Knowledge-Based Systems 236 (2022) 107680

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

X-CTRSNet: 3D cervical vertebra CT reconstruction and segmentation directly from 2D X-ray images

Rongjun Ge [a,d], Yuting He [b,c,d], Cong Xia [e], Chenchu Xu [f], Weiya Sun [b,c,d], Guanyu Yang [b,c,d], Junru Li [g], Zhihua Wang [c], Hailing Yu [c], Daoqiang Zhang [a,*], Yang Chen [b,c,d,**], Limin Luo [b,c,d], Shuo Li [h], Yinsu Zhu [i,*]

[a] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[b] Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Laboratory of Image Science and Technology, Southeast University, Nanjing, China
[c] School of Computer Science and Engineering, Southeast University, Nanjing, China
[d] Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China
[e] Jiangsu Key Laboratory of Molecular and Functional Imaging, Department of Radiology, Zhongda Hospital, Medical School of Southeast University, Nanjing, China
[f] School of Computer Science and Technology, Anhui University, Hefei, China
[g] College of Software Engineering, Southeast University, Nanjing, China
[h] Department of Medical Imaging, Western University, London, ON, Canada
[i] Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China

ARTICLE INFO

Article history:
Received 13 January 2021
Received in revised form 30 October 2021
Accepted 30 October 2021
Available online 11 November 2021

Keywords:
AI-based medical assistant tool
2D-to-3D
Reconstruction
Segmentation

ABSTRACT

Orthogonal 2D cervical vertebra (C-vertebra) X-ray images have the advantages of high imaging efficiency, low radiation risk, easy operation and low cost for rapid primary clinical diagnoses. Especially in emergency departments, this technique is known to be significantly useful in triage for trauma patients. However, the technique can only provide overlapping anatomic information from limited projection views and is unable to visually exhibit full-view anatomy and precise stereo structures without further CT examination. To promote "once is enough" for visualizing 3D anatomy & structures and reducing repetitive radiation as much as possible, we proposed X-CTRSNet for 2D X-ray images. This is the first powerful work that simultaneously and accurately enables 3D C-vertebra CT reconstruction and segmentation directly from orthogonally anteroposterior- and lateral-view 2D X-ray images. X-CTRSNet combines the reciprocally coupled SpaDRNet for reconstruction & MulSISNet for segmentation, and a RSC Learning for tasks consistency. The experiment shows that X-CTRSNet successfully reconstructs and segments the 3D C-vertebra CT from the 2D X-ray images with a PSNR of 24.58 dB, an SSIM of 0.749, and an average Dice of 80.44%. All these findings reveal the great potential of X-CTRSNet in clinical imaging and diagnosis to facilitate emergency triage by enabling precise 3D reconstruction and segmentation on 2D X-ray images.

1. Introduction

Accurate 3D cervical vertebra (C-vertebra) CT reconstruction and segmentation directly from orthogonal 2D C-vertebra X-ray images is clinically significant to distinctly enable a detailed 3D imaging diagnosis basis for clinicians, and effectively reduce repetitive radiation for patients, especially in assessing C-vertebra disease and surgery, and facilitating emergency triage. Because it is equipped with the efficiency in 2D X-ray imaging, the rich anatomic structure in 3D CT volume, and the intuitive visualization in 3D segmentation, detailed as following: **(1)** 2D C-vertebra X-ray images have the advantages of high imaging speed, low radiation risk, easy operation, low cost and portable development, and are widely used for fast primary clinical diagnoses [1–4]. But it can only provided overlapping anatomic information in a 2D plane from the limited projection view [5–7], as the anteroposterior- (AP) and lateral- (LA) view X-ray images in Figs. 1 (a1) & (a2) show. **(2)** Compared with 2D X-rays, 3D C-vertebra computed tomography (CT) scans are capable of providing full-view anatomy and precise stereo structures of pathologies [8,9]. It enables the volumetric data [10] as Fig. 1(b),

* Corresponding authors.
** Corresponding author at: Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Laboratory of Image Science and Technology, Southeast University, Nanjing, China.
E-mail addresses: dqzhang@nuaa.edu.cn (D. Zhang), chenyang.list@seu.edu.cn (Y. Chen), zhuyinsu@njmu.edu.cn (Y. Zhu).

