# Prior Guided 3D Medical Image Landmark Localization


***Yijie Pang[A], Pujin Cheng[A], Junyan Lyu[AC], Fan Lin[B] and Xiaoying Tang[A]***

Southern University of Science and Technology[A], Shenzhen Second People's Hospital[B], Shenzhen, China.

Queensland Brain Institute, The University of Queensland[C], St Lucia, QLD, Australia



## Abstract

Accurate detection of 3D landmarks is critical for evaluating and characterizing anatomical features and performing preoperative diagnostic screening. Detecting 3D landmarks can be challenging due to the local structural homogeneity of medical images. Physicians often annotate multiple landmarks in the key slice to address this issue, particularly when estimating 3D distance or volume.

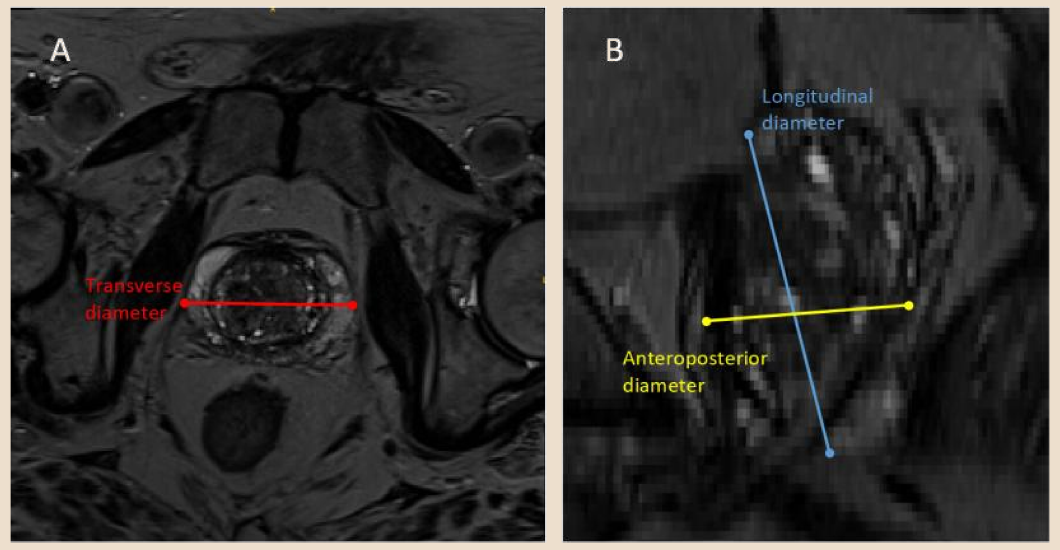
In this study, we present a prior guided coarse-to-fine framework for efficient and accurate 3D medical landmark detection; We utilize prior knowledge that in specific settings, physicians annotate multiple landmarks on the same slice. The coarse stage uses coordinate regression on downsampled 3D images . The fine stage categorizes landmarks as independent and correlated landmarks based on their annotation prior. Our method is extensively evaluated on two datasets, exhibiting superior performance with an average detection error of 3.29 mm and 2.13 mm, respectively.


## Introduction

3D landmark detection methods have advanced from key slice classification to coordinate and heatmap regression with 3D CNNs. Typically, 3D medical images are downsampled to of relatively low resolution to reduce network parameters which may cause inevitable errors. Coarse-to-fine strategies may relieve the accuracy degradation issue caused by downsampling. Although most previous methods delved into the model structure, the loss function, etc, they ignored the prior knowledge of medical annotation

