Beyond Single View: Anatomy-Aware Rib Fracture Diagnosis with Multi-View Aggregation

Lei Luo[1], Tao Luo[1], Jiancheng Yang[2,3], Dijia Wu[4], and Zhiming Cui[1]✉

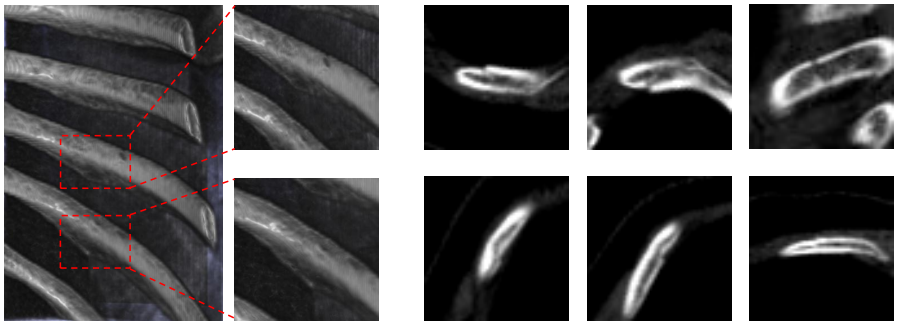
[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
cuizhm@shanghaitech.edu.cn
[2] ELLIS Institute Finland, Espoo, 02150, Finland
[3] Aalto University, Espoo, 02150, Finland
[4] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, 200232, China

Abstract. Rib fractures are prevalent traumatic injuries that may lead to severe complications if undetected. Precise detection and classification of fracture types on CT scans are therefore essential for timely clinical intervention. However, existing methods struggle with subtle fracture patterns and the intricate curvilinear anatomy of ribs, resulting in limited detection and classification accuracy. To address this, we propose **MVANet**, a two-stage **M**ulti-**V**iew **A**ggregation framework for rib fracture diagnosis. In the first stage, an anatomy-aware detection module localizes candidate fracture regions by leveraging rib masks as spatial priors, effectively suppressing background false positives. In the second stage, a trajectory-aligned multi-view sampler (TAMS) extracts geometry-aligned 2D projections along each candidate's rib centerline, which are encoded by a self-supervisedly adapted vision foundation model. A latent-query multi-view attention (LQMV-Attn) module further aggregates these multi-view features into a discriminative fracture-level representation for fine-grained classification. Experiments on our in-house MaskRib dataset and the public RibFrac dataset demonstrate that MVANet consistently outperforms existing methods in both detection sensitivity and classification accuracy, highlighting its potential for clinical translation. Code and models are publicly available at https://github.com/preciousluo/Beyond-Single-View.

Keywords: Rib Fracture Detection and Classification · Vision Foundation Model · Multi-view Aggregation

1 Introduction

Rib fractures are high-stakes injuries frequently complicated by life-threatening conditions such as pulmonary contusion and pneumothorax [4,17]. Accurate diagnosis requires not only detecting the presence of fractures but also classifying their morphological subtype, such as cortical twist, dislocation, or slight fracture, since treatment strategies differ substantially across categories. Although



(a) Standard 3D axis-aligned views (b) Trajectory-aligned views generated by TAMS

Fig. 1. Qualitative comparison: Geometry-aware slices in (b) unroll rib anatomy to reveal subtle fracture patterns obscured in (a).

computed tomography (CT) is the diagnostic gold standard, manual review of multi-slice CT volumes is labor-intensive, error-prone, and susceptible to diagnostic oversight [24]. These limitations motivate the development of an automated framework for rib fracture detection and fine-grained classification.

Existing deep learning approaches for rib fracture analysis broadly fall into two detection paradigms. Direct 3D volumetric methods preserve spatial context; for instance, FracNet [9] performs voxel-wise segmentation via a 3D UNet [18], while FracNet+ [23] integrates global point-cloud features to suppress false positives. Alternatively, 2D and projection-based methods such as Mask R-CNN [6] and YOLO [10] offer efficient localization [12]. Following detection, these pipelines typically feed the detected 3D regions into a cascaded classifier for multi-class categorization. Recent classification efforts have addressed complementary aspects, including temporal healing states [1] and hierarchical severity grading via hyperbolic embeddings [16], yet these focus on high-level diagnostic states rather than fine-grained morphological differentiation. A central limitation shared by most existing classifiers is their reliance on rigid 3D bounding-box crops, which, as illustrated in Fig. 1 (a), inevitably encapsulate excessive irrelevant tissue and dilute the subtle cortical features essential for distinguishing fracture subtypes. Moreover, directly operating in the 3D volumetric domain precludes these methods from leveraging vision foundation models (e.g., DINOv2 [15]) pretrained on large-scale natural images, which have demonstrated powerful feature extraction capabilities across diverse visual recognition tasks.