https://doi.org/10.1016/j.knosys.2021.107680
0950-7051/

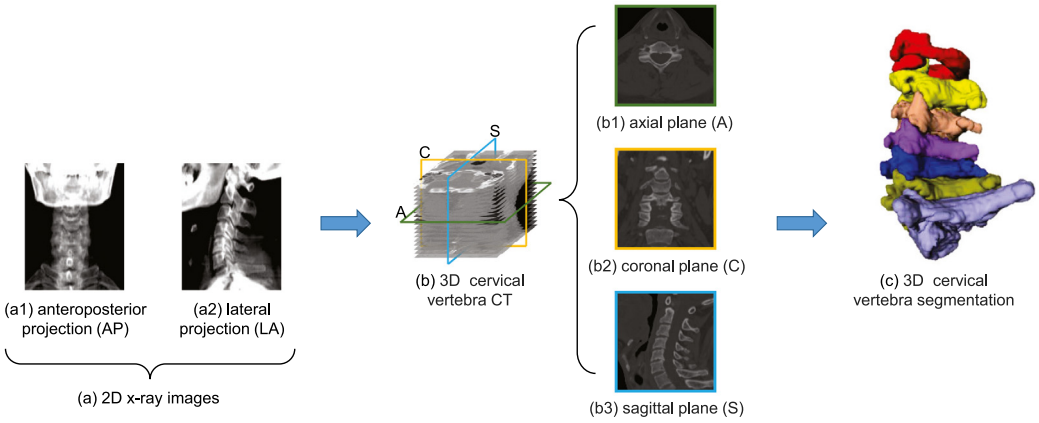

Fig. 1. 3D C-vertebra computed tomography (CT) scan can provide full-view anatomy and enable precise stereo structure, compared to the 2D X-ray images.

and the no-overlapping distinct anatomic structure from different views [11–13] such as the axial, coronal and sagittal planes in Fig. 1 (b1), (b2) & (b3). However, additional pricey CT scans may cause repetitive radiation and unnecessary medical resources occupation of over-treatment, and itself also has a high-dose radiation risk due to multi-slice dense projection [4]. Besides, for the rapid triage of trauma patients in emergency department, it also needs too much time in the race against time, and may overwhelm medical resources in a short time [14]. **(3)** For efficient image interpretation, surgical planning and objective assessment, 3D C-vertebra segmentation directly enables the precise stereo biological structures, as shown in Fig. 1(c). It effectively reflects the morphological shapes, relative locations, and physiological curves for the C-vertebras that has highly flexible anatomy vulnerable to injuries and degeneration. Therefore, it is of great clinical contribution to improve clinical diagnosis efficiency and speed up emergency triage, that with only rapid 2D X-ray image inputs and achieving 3D C-vertebra anatomy and structures as far as possible. It does not aim to replace CT examination completely, but can provide more 3D diagnostic basis on primary 2D X-ray imaging without additional time costs.

Although 3D C-vertebra CT reconstruction and segmentation directly from 2D X-ray images is clinically urgent, and deep learning has achieved great success in lots of clinical related tasks [15–21], the issue still has never been investigated. Some existing works [22–24] have attempted on 3D reconstruction from 2D X-ray images, but they ignored the inherent spatiality and lacked interactive segmentation. While these works show promising results on their issues, they still have difficulties to be applied here, due to the following reasons. **(1)** The C-vertebra has a high requirement of spatiality for stereo correspondence. Because it is important for the C-vertebra spatial evaluation of dislocation, physiological curves, and distance measurements [25]. But the existing works [22–24] just using only the 1D feature vector for 2D to 3D transformation ignore the inherent spatiality. **(2)** The lack of an interactive scheme that promotes multi-task learning with segmentation is needed to ensure reconstruction–segmentation consistency. So that can reciprocally make the shape constraint and enhance the biological CT anatomical texture for the reconstructed 3D CT, and strengthen the precise segmentation robustness on the reconstructed image. **(3)** Besides, for 3D reconstructed volumes from 2D projected images, content consistency is needed to express the anatomy, which is beyond the abilities of the former used voxel-to-voxel local optimization lacking assembled expressiveness. **(4)** The C-vertebra has multi-scale structure components such as the vertebral body, spinous process, transverse process, foramen transversarium, etc., and pathologically correlated locations inter C-vertebras, which requires multi-scale stereoscopical information for precise segmentation.

As shown in Fig. 2, we proposed the first powerful work, X-CTRSNet, which simultaneously and accurately enables 3D C-vertebra CT reconstruction and segmentation directly from widely accessible AP and LA view 2D X-ray images. X-CTRSNet is composed of three elements, including Spatial Decomposition-Reconstruction Net (SpaDRNet), Multi-scale Space Interoperability Segmentation Net (MulSISNet), and Reconstruction–segmentation Consistency (RSC) Learning. The effects of these three specially-designed elements can be summarized as follows: **(1) SpaDRNet** (Sect. 2.1) is used to achieve the 3D C-vertebra CT reconstruction directly from AP & LA projected 2D planes, as shown in Fig. 3(b). It consists of progressive 2D-to-3D conversion to multi-scale decompose the compressed space to the corresponding stereo location from the overlapped domain, hierarchical 3D fusion to interpretively sort out the 3D spatial information in consistency, and multi-view vgg loss to expressively guide the reconstructed scene content learning. And **(2) MulSISNet** (Sect. 2.2) further enables 3D C-vertebra semantic segmentation on reconstructed CT, and promotes shape constraints to SpaDRNet, as shown in Fig. 3(c). It comprehensively extracts rich stereo structure covering from the small-scale details to the large-scale distribution with the information densely interoperated among multi-scale 3D feature. Interactively, **(3) RSC Learning** (Sect. 2.3) enhances reconstruction–segmentation consistent with the ground truth (GT) for the multi-task learning, as shown in Fig. 3(d). It promotes segmenting on the CT ground truth (GT), and optimize the divergence of the reconstructed CT's segmentation with it.

