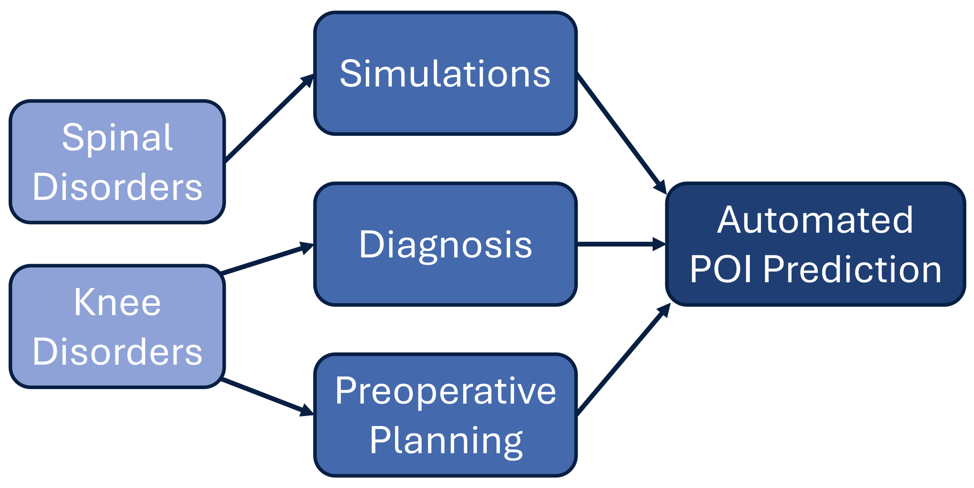
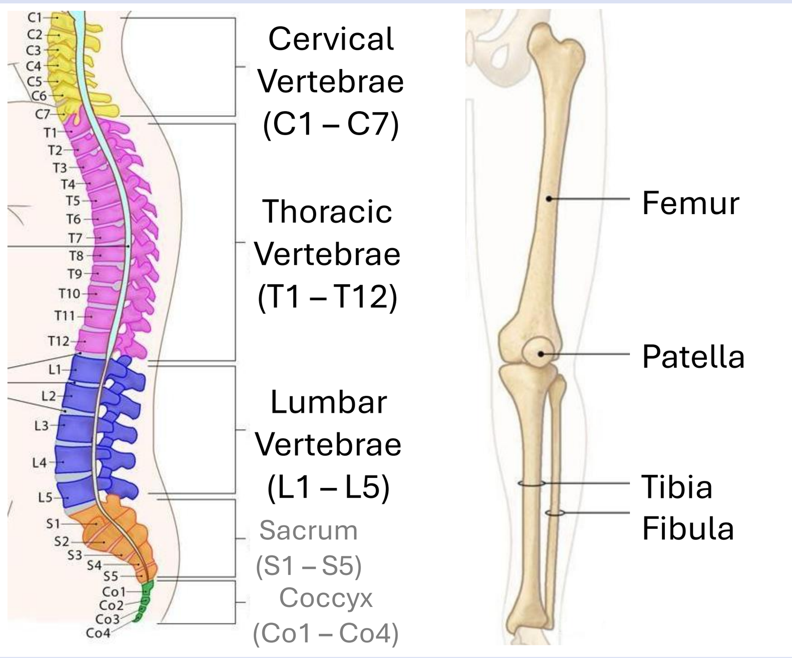
Deep Learning-Based Prediction of Anatomical Points-of-Interest from Sparse Annotations on Segmentation Masks