In this paper, we propose MVANet, a two-stage framework that bridges global volumetric screening with geometry-aware multi-view classification. The first stage performs anatomy-aware detection. By incorporating rib segmentation masks as spatial priors, the detector confines its search to skeletal regions, substantially reducing false positives from surrounding soft tissue. The second stage addresses the fundamental limitation of rigid 3D crops by introducing

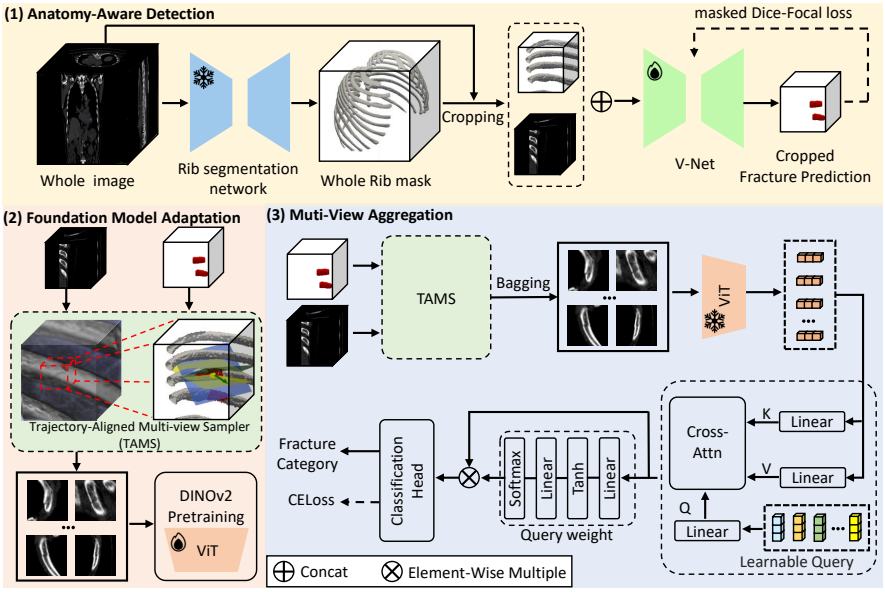


Fig. 2. Overview of MVANet. (1) Anatomy-aware detection localizes fracture candidates within rib-mask-constrained regions. (2) TAMS extracts trajectory-aligned multi-view slices, which are encoded by a self-supervisedly adapted DINOv2 backbone; (3) LQMV-Attn then aggregates multi-view features for fine-grained classification.

the trajectory-aligned multi-view sampler (TAMS), which extracts 2D oblique slices along each candidate's curvilinear rib centerline. This trajectory-aligned sampling effectively unrolls the rib cortex, exposing subtle discontinuities that axis-aligned views miss (Fig. 1 (b)). To bridge the modality gap between natural images and CT, we self-supervisedly adapt a vision foundation model (DINOv2 [15]) to encode these projections into rich feature representations. Finally, recognizing that fracture evidence is unevenly distributed across viewing angles, we introduce the latent-query multi-view attention (LQMV-Attn) module, which uses learnable queries to selectively attend to the most diagnostically informative views rather than relying on naive pooling. Extensive experiments on both an in-house dataset and the public RibFrac benchmark demonstrate that MVANet significantly outperforms existing methods in detection sensitivity and classification accuracy.

2 Method

The overall pipeline of MVANet is illustrated in Fig. 2. Given a CT volume $\mathbf{X} \in \mathbb{R}^{H\times W\times D}$ and an anatomical rib mask $\mathbf{m} \in \{0,1\}^{H\times W\times D}$, our goal is to detect all fracture locations and classify each into its morphological subtype. The rib mask $\mathbf{m}$ can be obtained from dataset annotations or, for fully automated

deployment, generated by an off-the-shelf rib segmentation network [22], as rib segmentation is a well-established task with readily available pretrained models. Ground-truth annotations include a voxel-wise fracture mask $\mathbf{y} \in \{0,1\}^{H\times W\times D}$ for localization and instance-level category labels $\mathcal{C}$ for classification. MVANet decomposes the task into two sequential stages: anatomy-aware 3D detection followed by multi-view classification.