The contributions of this work are summarized as following:

- For the first time, the proposed X-CTRSNet promotes "once is enough" for achieving 3D C-vertebra reconstruction and segmentation directly from AP and LA view 2D X-ray images. It efficiently gains detailed 3D imaging diagnosis basis with only 2D X-ray imaging, so that effectively improves clinical diagnosis and imaging efficiency of C-vertebra, as well as greatly reduces repetitive radiation.
- The newly designed SpaDRNet enables 2D-to-3D reconstruction of C-vertebra from overlapping 2D projections for clear 3D anatomy. It novelly designs the progressive 2D-to-3D conversion to multi-scale decompose the compressed space to the spatially corresponding stereo location from the overlapped 2D domain, and the hierarchical 3D fusion to interpretively sort out the multi-scale 3D spatial information in consistency. During network learning, it creatively develops multi-view vgg loss to guide the reconstructed scene content of expressing anatomy, for volumetric images.

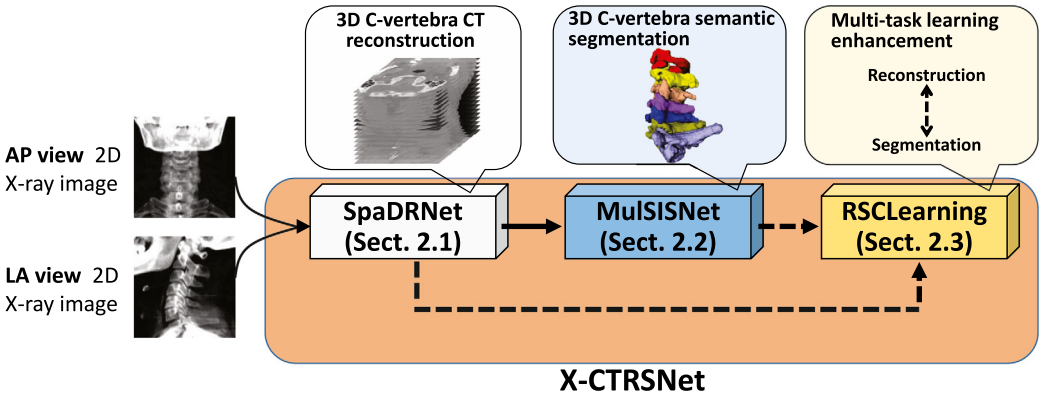

Fig. 2. Workflow diagram of X-CTRSNet.

