-the-Art Methods

We evaluate FracSegmentator on three independent tasks: SacrumFrac, HipFrac, and FemurFrac. Each model is trained from scratch using the corresponding dataset and tested only within that task. We compare against eleven state-of-the-art segmentation models. These include (1) CNN-based models: 3D UNet, ResUNet, nnUNet (Çiçek et al. 2016; Diakogiannis et al. 2020; Isensee et al. 2021); (2) transformer-based models: TransUNet, swin-UNETR, MedNext (Chen et al. 2024; He et al. 2023; Roy et al. 2023); (3) state space sequence model-based method: U-Mamba (Ma, Li, and Wang 2024); and (4) task-specific methods for pelvic fractrue segmentation: FDM-UNet, FDD-Net (Liu et al. 2023; Zeng et al. 2024).

Figure 3 provides qualitative comparisons in both 3D and

Variant	IoU↑	DSC↑	HD_{95}↓	ASSD↓
FracSegmentator	**96.24**	**97.37**	**4.40**	**1.28**
w/o Feature Calibration	95.41	96.61	5.67	1.55
w/o Contrastive	95.20	96.48	5.82	1.55
w/o TPMA	92.60	94.32	7.65	2.07
w/o TPMA & Contrastive	91.16	93.06	8.64	2.26

Table 2: Module-level ablation results on the merged validation set.

2D slice views. Notably, for displaced fractures, all methods can successfully isolate the fragments. However, for non-displaced or incomplete fractures, some methods fail to separate the fragments clearly, showing confusion near the CFS. In cases where fragments collide, the proposed FracSegmentator maintains excellent consistency, while other methods show significant performance degradation near the collision zones. Mean inference time is approximately 3 s per case and scales with input volume size.

As shown in table 1, FracSegmentator outperforms all other methods across all metrics. FracSegmentator achieves the highest DSC of 97.59%, 96.78%, and 97.74% on SacrumFrac, HipFrac, and FemurFrac, outperforming the second best-performing MedNext by 2.99%, 1.91%, and 0.71%. These gains are statistically significant ($p < 0.05$ in paired t-test), underscoring the consistent advantage of our method across diverse anatomical challenges. To further assess generalizability, we evaluate FracSegmentator on the PENGWIN, which comprises a combination of hip and sacrum fractures. FracSegmentator achieves an IoU of 95.72%, HD_{95} of 4.79 mm, and ASSD of 1.52 mm, significantly outperforming the existing state-of-the-art method in PENGWIN, which reports 92.96%, 5.87 mm, and 1.84 mm, respectively. These results correspond to absolute improvements of 2.76%, 1.08 mm, and 0.32 mm.

Ablation Study

To better understand the contributions of individual components in FracSegmentator, we conducted comprehensive ablation studies focusing on module design, annotation strategy, and post-processing parameters. All experiments were performed on the merged validation set using consistent training and evaluation protocols.

Module-Level Ablation. As shown in table 2, removing any major component-Feature Calibration, Contrastive Learning, or the TPMA-leads to clear performance degradation, with the largest impact observed when TPMA is excluded. Notably, removing both TPMA and contrastive learning results in compounded performance loss, indicating their complementary effects.

Annotation Strategy Ablation. Table 3 shows that the BICM strategy yields the best results. When boundaries are merged with either interior or contact regions, performance drops significantly. This supports that explicit structural supervision is essential for fragment-level discrimination. The CFS, in particular, serves as a critical separator in non-displaced fractures.

Mask Strategy	IoU↑	DSC↑	HD_{95}↓	ASSD↓
BICM	**96.24**	**97.37**	**4.40**	**1.28**
Merge B+I	88.56	90.98	12.82	3.15
Merge B+C	88.27	90.73	11.40	2.77

Table 3: Annotation strategy ablation on the merged validation set.

Radius	IoU↑	DSC↑	HD_{95}↓	ASSD↓
1mm	95.33	96.58	5.72	1.54
3mm	95.88	97.12	4.94	1.37
5mm	**96.24**	**97.37**	**4.40**	**1.28**
7mm	96.05	97.20	4.83	1.36

Table 4: Effect of CFS dilation radius on final instance accuracy.