2.1 Anatomy-Aware Detection

The detection stage leverages anatomical priors to constrain the search space and suppress background false positives. Specifically, a 3D V-Net receives as input the concatenation of the CT volume $\mathbf{X}$ and the rib mask $\mathbf{m}$, forming a dual-channel input $[\mathbf{X};\mathbf{m}] \in \mathbb{R}^{H\times W\times D\times 2}$. This explicit spatial priors restricts learning to skeletal regions, substantially reducing the overwhelming class imbalance between fracture voxels and background. The V-Net encoder extracts hierarchical features through residual convolutional blocks with progressive downsampling, while the decoder recovers spatial resolution via skip connections and transposed convolutions to produce a dense fracture probability map $\hat{\mathbf{y}} \in [0,1]^{H\times W\times D}$.

To further address the extreme sparsity of fractures, we optimize the network using a masked Dice-Focal loss. The rib mask $\mathbf{m}$ restricts both the Dice and Focal loss [19,11] computations to clinically relevant skeletal regions:

$$\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{Dice}}(\hat{\mathbf{y}} \odot \mathbf{m}, \mathbf{y} \odot \mathbf{m}) + \lambda\, \mathcal{L}_{\text{Focal}}(\hat{\mathbf{y}} \odot \mathbf{m}, \mathbf{y} \odot \mathbf{m}), \tag{1}$$

where $\odot$ denotes element-wise multiplication and λ balances the two loss terms. This masking strategy ensures that gradient signals are concentrated on diagnostically meaningful voxels rather than being diluted by the vast background. During inference, the predicted probability map is binarized with a threshold of 0.5, and connected-component analysis on the thresholded output yields a set of candidate fracture regions for subsequent classification.

2.2 Foundation Model Adaptation via Trajectory-Aligned Sampling

Following detection, each candidate region requires fine-grained classification. As discussed in Sec. 1, conventional 3D bounding-box crops are ill-suited because the curvilinear geometry of ribs causes axis-aligned volumes to include excessive irrelevant tissue. We address this with a two-step strategy: (1) extracting geometry-aligned 2D projections via trajectory-aligned sampling, and (2) encoding these projections with a domain-adapted vision foundation model.

Trajectory-Aligned Multi-View Sampler (TAMS). The key insight behind TAMS is that fracture-discriminative features, such as cortical discontinuities, angular deformations, and callus formation, are best visualized on cross-sectional planes oriented perpendicular to the local rib trajectory, rather than along arbitrary global axes. As shown in Fig. 2 (2), for each detected candidate, we first extract the local rib centerline $\mathcal{T}$ via morphological skeletonization of the rib mask. Let $\mathbf{p}_c \in \mathbb{R}^3$ denote the candidate's centroid and $\mathbf{t} \in \mathbb{R}^3$ the unit tangent

vector along $\mathcal{T}$ at $\mathbf{p}_c$. We initialize a unit normal vector $\mathbf{n} \perp \mathbf{t}$ and systematically rotate it around $\mathbf{t}$ in increments of 9° over $[0^\circ, 180^\circ)$, generating $K = 20$ multi-view oblique slices of 56×56 pixels. The k-th slice $\mathbf{S}_k(u, v)$ is sampled via:

$$\mathbf{S}_k(u, v) = \mathbf{X}\Big(\mathbf{p}_c + u \cdot \mathbf{t} + v \cdot \mathbf{n}_k\Big), \tag{2}$$

where $\mathbf{n}_k = \mathbf{R}(\mathbf{t}, \theta_k)\,\mathbf{n}$ is the rotated normal and $\mathbf{R}(\mathbf{t}, \theta_k) \in \mathbb{R}^{3\times3}$ denotes the rotation matrix about $\mathbf{t}$ by angle θ_k. Trilinear interpolation ensures smooth subvoxel sampling during slice extraction. By sampling along this trajectory-aligned coordinate system, TAMS unrolls the curvilinear cortical surface, maximizing the conspicuity of subtle fractures that overlapping tissues would otherwise obscure.