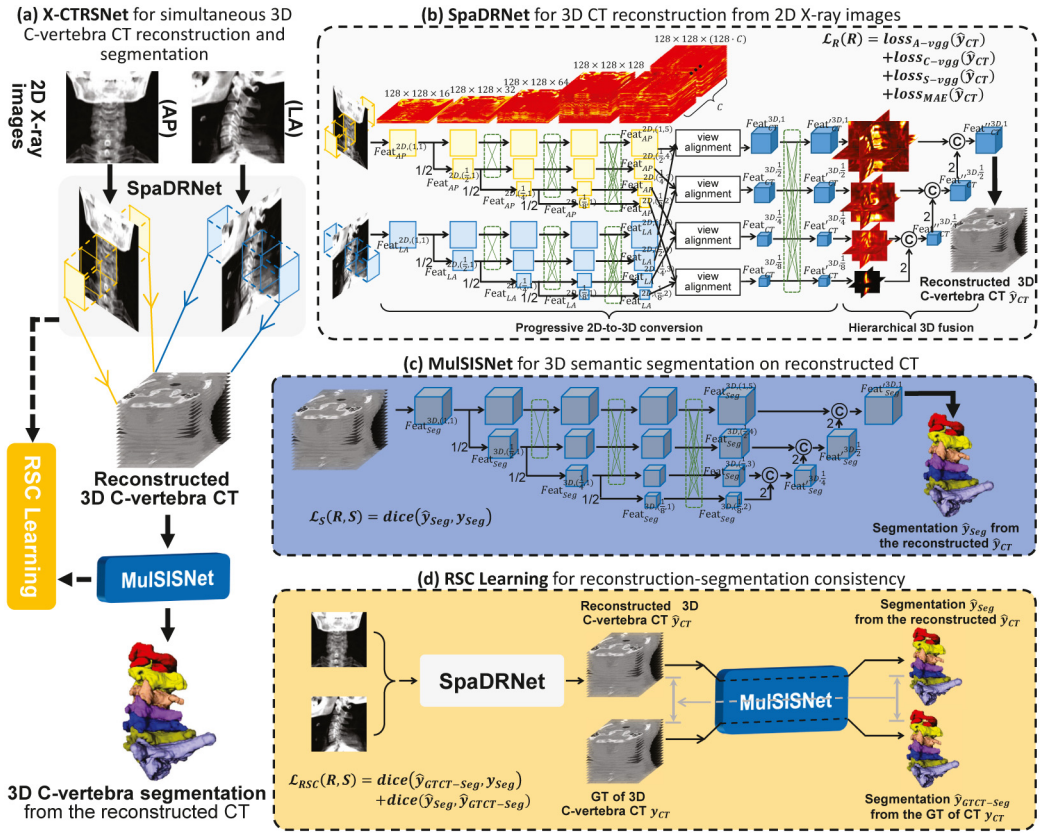

Fig. 3. X-CTRSNet is achieved via reciprocally coupled SpaDRNet of reconstruction & MulSISNet of Segmentation, and a RSC Learning of tasks consistency, to simultaneously enable 3D C-vertebra CT reconstruction and segmentation directly from the 2D X-ray images.

- The novel MulSISNet promotes 3D segmentation from reconstructed CT to distill the 3D structure of the C-vertebra, and establishes shape constraints to SpaDRNet. It robustly extracts the multi-scale stereo structures and distribution for the fine-gained spatial representation, and densely interoperates among the multi-scale 3D features to fully exploit the learned information for enhancing each other and compensating the lost. Besides, it also feeds back the segmented shape error to guide SpaDRNet reconstruction.
- The creative RSC Learning enhances the multi-task learning of reconstruction and segmentation. It forcefully introduces the CT GT to strengthen the segmentation effectiveness, reducing the misleading of MulSISNet from the early-stage unstable reconstruction; and deeply feeds back the divergence between two segmentations for the reconstructed biological CT anatomy.

2. Methodology

As shown in Fig. 3(a), the proposed X-CTRSNet is conducted on the AP and LA views 2D X-ray images to directly make the 3D C-vertebra CT reconstruction and segmentation. So that it achieves the full-view anatomy and precise stereo structure, making up the shortage in 2D imaging. It is built by three collaborate elements: **(1)** SpaDRNet (R, Section 2.1) combines progressive 2D-to-3D multi-paths, hierarchical 3D fusion and multi-view vgg loss to decompose the overlapped 2D X-ray images into reconstructing the detailed 3D CT. **(2)** MulSISNet (S, Section 2.2) extracts the

robust multi-scale stereo features for the reciprocal 3D semantic segmentation on reconstructed CT and shape constraint feedback. **(3)** RSC Learning ($\mathcal{L}_{RSC}$, Section 2.3) promotes segmenting on the CT GT and optimizes the divergence between two segmentations to drive the reconstruction–segmentation consistency. Given the AP & LA views 2D X-ray images x_{AP}&x_{LA}, the GT of 3D CT y_{CT} and segmentation y_{Seg}, the target of X-CTRSNet is formulated as:

$$\min_{R,S} \mathcal{L}_{X-CTRSNet} = \mathcal{L}_R(R) + \mathcal{L}_S(R,S) + \mathcal{L}_{RSC}(R,S) \tag{1}$$

2.1. SpaDRNet for 3D CT reconstruction from 2D X-ray images

Our SpaDRNet in Fig. 3(b) innovatively uses **progressive 2D-to-3D conversion**, **hierarchical 3D fusion**, and **multi-view vgg loss**, to decompose the latent dimension space in the overlapped 2D projection, and spatially correspondingly reconstruct into 3D CT images, directly from the AP and LA views X-ray images.

2.1.1. Progressive 2D-to-3D conversion

As shown in Fig. 3(b), the **progressive 2D-to-3D conversion** exploits the 2D collateral multi-paths and the view alignment, to extract the 3D spatial information existing in 2D projection.

(1) The *2D collateral multi-paths* progressively extract the projected information in multi-scale with spatial correspondence instead of spatialless 1D abstraction. The beginnings of each path $Feat^{2D,(S,1)}$, $S \in \left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}\right\}$ for different scales, are calculated as:

$$\begin{aligned}
Feat^{2D,(1,1)} &= \mathrm{SeqConv}^{2D}(I) \\
Feat^{2D,(\frac{1}{2},1)} &= \mathrm{SeqConv}^{2D}(\mathrm{DConv}(Feat^{2D,(1,1)})) \\
Feat^{2D,(\frac{1}{4},1)} &= \mathrm{SeqConv}^{2D}(\mathrm{DConv}(Feat^{2D,(\frac{1}{2},1)})) \\
Feat^{2D,(\frac{1}{8},1)} &= \mathrm{SeqConv}^{2D}(\mathrm{DConv}(Feat^{2D,(\frac{1}{4},1)}))
\end{aligned} \tag{2}$$

where $\mathrm{SeqConv}^{2D}(\cdot)$ means sequentially conducts the 2D 3×3 convolution with Leaky ReLU activation function. $\mathrm{DConv}(\cdot)$ is the down-sampling operation by using the 3×3 convolutions with stride 2, and I is the inputting 2D X-ray images.