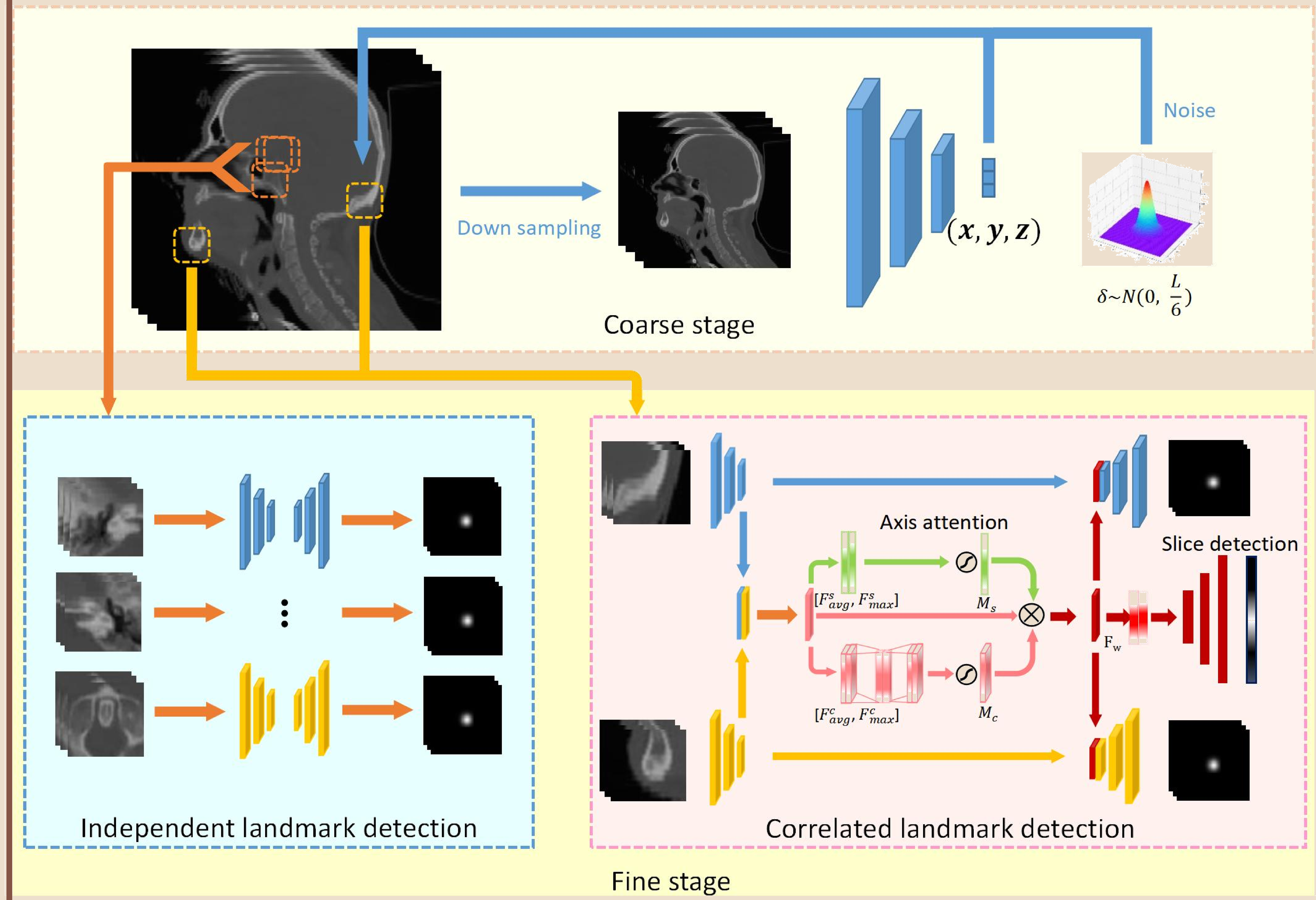
Medical images' local structural similarity often leads to ambiguity in landmark localization, especially in 3D cases wherein adjacent slices have similar structural and intensity profiles. In such context, physicians often annotate landmarks with specific landmarks serving as references or annotate multiple landmarks on a same key slice as shown in panel (b).

We here propose a prior guided coarse-to-fine landmark localization framework to effectively combine the advantages of heatmap regression and coordinate regression and integrate the prior knowledge from physicians' annotation process.




(b) Prostate Landmarks

## Methods




The coarse stage takes the entire downsampled image as input and incorporates structural knowledge. Given an entire image containing n landmarks with ground truth (**x, y, z**), we employ a ResNet-34 for coordinate regression. We modify the output length of the fully connected layer to 3 × N and use regression loss to train the network.

The fine stage focuses on extracting local features around multiple landmarks using patch based Unets. To effectively exploit physicians' prior knowledge, we categorize the landmarks into independent and correlated according to the physicians' annotation practice. For independent landmarks, physicians separately annotate them based on local texture features, each U-Net will extract features and predict heatmaps for the corresponding patch. For the correlated landmarks, physicians typically identify particular slices containing the complete organ's characteristics and annotate correlated landmarks on those slices. We design an axis attention module and key slice detection for key slice querying and landmark detection

**Axis attention:** we utilize different encoders to handle the correlated patches separately, we deeply fuse axis features for key slice detection. Inspired by the CBAM module, we propose axis channel attention and spatial attention to perform dynamically weighted refining $F_w = F \times M_s \times M_c$. Axis is defined as the direction perpendicular to the key slice.