Domain Adaptation of Vision Foundation Model. Despite offering powerful representations, vision foundation models exhibit a substantial modality gap between natural photographs and CT bone textures [13]. Directly applying such models to medical images yields suboptimal performance due to differences in intensity distributions, texture statistics, and semantic content. To bridge this gap, we adopt DINOv2 [15] (ViT-B/14) as our feature encoder and self-supervisedly fine-tune all layers on the TAMS-extracted multi-view slices from our training data. During adaptation, we employ the original DINO self-distillation objective, in which a student network learns to match the output distribution of a momentum-updated teacher network across augmented views of the same slice. Each single-channel CT slice is replicated to three channels for compatibility with the pretrained architecture. This adaptation calibrates the pretrained visual features to CT-specific bone textures and fracture morphologies, producing embeddings $\mathbf{z}^{(k)} \in \mathbb{R}^d$ that are sensitive to fine-grained anatomical nuances while retaining the generalization benefits of large-scale pretraining.

2.3 Multi-View Aggregation

In clinical practice, radiologists confirm fractures by synthesizing visual evidence across multiple imaging planes. We formalize this multi-planar reasoning within a multiple-instance learning (MIL) [8,5] framework: the K view-specific embeddings for a candidate are treated as an instance bag $\mathbf{Z} = [\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(K)}]^\top \in \mathbb{R}^{K\times d}$. Since fracture evidence is unevenly distributed across viewing angles, with some views clearly revealing a cortical break while others show only normal anatomy, naive pooling (mean or max) dilutes diagnostically critical signals. This motivates a learned, attention-based aggregation mechanism.

Latent-Query Multi-View Attention (LQMV-Attn). As illustrated in Fig. 2 (3), the LQMV-Attn module introduces M learnable latent queries $\mathbf{Q} \in \mathbb{R}^{M\times d}$, each designed to capture a distinct morphological pattern (e.g., cortical disruption, angular deformation). These queries attend to the multi-view feature bag via scaled dot-product cross-attention:

$$\mathbf{H} = \mathrm{Softmax}\left(\frac{\mathbf{Q}\cdot\mathbf{K}^\top}{\sqrt{d}}\right)\cdot\mathbf{V}, \tag{3}$$

where $\mathbf{K} = \mathbf{Z}\mathbf{W}_K$ and $\mathbf{V} = \mathbf{Z}\mathbf{W}_V$ are linear projections of $\mathbf{Z}$ with learnable weight matrices $\mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d\times d}$, and $\mathbf{H} \in \mathbb{R}^{M\times d}$ encapsulates M morphological representations aggregated from the bag. By using M independent queries rather than a single aggregation vector, the module can simultaneously probe for multiple fracture-related patterns across different views.

Since not all latent responses contribute equally to diagnosis, a gating mechanism dynamically evaluates their importance. Specifically, $\mathbf{H}$ is projected into a normalized importance distribution $\mathbf{a} \in \mathbb{R}^M$ via a bottleneck MLP:

$$\mathbf{a} = \mathrm{Softmax}\big(\tanh(\mathbf{H}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{w}_2 + b_2\big), \tag{4}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d\times d'}$, $\mathbf{w}_2 \in \mathbb{R}^{d'}$ are learnable weight matrices and $\mathbf{b}_1 \in \mathbb{R}^{d'}$, $b_2 \in \mathbb{R}$ are bias terms. The bottleneck dimension d' is set to $\frac{d}{2}$ to encourage compact importance estimation. The final fracture-level representation is obtained by weighted aggregation:

$$\mathbf{g} = \mathbf{H}^\top \mathbf{a}, \tag{5}$$

the resulting vector $\mathbf{g} \in \mathbb{R}^d$ is passed to a linear classification head that outputs logits over the fracture categories. The classification stage is trained end-to-end (excluding the frozen DINOv2 backbone) with cross-entropy loss. By selectively attending to views that reveal diagnostically salient cortical features and suppressing uninformative background projections, LQMV-Attn mirrors the multiplanar reasoning of expert radiologists.