In addition, the multi-scale fusion [26] is used during the scale progression for information compensation and representation enhancement. This procedure can be formulated as:

$$\begin{aligned}
Feat^{2D,(1,l+1)} &= \mathrm{SeqConv}^{2D}(Feat^{2D,(1,l)} + \mathrm{U}(Feat^{2D,(\frac{1}{2},l-1)}) \\
&\quad + \mathrm{U}(Feat^{2D,(\frac{1}{4},l-2)}) + \mathrm{U}(Feat^{2D,(\frac{1}{8},l-3)})) \\
Feat^{2D,(\frac{1}{2},l+1)} &= \mathrm{SeqConv}^{2D}(Feat^{2D,(\frac{1}{2},l)} + \mathrm{D}(Feat^{2D,(1,l+1)}) \\
&\quad + \mathrm{U}(Feat^{2D,(\frac{1}{4},l-1)}) + \mathrm{U}(Feat^{2D,(\frac{1}{8},l-2)})) \\
Feat^{2D,(\frac{1}{4},l+1)} &= \mathrm{SeqConv}^{2D}(Feat^{2D,(\frac{1}{4},l)} + \mathrm{D}(Feat^{2D,(1,l+2)}) \\
&\quad + \mathrm{D}(Feat^{2D,(\frac{1}{2},l+1)}) + \mathrm{U}(Feat^{2D,(\frac{1}{8},l-1)})) \\
Feat^{2D,(\frac{1}{8},l+1)} &= \mathrm{SeqConv}^{2D}(Feat^{2D,(\frac{1}{8},l)} + \mathrm{D}(Feat^{2D,(1,l+3)}) \\
&\quad + \mathrm{D}(Feat^{2D,(\frac{1}{2},l+2)}) + \mathrm{D}(Feat^{2D,(\frac{1}{4},l+1)}))
\end{aligned} \tag{3}$$

where l is the layer number of the corresponding path, $\mathrm{D}(\cdot)$ is the down-sampling operation composed of consecutive strided 3×3 convolutions with stride 2, $\mathrm{U}(\cdot)$ represents the simple nearest neighbor sampling following a 1×1 convolution.

Progressively, the 2D overlapping is spatially converted into 3D decomposition, extracting the projected dimension and keeping the plane spatiality, as $128 \times 128 \times 1 \rightarrow 128 \times 128 \times 16 \rightarrow 128 \times 128 \times 32 \rightarrow 128 \times 128 \times 64 \rightarrow 128 \times 128 \times 128$ (the first path for instance), so that promotes the 2D–3D spatial correspondence in conversion. At the end, a further 2D convolution with pixelshuffle [27,28] is used to convert into $128 \times 128 \times (128 \cdot C)$ to enlarge the feature channels for more expression in the 3D domain, formulated as:

$$Feat^{3D,S} = \mathrm{PS}(Feat^{2D,(S,L)}) \tag{4}$$

where $\mathrm{PS}(\cdot)$ is pixelshuffle operation.

Thanks to these, the 2D collateral multi-paths are able to effectively decompose the overlapped anatomical structures.

(2) The *view alignment* combines the AP and the LA information to construct the stereo anatomical structure. It transforms the AP and LA features to be consistent in orientation, and integrates the orthogonal spatial anatomy. The further 3D convolution with multi-scale fusion is used to enhance the correlation of the voxel for the stereo structure.

In detail, the 3D features with projected information extracted from AP and LA views are firstly permuted to be assigned to each other and 3D CT orientation. For LA-view 3D feature $Feat^{3D}_{LA}(X_{LA}, Y_{LA}, Z_{LA})$, the first dimension X_{LA} represents the length dimension of LA-view X-ray image, and corresponds to the width dimension Y_{CT} of 3D CT data. Such real-word geometric correspondence of all dimensions between 3D features of X-ray images and CT data can be described as:

$$\begin{aligned}
X_{LA} &\leftrightarrow Y_{CT} \\
Y_{LA} &\leftrightarrow Z_{CT} \\
Z_{LA} &\leftrightarrow X_{CT} \\
X_{AP} &\leftrightarrow X_{CT} \\
Y_{AP} &\leftrightarrow Z_{CT} \\
Z_{AP} &\leftrightarrow Y_{CT}
\end{aligned} \tag{5}$$