Post-Processing Parameter Selection

Table 4 evaluates the effect of varying the dilation radius for CFS propagation. A 5 mm radius achieves optimal accuracy, while smaller radii under-connect fragments and larger ones over-merge them. The optimal radius balances anatomical fidelity and topological separation, especially in regions with high fragment density.

Conclusion

We introduce FracSegmentator, a robust framework for anatomically precise fracture instance segmentation in CT scans. By first extracting bone structures and then isolating each fracture fragment using a trauma-aware contrastive network, our method effectively handles diverse fracture patterns. A key contribution of this work is the integration of trauma-informed priors through a contrastive encoder, which enhances the model's ability to capture subtle morphological variations often missed by conventional methods. In addition, we reformulate the segmentation task using a BICM representation. This provides fine-grained structural supervision and improves the separation of closely positioned fragments. We evaluated FracSegmentator on three anatomical sites: sacrum, hipbone, and femur. The results show consistent performance improvements over eleven state-of-the-art methods. Ablation studies further verify the effectiveness of each component, including trauma prior memory attention module, feature calibration attention decoder, and the proposed structural annotation strategy. In summary, FracSegmentator offers an accurate and generalizable solution for fracture instance segmentation. It demonstrates strong potential for clinical adoption and can be extended to other skeletal structures and trauma types.

In future work, we will extend FracSegmentator to additional anatomies and evaluate its generalization to related subtasks such as fracture detection and localization (Yang et al. 2025). We will also integrate it into clinical workflows to quantify its impact on surgical planning, fixation strategy selection, longitudinal monitoring, and overall decision making and outcomes.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant NSFC6247010104, in part by the National Key Research and Development Program of China under Grant 2022YFC2504304, in part by the Natural Science Foundation of Beijing under Grant L222136, in part by Beijing Science and Technology Project under Grant Z221100003522007 and Z241100009024030.

References

Bittner-Frank, M.; Strassl, A.; Unger, E.; Hirtler, L.; Eckhart, B.; Koenigshofer, M.; Stoegner, A.; Nia, A.; Popp, D.; Kainberger, F.; et al. 2024. Accuracy Analysis of 3D Bone Fracture Models: Effects of Computed Tomography (CT) Imaging and Image Segmentation. *Journal of Imaging Informatics in Medicine*, 1–13.

Chen, J.; Mei, J.; Li, X.; Lu, Y.; Yu, Q.; Wei, Q.; Luo, X.; Xie, Y.; Adeli, E.; Wang, Y.; et al. 2024. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97: 103280.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PmLR.

Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; and Ronneberger, O. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 424–432. Springer.

Diakogiannis, F. I.; Waldner, F.; Caccetta, P.; and Wu, C. 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162: 94–114.

Han, R.; Uneri, A.; Vijayan, R. C.; Wu, P.; Vagdargi, P.; Sheth, N.; Vogt, S.; Kleinszig, G.; Osgood, G.; and Siewerdsen, J. H. 2021. Fracture reduction planning and guidance in orthopaedic trauma surgery via multi-body image registration. *Medical image analysis*, 68: 101917.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

He, Y.; Nath, V.; Yang, D.; Tang, Y.; Myronenko, A.; and Xu, D. 2023. Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 416–426. Springer.

Isensee, F.; Jaeger, P. F.; Kohl, S. A.; Petersen, J.; and Maier-Hein, K. H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2): 203–211.

Kalinic, H. 2009. Atlas-based image segmentation: A Survey. *Croatian Scientific Bibliography*, 1–7.

Liang, C.; Wang, W.; Miao, J.; and Yang, Y. 2022. Gmmseg: Gaussian mixture based generative semantic segmentation models. *Advances in Neural Information Processing Systems*, 35: 31360–31375.

Liu, J.; Li, H.; Zeng, B.; Wang, H.; Kikinis, R.; Joskowicz, L.; and Chen, X. 2024. An end-to-end geometry-based pipeline for automatic preoperative surgical planning of pelvic fracture reduction and fixation. *IEEE Transactions on Medical Imaging*.