$$M_c = \sigma(Conv_{1\times1\times1}(F^c_{max}) + Conv_{1\times1\times1}(F^c_{avg})).$$

$$M_s = \sigma(Conv_{1\times1\times1}[F^s_{max}, F^s_{avg}]).$$

**Key slice detection:** To improve the localization accuracy in the axial direction, we introduce a slice detection branch to determine the key slice and constrain the attention map. We use 1D convolutions to decode the high-level features and use a Gaussian heatmap to represent the probability of the key slice's location

## Results

Table 1: Performance comparisons in PDDCA dataset and prostate dataset. † and ‡ respectively indicates one-stage methods and coarse-to-fine methods.

| Dataset | method | MRE (SD) | SDR (%) | | | | |
|---|---|---|---|---|---|---|---|
| | | | 2mm | 2.5mm | 3mm | 4mm | 8mm |
| PDDCA | 3D-Unet † | 7.69 (5.24) | 2.03 | 3.25 | 5.05 | 16.28 | 67.90 |
| | SCN † | 7.44 (4.26) | 2.65 | 6.74 | 10.98 | 21.36 | 69.30 |
| | DRM ‡ | 6.39 (3.37) | 7.27 | 12.72 | 16.36 | 29.09 | 74.54 |
| | LA-GCN ‡ | 3.23 (2.52) | 35.68 | 46.76 | 58.19 | 69.48 | 94.74 |
| | SA-LSTM ‡ | 2.37 (1.60) | **56.36** | **71.60** | 80.00 | 89.99 | 95.91 |
| | **Proposed ‡** | **2.13 (1.18)** | 55.23 | 70.12 | **86.20** | **93.50** | **99.40** |
| Prostate | 3D-Unet † | 3.57 (2.27) | 23.84 | 36.82 | 48.23 | 68.12 | 96.32 |
| | SCN † | 3.48 (2.31) | 25.68 | 39.34 | 51.74 | 69.57 | 95.73 |
| | DRM ‡ | 3.44 (**2.21**) | 26.74 | 38.24 | 52.13 | 70.54 | **97.58** |
| | **Proposed ‡** | **3.29** (2.26) | **31.17** | **41.67** | **54.13** | **73.22** | 95.62 |

Table 2: Ablation study results for PDDCA dataset by evaluation metrics of MRE(SD).

| Component | | | Correlated | | Independent | | |
|---|---|---|---|---|---|---|---|
| Fusion | Axis attention | Slice detection | Chin | Occ | Mand_l | Mand_r | Odont_p |
| | | | 2.00 (0.65) | 4.32 (3.10) | 2.05 (0.53) | 2.18 (0.74) | 1.68 (1.14) |
| ✓ | | | 1.87 (**0.57**) | 2.18 (1.18) | | | |
| ✓ | ✓ | | **1.76** (0.66) | 2.28 (1.56) | | | |
| ✓ | ✓ | ✓ | 1.85 (0.62) | **1.79** (**1.15**) | | | |

Among all methods, our proposed method performs the best on PDDCA dataset in terms of MRE (2.13 mm) and SD (1.18 mm). In terms of SDR, our method achieves much higher successful detection rates than the second best-performing method SA-LSTM for the target radii of 3mm, 4mm, and 8mm. For the other two target radii, namely 2mm and 2.5mm, the proposed method's performance is on par with SA-LSTM. On prostate dataset, our proposed method still obtains the best localization performance, in terms of almost all evaluation metrics. This may be because our method has a stronger local feature extraction capability.

shown by Table 2, the implementation of the axial attention module and the slice detection branch aggregate axis features to further guide the localization of ambiguous landmarks. These modules enable our model to achieve a lower MSE (1.82 mm) on correlated landmarks. Collectively, our proposed method can effectively use prior information from physicians' annotation practice and constrains the location of correlated landmarks. By leveraging this prior information, our model exhibits superior performance on multiple datasets.

## Conclusion

We propose a coarse-to-fine framework for localizing independent and correlated anatomical landmarks from 3D medical images. In the fine stage, we employ multiple Unet models for landmarks regression for independent landmarks, ensuring that each model solely focuses on a patch centering the specific single landmark of interest. For the correlated landmarks, we propose a feature fusion module and a key slice detection module. It successfully identifies the position of the key slice from multiple patches and uses the fused features to assist in landmark localization.

Our method outperforms state-of-the-art methods, according to extensive experiments on the publicly-accessible PDDCA dataset and our in-house prostate dataset. We shall further incorporate more types of medical prior knowledge, such as the shape prior of the anatomical region of interest and other prior knowledge utilized in the manual annotation process in future work.