where X_{LA}, Y_{LA}, Z_{LA} are dimensions in LA-view 3D feature $Feat^{3D}_{LA}(X_{LA}, Y_{LA}, Z_{LA})$, X_{AP}, Y_{AP}, Z_{AP} mean dimensions in AP-view 3D feature $Feat^{3D}_{AP}(X_{AP}, Y_{AP}, Z_{AP})$, and X_{CT}, Y_{CT}, Z_{CT} denote dimensions in CT feature $Feat^{3D}_{CT}(X_{CT}, Y_{CT}, Z_{CT})$. Therefore, to promote view alignment, the permutations of $Feat^{3D}_{LA}$ and $Feat^{3D}_{AP}$ are formulated as:

$$\begin{aligned}
Feat'^{3D}_{LA} &= \mathrm{P}(Feat^{3D}_{LA}, [Z_{LA}, X_{LA}, Y_{LA}]) \\
Feat'^{3D}_{AP} &= \mathrm{P}(Feat^{3D}_{AP}, [X_{AP}, Z_{AP}, Y_{AP}])
\end{aligned} \tag{6}$$

where P is the permutation operation, according to the order $[\cdot]$. Then, the permuted 3D features $Feat'^{3D}_{LA}$ and $Feat'^{3D}_{AP}$ are concatenated along channel as:

$$Feat^{3D}_{CT} = \mathrm{Concat}(Feat'^{3D}_{LA}, Feat'^{3D}_{AP}) \tag{7}$$

And a following multi-scale 3D fusion is conducted to further merge the multi-view decomposed stereo information and integrate multi-scale structure. Given the permuted 3D features $Feat^{3D,\frac{1}{8}}_{CT}$, $Feat^{3D,\frac{1}{4}}_{CT}$, $Feat^{3D,\frac{1}{2}}_{CT}$ and $Feat^{3D,1}_{CT}$ from the multi-paths with scales of $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$ and 1, respectively, the procedures are depicted as

$$\begin{aligned}
Feat'^{3D,1}_{CT} &= \mathrm{SeqConv}^{3D}(Feat^{3D,1}_{CT} + \mathrm{U}(Feat^{3D,\frac{1}{2}}_{CT}) + \mathrm{U}(Feat^{3D,\frac{1}{4}}_{CT}) + \mathrm{U}(Feat^{3D,\frac{1}{8}}_{CT})) \\
Feat'^{3D,\frac{1}{2}}_{CT} &= \mathrm{SeqConv}^{3D}(Feat^{3D,\frac{1}{2}}_{CT} + \mathrm{D}(Feat^{3D,1}_{CT}) + \mathrm{U}(Feat^{3D,\frac{1}{4}}_{CT}) + \mathrm{U}(Feat^{3D,\frac{1}{8}}_{CT})) \\
Feat'^{3D,\frac{1}{4}}_{CT} &= \mathrm{SeqConv}^{3D}(Feat^{3D,\frac{1}{4}}_{CT} + \mathrm{D}(Feat^{3D,1}_{CT}) + \mathrm{D}(Feat^{3D,\frac{1}{2}}_{CT}) + \mathrm{U}(Feat^{3D,\frac{1}{8}}_{CT})) \\
Feat'^{3D,\frac{1}{8}}_{CT} &= \mathrm{SeqConv}^{3D}(Feat^{3D,\frac{1}{8}}_{CT} + \mathrm{D}(Feat^{3D,1}_{CT}) + \mathrm{D}(Feat^{3D,\frac{1}{2}}_{CT}) + \mathrm{D}(Feat^{3D,\frac{1}{4}}_{CT}))
\end{aligned} \tag{8}$$

where $\mathrm{SeqConv}^{3D}(\cdot)$ represents sequentially conducting the 3D $3 \times 3 \times 3$ convolution with Leaky ReLU activation function. $\mathrm{D}(\cdot)$ is the down-sampling operation composed of consecutive strided $3 \times 3 \times 3$ convolutions with stride 2, $\mathrm{U}(\cdot)$ represents the simple nearest neighbor sampling following a $1 \times 1 \times 1$ convolution.

2.1.2. Hierarchical 3D fusion

To interpretively sort out the 3D spatial information in consistency, the **hierarchical 3D fusion** in Fig. 3(b) gradually merges the 3D features between the adjacent scale expressions. The procedure uses the transpose convolution to up-sample for arranging

the adjacent scales in a learnable way, and fuse the grouped adjacent scales with successive 3D convolutions, formulaically described as:

$$\begin{aligned} Feat''^{3D,1}_{CT} &= \text{SeqConv}^{3D}(\text{Concat}(Feat'^{3D,1}_{CT}, \text{UConv}(Feat''^{3D,\frac{1}{2}}_{CT}))) \\ Feat''^{3D,\frac{1}{2}}_{CT} &= \text{SeqConv}^{3D}(\text{Concat}(Feat'^{3D,\frac{1}{2}}_{CT}, \text{UConv}(Feat''^{3D,\frac{1}{4}}_{CT}))) \\ Feat''^{3D,\frac{1}{4}}_{CT} &= \text{SeqConv}^{3D}(\text{Concat}(Feat'^{3D,\frac{1}{4}}_{CT}, \text{UConv}(Feat'^{3D,\frac{1}{8}}_{CT}))) \end{aligned} \tag{9}$$

where is UConv$(\cdot)$ is the up-sampling operation by using the $3 \times 3 \times 3$ transpose convolution with stride 2.