Liu, P.; Han, H.; Du, Y.; Zhu, H.; Li, Y.; Gu, F.; Xiao, H.; Li, J.; Zhao, C.; Xiao, L.; et al. 2021. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. *International Journal of Computer Assisted Radiology and Surgery*, 16: 749–756.

Liu, Y.; Yibulayimu, S.; Sang, Y.; Zhu, G.; Shi, C.; Liang, C.; Cao, Q.; Zhao, C.; Wu, X.; and Wang, Y. 2025. Preoperative fracture reduction planning for image-guided pelvic trauma surgery: A comprehensive pipeline with learning. *Medical Image Analysis*, 103506.

Liu, Y.; Yibulayimu, S.; Sang, Y.; Zhu, G.; Wang, Y.; Zhao, C.; and Wu, X. 2023. Pelvic fracture segmentation using a multi-scale distance-weighted neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 312–321. Springer.

Ma, J.; Li, F.; and Wang, B. 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*.

Neubauer, A.; Bühler, K.; Wegenkittl, R.; Rauchberger, A.; and Rieger, M. 2005. Advanced virtual corrective osteotomy. In *International congress series*, volume 1281, 684–689. Elsevier.

Parulekar, A.; Collins, L.; Shanmugam, K.; Mokhtari, A.; and Shakkottai, S. 2023. Infonce loss provably learns cluster-preserving representations. In *The Thirty Sixth Annual Conference on Learning Theory*, 1914–1961. PMLR.

Pettersson, J.; Knutsson, H.; and Borga, M. 2006. Non-rigid registration for automatic fracture segmentation. In *2006 International Conference on Image Processing*, 1185–1188. IEEE.

Roy, S.; Koehler, G.; Ulrich, C.; Baumgartner, M.; Petersen, J.; Isensee, F.; Jaeger, P. F.; and Maier-Hein, K. H. 2023. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 405–415. Springer.

Sang, Y.; Liu, Y.; Yibulayimu, S.; Wang, Y.; Killeen, B. D.; Liu, M.; Ku, P.-C.; Johannsen, O.; Gotkowski, K.; Zenk, M.; et al. 2025. Benchmark of Segmentation Techniques for Pelvic Fracture in CT and X-ray: Summary of the PENGWIN 2024 Challenge. *arXiv preprint arXiv:2504.02382*.

Tomazevic, M.; Kreuh, D.; Kristan, A.; Puketa, V.; and Cimerman, M. 2010. Preoperative planning program tool in treatment of articular fractures: process of segmentation procedure. In *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010: May 27–30, 2010 Chalkidiki, Greece*, 430–433. Springer.

Wasserthal, J.; Breit, H.-C.; Meyer, M. T.; Pradella, M.; Hinck, D.; Sauter, A. W.; Heye, T.; Boll, D. T.; Cyriac, J.; Yang, S.; et al. 2023. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5): e230024.

Wu, L.; Zhuang, J.; and Chen, H. 2024. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22873–22882.

Yang, J.; Shi, R.; Jin, L.; Huang, X.; Kuang, K.; Wei, D.; Gu, S.; Liu, J.; Liu, P.; Chai, Z.; et al. 2025. Deep rib fracture instance segmentation and classification from ct on the ribfrac challenge. *IEEE Transactions on Medical Imaging*.

You, X.; He, J.; Yang, J.; and Gu, Y. 2024. Learning with explicit shape priors for medical image segmentation. *IEEE Transactions on Medical Imaging*.

Zeng, B.; Wang, H.; Joskowicz, L.; and Chen, X. 2024. Fragment distance-guided dual-stream learning for automatic pelvic fracture segmentation. *Computerized Medical Imaging and Graphics*, 116: 102412.

Zeng, B.; Wang, H.; Xu, J.; Tu, P.; Joskowicz, L.; and Chen, X. 2023. Two-stage structure-focused contrastive learning for automatic identification and localization of complex pelvic fractures. *IEEE Transactions on Medical Imaging*, 42(9): 2751–2762.