3 Experiments

3.1 Dataset and Evaluation Metrics

MaskRib is an in-house dataset comprising 951 paired samples, each containing a chest CT volume, a rib segmentation mask, and voxel-wise fracture annotations. Fractures are classified into four clinically significant categories: callus (CA), cortical twist (TW), dislocation fracture (DF), and slight fracture (SL).
RibFrac [23] is a public benchmark consisting of 420 training, 80 validation, and 160 test scans. It defines four classification categories: displaced (DP), non-displaced (ND), buckle (BK), and segmental (SG) rib fractures. An additional ambiguous category is excluded from classification evaluation.
Evaluation Metrics. For detection, we employ Free Response Receiver Operating Characteristic (FROC) analysis. A detection is counted as a true positive (TP) if its 3D IoU with a ground-truth annotation satisfies IoU ≥ 0.1; unmatched predictions and missed lesions are counted as false positives (FP) and false negatives (FN), respectively. The primary detection metric is the mean sensitivity at 4, 8, 16, and 32 FP per scan, estimated via linear interpolation; we also report average FP per scan (Avg FP). For classification, we adopt Target-Aware F1 as the primary metric, computed exclusively on ground-truth regions (both correctly localized TP and missed FN) to decouple classification performance from detection quality.

Table 1. Comprehensive performance of the proposed MVANet. We report FROC and Avg FP in the detection stage and classwise F1 score in the classification stage.

Dataset	Method	Stage 1: Detection		Stage 2: Classification (F1 %)				
		FROC↑	Avg. FP↓	C1	C2	C3	C4	Avg.↑
MaskRib	DCC [2]	60.45	76.38	74.04	48.13	44.29	52.62	54.77
	UCI [21]	67.75	76.19	74.00	46.07	53.95	45.89	54.98
	MVANet	**84.56**	**39.74**	**81.75**	**55.23**	**80.88**	**68.01**	**71.47**
RibFrac [23]	DCC [2]	71.86	85.56	62.92	47.71	48.57	17.14	44.09
	UCI [21]	81.76	41.24	56.18	52.89	56.25	0.00	41.33
	MVANet	**86.40**	**32.41**	**67.76**	**58.73**	**66.67**	**26.42**	**54.89**

Note: C1–C4 denote CA, TW, DF, SL on MaskRib; and DP, ND, BK, SG on RibFrac.

3.2 Implementation Details

CT volumes are windowed to $[-50, 500]$ HU and resampled to $0.68 \times 0.68 \times 1.0$ mm^3. For detection, we sample 128^3 patches with a $1:3$ positive-to-negative ratio, confining negative samples within the rib masks. The V-Net is optimized with AdamW ($\mathrm{lr} = 10^{-3}$, batch size 16). During inference, sliding-window evaluation with 25% overlap is applied under rib-mask constraints, with a binarization threshold of 0.5. For classification, TAMS extracts $K = 20$ slices of 56×56 pixels at 9° intervals. We adopt the ViT-B/14 variant of DINOv2 ($d = 768$) with a model size of 342.2 MB as the fixed feature encoder, which is frozen after self-supervised adaptation. The classification head, with a model size of approximately 9.5 MB, is fine-tuned for 50 epochs using AdamW ($\mathrm{lr} = 5 \times 10^{-5}$, batch size 128). For a whole CT volume, detection requires 11.47 s per case, while each detected fracture candidate incurs 0.7 s for multi-view sampling and only 0.002 ms for classification.

3.3 Quantitative Results

Recent studies prioritize clinical status or severity grading over fine-grained morphological discrimination. As the sole public benchmark providing detailed annotations, the MICCAI RibFrac challenge serves as the standard reference. Consequently, we compare MVANet against two established end-to-end pipelines: DCC [2], which pairs Mask R-CNN detection with a NoduleNet [20] classifier, and UCI [21], which combines a 3D UNet detector with a pretrained 3D Medical ResNet [3,7].

As reported in Table 1, MVANet achieves substantial improvements in both stages across both datasets. In detection, MVANet attains 84.56% FROC on MaskRib and 86.40% on RibFrac, outperforming the best baseline by 16.81 and 4.64 percentage points, respectively. Notably, false positives are reduced by nearly half, from 76.19 to 39.74 Avg FP on MaskRib, confirming that anatomy-aware screening effectively suppresses non-rib candidates that lack structural priors. In classification, MVANet achieves an average F1 of 71.47% on MaskRib

Table 2. Ablation study of the MVANet. In the detection stage, we report FROC values at average 4, 8, 16, and 32 FP per scan. In the classification stage, we report F1 score (%) with varying sampling view number K.