Alissa Yuxuan Wang

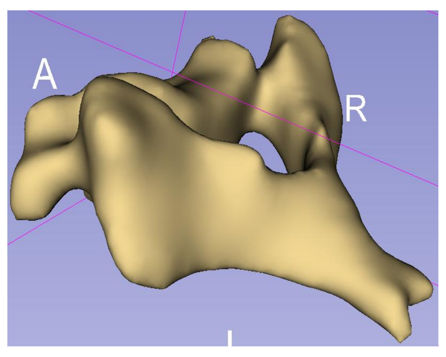
1 Motivation



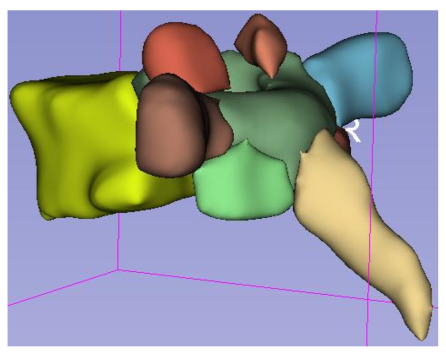


2 Datasets and Data Preprocessing

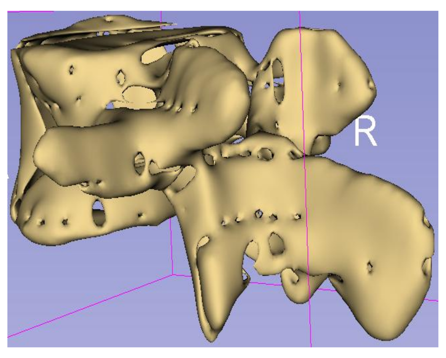
Vertebral Landmark Dataset



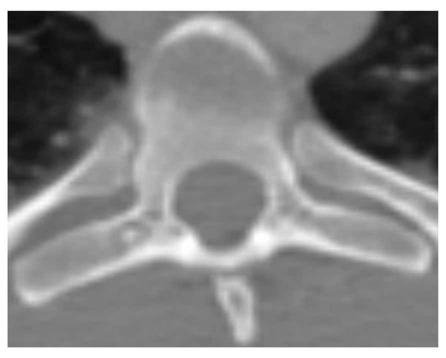
Instance Segmentation labels each vertebra on the spine.



Semantic Segmentation labels each subregion of the vertebra.



Surface Mask labels the surface of each vertebra.



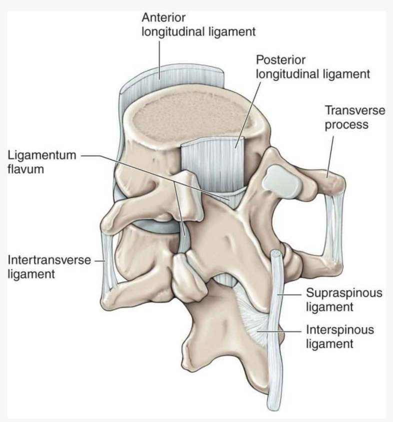
CT Scan 3D grayscaled values.

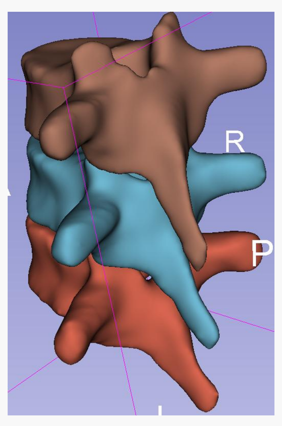
Neighbor

- POIs correspond to ligament attachment points
- POI positions depend on neighboring vertebrae → when connected a smooth line should be formed

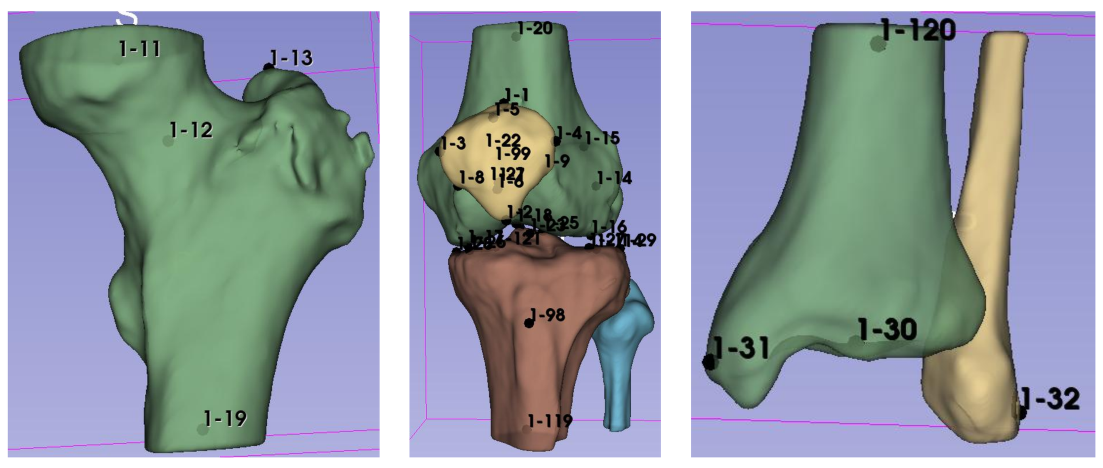
Idea:

- Load each vertebra along with its neighbors, so the model can learn inter-vertebral relationships





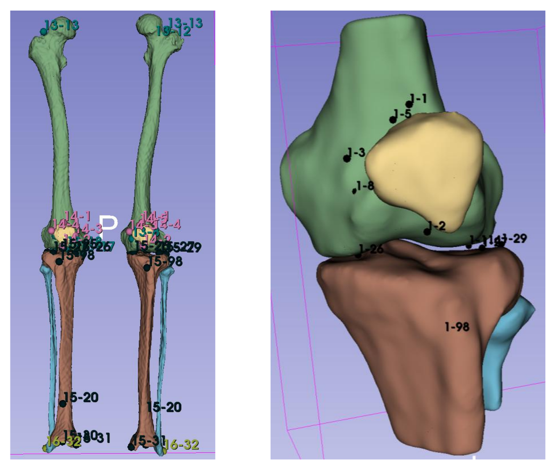
Lower Limb Dataset



- Leg is divided into three FOVs to handle large input size
- Femoral- and lower leg shaft excluded (no POIs present)
- Separate dataloaders per FOV
- Independent training due to anatomical differences and POI distributions

3 Data Augmentation – Elastic Deformation

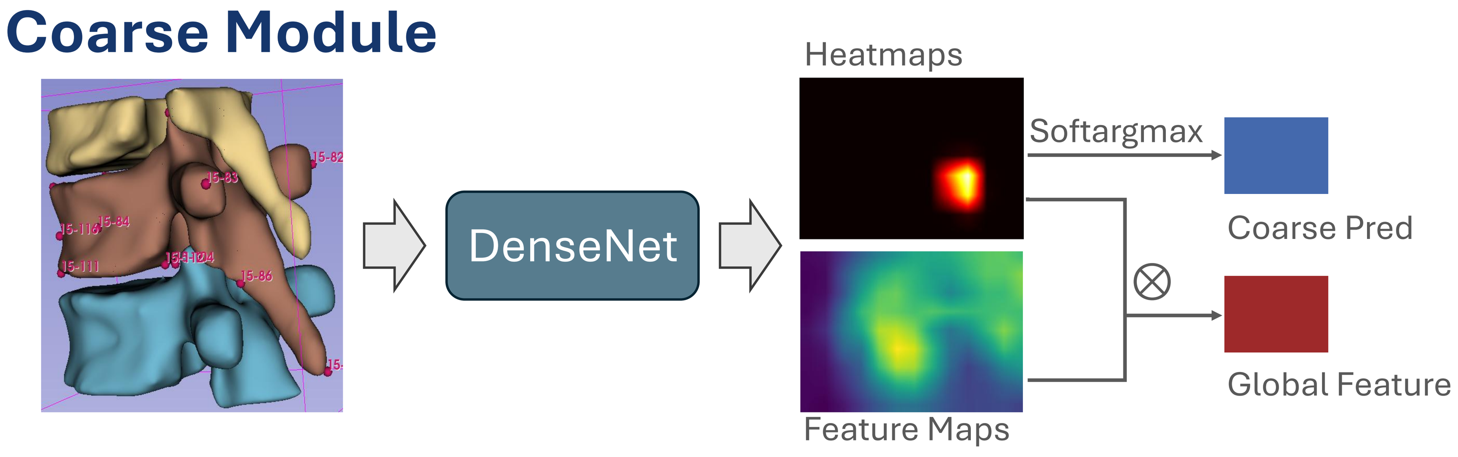
- Need for artificial data due to data scarcity
- apply random displacement fields to deform data
- 10 deformations for each of the 13 subjects were generated (with different degrees of deformation)
- Interpolation Issue: segmentation masks (order 0) vs POI-coordinates (order 3) → some POIs are shifted off the surface
- Cleaning Steps: exclude distant surface POIs



4 Model Architecture

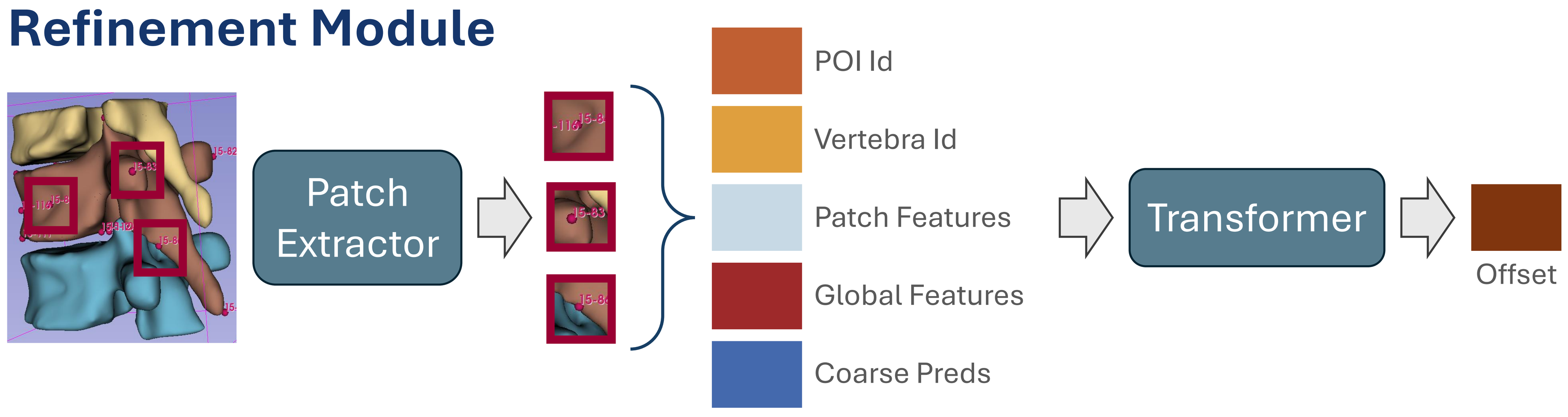
- Two-step process:
  - Initial **coarse predictions** on lower resolution using heatmap regression
  - **Refine predictions** using attention and contextual relationships

Coarse Module



- generates initial coordinate estimates of each landmark using a **DenseNet** backbone
- Extracts global features and produces heatmaps for each POI to capture spatial information