Dataset	Method	Detection Ablation					
		4↑	8↑	16↑	32↑	Max Sens.↑	Avg. FP↓
MaskRib	V-Net [14]	56.02	68.16	72.76	74.77	75.17	48.31
	MVANet	**78.38**	**82.81**	**87.06**	**89.98**	**90.38**	**39.74**
RibFrac [23]	V-Net [14]	76.19	83.72	86.18	85.28	87.19	39.05
	MVANet	**83.10**	**85.74**	**87.57**	**89.17**	**89.24**	**32.41**
		Classification Ablation					
		$K = 1$	$K = 5$	$K = 10$	$K = 20$	$K = 25$	$K = 30$
MaskRib		65.86	71.13	70.91	**71.47**	70.20	70.13
RibFrac [23]		48.63	52.77	53.58	**54.89**	54.72	54.59

and 54.89% on RibFrac, surpassing the strongest baseline by 16.49 and 10.80 points, respectively. The gains are most pronounced for morphologically challenging subtypes: on MaskRib, dislocation fracture (C3) F1 improves from 53.95% to 80.88%, while on RibFrac, MVANet recovers a 26.42% F1 for the segmental rib fracture (C4), where the UCI baseline scores 0.00%. These results demonstrate that trajectory-aligned multi-view sampling, combined with attention-based aggregation, captures the geometric features that conventional 3D-crop classifiers miss, particularly for rare and subtle fracture patterns.

3.4 Ablation Study

Table 2 validates the contribution of each component.

Detection stage. Incorporating rib-mask-guided anatomy priors into the V-Net yields consistent FROC gains across all FP thresholds. On MaskRib, maximum sensitivity improves by 15.21 percentage points, from 75.17% to 90.38%, with a simultaneous reduction in Avg FP from 48.31 to 39.74. This confirms that explicit anatomical constraints focus the detector on skeletal regions, reducing false alarms from visually similar soft-tissue structures.

Classification stage. Performance improves steadily as the number of sampled views K increases, reflecting the benefit of richer geometric context. However, performance peaks at $K = 20$ and slightly degrades beyond this point, from 71.47% to 70.13% at $K = 30$ on MaskRib. We attribute this to the introduction of redundant or highly overlapping slices that dilute discriminative features through geometric noise. Accordingly, $K = 20$ is selected as the optimal trade-off between spatial coverage and feature quality.

4 Conclusion

We presented MVANet, a two-stage framework that bridges global volumetric screening with multi-view classification for rib fracture diagnosis. In the detection stage, rib-mask-guided spatial priors reduce false positives by nearly half compared to baselines. In the classification stage, TAMS unrolls curvilinear rib anatomy into geometry-aligned cross-sections, a vision foundation model encodes fine-grained bone features, and LQMV-Attn selectively fuses the most diagnostically informative views. Evaluations on both in-house and public datasets demonstrate consistent improvements in detection sensitivity and classification F1, confirming the clinical potential of combining anatomical priors with foundation-model-based multi-view reasoning.

Acknowledgments. This work was supported by Capital's Funds for Health Improvement and Research under Grant No. CFH2024-1-4101, the National Natural Science Foundation of China (NSFC) under Grant Nos. 6230012077 and 62531024, Shanghai Municipal Central Guided Local Science and Technology Development Fund Project under Grant No. YDZXX20233100001001, and Beijing Natural Science Foundation under Grant No. L242133. The authors also acknowledge the HPC Platform of ShanghaiTech University for providing computational resources.

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

References

1. Castro-Zunti, R., Li, K., Vardhan, A., Choi, Y., Jin, G.Y., bum Ko, S.: Ribfracturesys: A gem in the face of acute rib fracture diagnoses. Computerized Medical Imaging and Graphics **117**, 102429 (2024). https://doi.org/https://doi.org/10.1016/j.compmedimag.2024.102429, https://www.sciencedirect.com/science/article/pii/S089561112400106X
2. Chai, Z., Xiao, Y., Chen, H.: Deep cascaded learning for automated rib fracture detection and subtype classification. MICCAI RibFrac Challenge Technical Report (2020), available at https://ribfrac.grand-challenge.org/
3. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
4. Franssen, A.J., Daemen, J.H., Luyten, J.A., Meesters, B., Pijnenburg, A.M., Reisinger, K.W., van Vugt, R., Hulsewé, K.W., Vissers, Y.L., de Loos, E.R.: Treatment of traumatic rib fractures: an overview of current evidence and future perspectives. Journal of Thoracic Disease **16**(8), 5399–5408 (2024)
5. Han, Z., Wei, B., Hong, Y., Li, T., Cong, J., Zhu, X., Wei, H., Zhang, W.: Accurate screening of covid-19 using attention-based deep 3d multiple instance learning. IEEE Transactions on Medical Imaging **39**(8), 2584–2594 (2020). https://doi.org/10.1109/TMI.2020.2996256
6. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
7. 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)

8. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
9. Jin, L., Yang, J., Kuang, K., Ni, B., Gao, Y., Sun, Y., Gao, P., Ma, W., Tan, M., Kang, H., Chen, J., Li, M.: Deep-learning-assisted detection and segmentation of rib fractures from ct scans: Development and validation of fracnet. eBioMedicine (2020)
10. van Leeuwen, M.P., Ralf van Oudheusden, T., Oei, S.P., Pieter Theodoor Bosma, G., Haak, K.V., Saygili, G., Ong, L.L.S.: Automatic identification of anatomical locations for bone abnormalities in ct imaging: A multiplanar yolov5 detection approach. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1–7 (2025). https://doi.org/10.1109/EMBC58623.2025.11254481
11. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
12. Lopez-Melia, M., Magnin, V., Schranz, S., Andrearczyk, V., Depeursinge, A., Marchand-Maillet, S., Grabherr, S.: Automatic rib fracture detection on post-mortem ct data using deep learning. International Journal of Legal Medicine (2025). https://doi.org/10.1007/s00414-025-03669-x
13. Luo, T., Wu, H., Yang, T., Shen, D., Cui, Z.: Adapting foundation model for dental caries detection with dual-view co-training. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 46–55. Springer Nature Switzerland, Cham (2026)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
15. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. Transactions on Machine Learning Research Journal (2024)
16. Pate, S., Farooq, A., Datta, S., Sheikh, M.A., Kumar, A., Mishra, D.: Fine-grained rib fracture diagnosis with hyperbolic embeddings: A detailed annotation framework and multi-label classification model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 218–227. Springer (2025)
17. Peek, J., Beks, R.B., Hietbrink, F., De Jong, M.B., Heng, M., Beeres, F.J., IJpma, F.F., Leenen, L.P., Groenwold, R.H., Houwert, R.M.: Epidemiology and outcome of rib fractures: a nationwide study in the netherlands. European journal of trauma and emergency surgery **48**(1), 265–271 (2022)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
19. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 240–248. Springer (2017). https://doi.org/10.1007/978-3-319-67558-9_28, https://doi.org/10.1007/978-3-319-67558-9_28
20. Tang, H., Zhang, C., Xie, X.: Nodulenet: Decoupled false positive reductionfor pulmonary nodule detection and segmentation (2019)

21. Yan, X., Tang, H., Xiang, Y., Sun, Q., Zan, J., Naushad, J., Xie, X.: Solution to miccai 2020 ribfrac challenge. MICCAI RibFrac Challenge Technical Report (2020), available at `https://ribfrac.grand-challenge.org/`
22. Yang, J., Gu, S., Wei, D., Pfister, H., Ni, B.: Ribseg dataset and strong point cloud baselines for rib segmentation from ct scans. In: International conference on medical image computing and computer-assisted intervention. pp. 611–621. Springer (2021)
23. Yang, J., Shi, R., Jin, L., Huang, X., Kuang, K., Wei, D., Gu, S., Liu, J., Liu, P., Chai, Z., Xiao, Y., Chen, H., Xu, L., Du, B., Yan, X., Tang, H., Alessio, A., Holste, G., Zhang, J., Wang, X., He, J., Che, L., Pfister, H., Li, M., Ni, B.: Deep rib fracture instance segmentation and classification from ct on the ribfrac challenge. IEEE Transactions on Medical Imaging (2025)
24. Yao, L., Guan, X., Song, X., Tan, Y., Wang, C., Jin, C., Chen, M., Wang, H., Zhang, M.: Rib fracture detection system based on deep learning. Scientific reports **11**(1), 23513 (2021)