Refinement Module

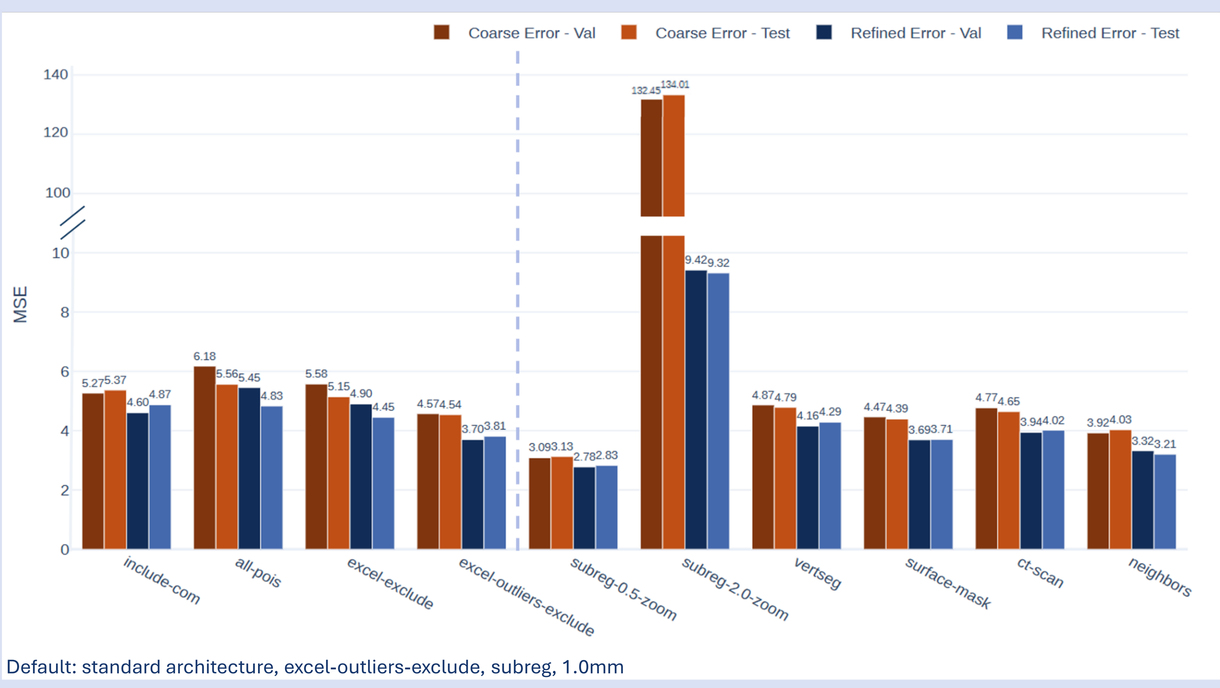


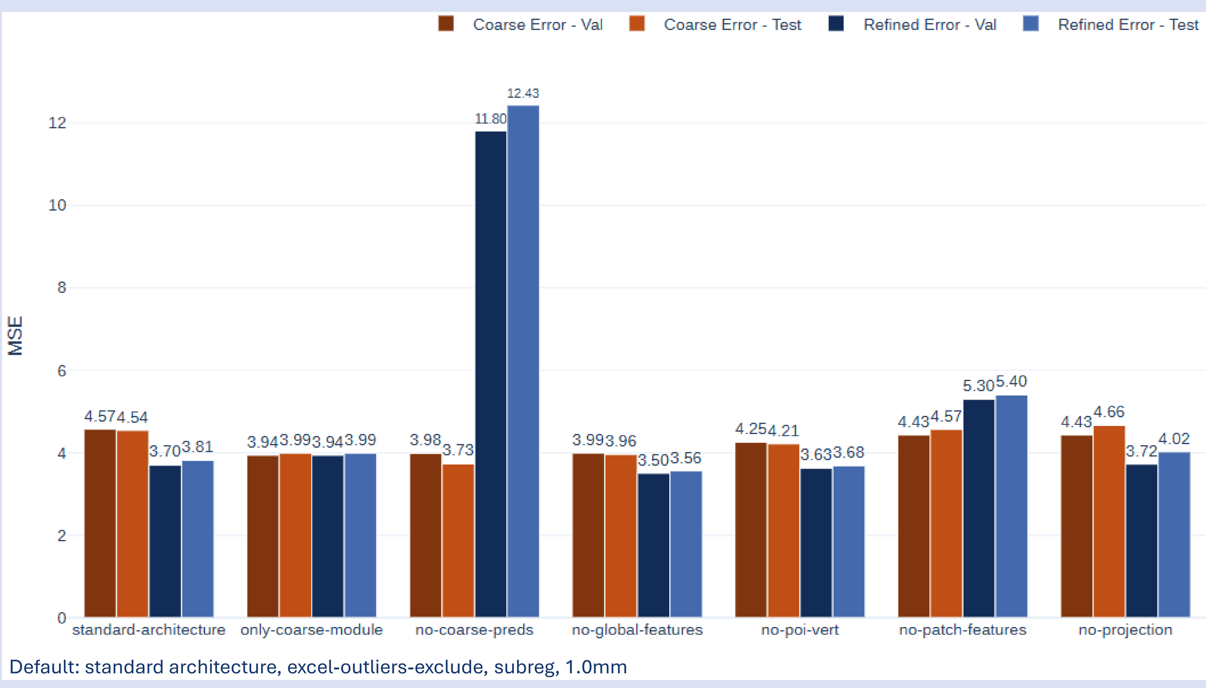
- Extracts patches around estimated landmarks, producing high-resolution patch features that capture local appearances and spatial context
- Improves initial coarse predictions using a **transformer-based** model
- Integrates spatial information with semantic features
- Applies self-attention mechanisms to model contextual relationships and inter-landmark dependencies

Experiments and Results

Ablation Study on Vertebral Dataset

An ablation study was performed to analyze the contributions of individual components to the overall performance.





Dataloader: effect of input data on performance

- POI selection: data cleaning, incl. contextual POIs
- Input type and scaling: different segmentation mask types

> - Data Cleansing → higher performance
> - Neighboring vertebrae → improved accuracy
> - Low resolution data → poor coarse performance

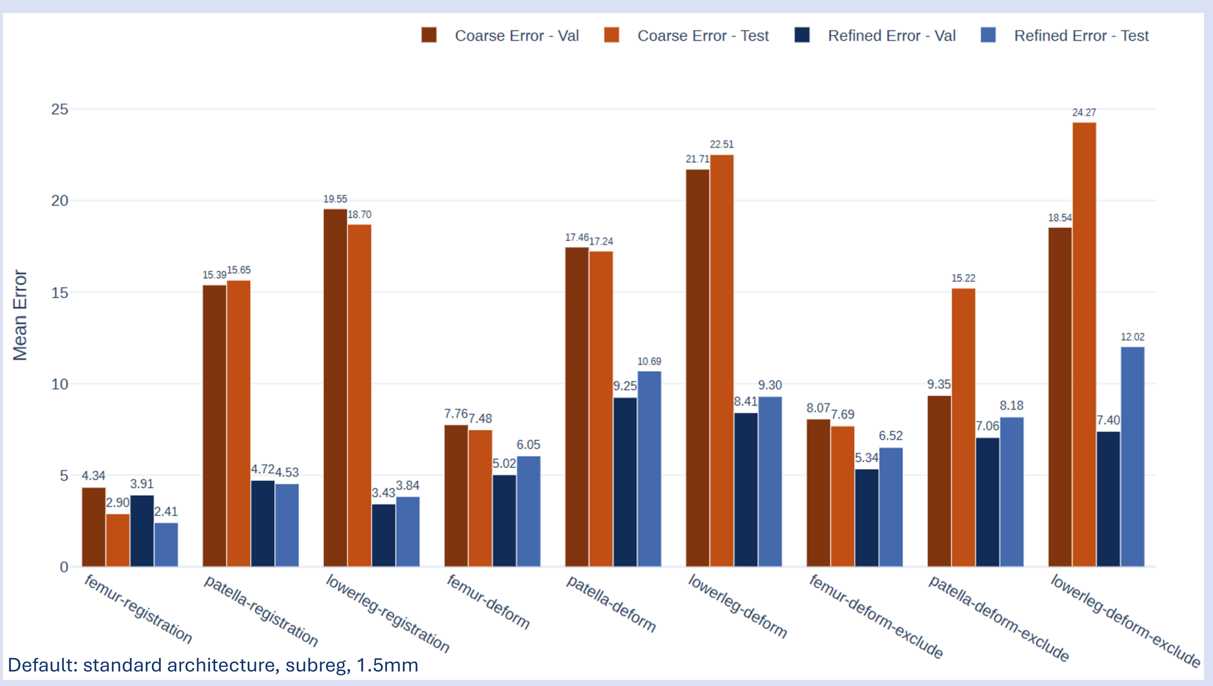
Architecture: effect of individual architectural components

- Analyze which inputs of the refinement module contribute to a better performance

> - Refinement module → increases performance
> - Patch extractor → key component for refinement performance; other inputs show little benefits

POI Prediction on the Lower Limbs

The architecture was applied to predict POIs on the lower limb.



Registration: align image to a different coordinate system (236 subjects, 1.5mm scale)

Elastic deformation (133 subjects, 1.5mm scale)

> - Refinement module → large accuracy gain
> - Registration outperforms elastic deformation *(note: more registration data available)*

Conclusion

- Architecture is easily adaptable to other anatomical regions
- Refinement module produces great performance gains → especially for lower limb POI prediction
- Data quality is crucial for performance improvements
  - Elastic deformation insufficient → better data cleaning/ correction is required
  - Registration-based augmentation → more robust, but still impacted by annotation errors
- Comparable performance to state-of-the-art on vertebral dataset