Hierarchical 3D fusion thus explicitly interprets the consistent multi-scale structure in 3D C-vertebra CT, and reconstructs the stereo anatomy according to the strong spatial relation inter adjacent scale.

2.1.3. Multi-view vgg loss

For the assembled expressiveness of the voxels in 3D reconstructed CT, **multi-view vgg** loss is creatively developed on all planes along the axial, the coronal and the sagittal views. It guides content consistency among the voxels in the 3D scene to express the anatomy. It is directly transferred from the widely accepted pre-trained VGGNet. With the pre-trained VGGNet [29], the loss extracts the high-level feature representations of expressing [30] for the scene interpretation, and further constraints in multi-views for the stereo context. Given the reconstructed 3D CT $\hat{y}_{CT}$ and its GT y_{CT}, multi-view vgg loss is defined as:

$$\begin{aligned} loss_{MV-vgg} &= loss_{A-vgg}(\hat{y}_{CT}) + loss_{C-vgg}(\hat{y}_{CT}) + loss_{S-vgg}(\hat{y}_{CT}) \\ &= \frac{1}{L}\sum_i \left\| vgg(\hat{y}^i_A) - vgg(y^i_A) \right\|_2^2 \\ &\quad + \frac{1}{W}\sum_j \left\| vgg(\hat{y}^j_C) - vgg(y^j_C) \right\|_2^2 \\ &\quad + \frac{1}{H}\sum_k \left\| vgg(\hat{y}^k_S) - vgg(y^k_S) \right\|_2^2 \end{aligned} \tag{10}$$

where $\hat{y}^i_A$, $\hat{y}^j_C$ and $\hat{y}^k_S$ are planes exported from $\hat{y}_{CT}$ along the axial, the coronal and the sagittal views. L, W and H are the length, width and height of $\hat{y}_{CT}$.

Furthermore, the loss function of SpaDRNet consists of $loss_{MV-vgg}$ for anatomy expression and $loss_{MAE}$ for voxel details, as follows:

$$\begin{aligned} \mathcal{L}_R(R) &= loss_{MV-vgg} + loss_{MAE} \\ &= loss_{MV-vgg} + \left\| \hat{y}_{CT} - y_{CT} \right\| \end{aligned} \tag{11}$$

2.2. MulSISNet for 3D semantic segmentation on reconstructed CT

Our MulSISNet in Fig. 3(c) robustly utilizes multi-scale 3D extraction and interoperation to interactively make the 3D C-vertebra semantic segmentation from the reconstructed CT, and transfers the shape constraint to SpaDRNet.

2.2.1. MulSISNet architecture

The *MulSISNet architecture* is inspired by the robust representation of HRNet [26], and the efficient architecture of 3D UNet [31]. As shown Fig. 3(c), MulSISNet further lengthens each encoded layer to connect the decoder, instead of skip connection, for avoiding the low-level representation of larger size feature with shallow extraction and the information loss of smaller size feature with downsample in 3D UNet. So that it develops 3D multi-scale paths in 3D UNet for the different-range stereo structure correlation acquirement for C-vertebra shape. To make full use of the learned information and compensate the lost of the feature extraction in each scale path, MulSISNet deploys the interoperation among multi-scales paths to enhance the representation of each other. During the interoperation, MulSISNet develops the multi-scale fusion [26] into stereo by using $3 \times 3 \times 3$ convolution with *stride* $= 2$ to downsample, nearest neighbor interpolation$+1 \times 1 \times 1$ convolution to upsample, and summation to merge the arranged multi-scales. The procedure can be formulated as:

$$\begin{aligned} Feat^{3D,(1,l+1)}_{Seg} &= \text{SeqConv}^{3D}(Feat^{3D,(1,l)}_{Seg} + \text{U}(Feat^{3D,(\frac{1}{2},l-1)}_{Seg}) \\ &\quad + \text{U}(Feat^{3D,(\frac{1}{4},l-2)}_{Seg}) + \text{U}(Feat^{3D,(\frac{1}{8},l-3)}_{Seg})) \\ Feat^{3D,(\frac{1}{2},l+1)}_{Seg} &= \text{SeqConv}^{3D}(Feat^{3D,(\frac{1}{2},l)}_{Seg} + \text{D}(Feat^{3D,(1,l+1)}_{Seg}) \\ &\quad + \text{U}(Feat^{3D,(\frac{1}{4},l-1)}_{Seg}) + \text{U}(Feat^{3D,(\frac{1}{8},l-2)}_{Seg})) \\ Feat^{3D,(\frac{1}{4},l+1)}_{Seg} &= \text{SeqConv}^{3D}(Feat^{3D,(\frac{1}{4},l)}_{Seg} + \text{D}(Feat^{3D,(1,l+2)}_{Seg}) \\ &\quad + \text{D}(Feat^{3D,(\frac{1}{2},l+1)}_{Seg}) + \text{U}(Feat^{3D,(\frac{1}{8},l-1)}_{Seg})) \\ Feat^{3D,(\frac{1}{8},l+1)}_{Seg} &= \text{SeqConv}^{3D}(Feat^{3D,(\frac{1}{8},l)}_{Seg} + \text{D}(Feat^{3D,(1,l+3)}_{Seg}) \\ &\quad + \text{D}(Feat^{3D,(\frac{1}{2},l+2)}) + \text{D}(Feat^{3D,(\frac{1}{4},l+1)}_{Seg})) \end{aligned} \tag{12}$$

and

$$\begin{aligned} Feat^{3D,(1,1)}_{Seg} &= SeqConv^{3D}(\hat{y}_{CT}) \\ Feat^{3D,(\frac{1}{2},1)}_{Seg} &= \text{SeqConv}^{3D}(\text{DConv}(Feat^{3D,(1,1)}_{Seg})) \\ Feat^{3D,(\frac{1}{4},1)}_{Seg} &= \text{SeqConv}^{3D}(\text{DConv}(Feat^{3D,(\frac{1}{2},1)}_{Seg})) \\ Feat^{3D,(\frac{1}{8},1)}_{Seg} &= \text{SeqConv}^{3D}(\text{DConv}(Feat^{3D,(\frac{1}{4},1)}_{Seg})) \end{aligned} \tag{13}$$

Finally, the multi-scale stereo features are transmitted into the decoder to summarize the semantics for segmentation, formulaically described as:

$$\begin{aligned} Feat'^{3D,1}_{Seg} &= \text{SeqConv}^{3D}(\text{Concat}(Feat^{3D,(1,5)}_{Seg}, \text{UConv}(Feat'^{3D,\frac{1}{2}}_{Seg}))) \\ Feat'^{3D,\frac{1}{2}}_{Seg} &= \text{SeqConv}^{3D}(\text{Concat}(Feat^{3D,(\frac{1}{2},4)}_{Seg}, \text{UConv}(Feat'^{3D,\frac{1}{4}}_{Seg}))) \\ Feat'^{3D,\frac{1}{4}}_{Seg} &= \text{SeqConv}^{3D}(\text{Concat}(Feat^{3D,(\frac{1}{4},3)}_{Seg}, \text{UConv}(Feat^{3D,(\frac{1}{8},2)}_{Seg}))) \end{aligned} \tag{14}$$

2.2.2. Dice loss

Reciprocally, on the basis of the reconstructed CT, there is a strong relation inter the tasks, so that MulSISNet further transfers the segmentation learning to make shape constraint on the reconstruction in SpaDRNet. The segmentation learning loss is calculated with dice as Eq. (15), where $\hat{y}^{m,n,p}_{Seg}$ and $y^{m,n,p}_{Seg}$ are voxel segmentation and its GT at (m, n, p).

$$\begin{aligned} \mathcal{L}_S(R,S) &= 1 - dice(\hat{y}_{Seg}, y_{Seg}) \\ &= 1 - \frac{2\sum_{m,n,p}\hat{y}^{m,n,p}_{Seg} \cdot y^{m,n,p}_{Seg}}{\sum_{m,n,p}\hat{y}^{m,n,p}_{Seg} + \sum_{m,n,p} y^{m,n,p}_{Seg}} \end{aligned} \tag{15}$$

2.3. RSC learning for reconstruction–segmentation consistency

Our RSC Learning in Fig. 3(d) creatively introduces CT GT-to-segmentation learning, to drive the reconstruction–segmentation consistency of the interactive tasks.

(1) RSC Learning on segmentation: It strengthens the segmentation by enforcing the segmentation learning on the CT GT during X-CTRSNet training, so that robustly reduces the misleading and mess from the early-stage unstable reconstruction, and stabilizes the learning direction of MulSISNet. Moreover, with such segmentation optimization on CT GT, it further guides the effectiveness of MulSISNet on real CT data, thus refining the segmentation feedback and driving consistency with CT.

Fig. 4. X-CTRSNet anatomically enables a 3D C-vertebra physiological structure illustration.

(2) RSC Learning on reconstruction: It enhances the reconstruction as penalizing the reconstruction training with the divergence between the two segmentations of reconstructed CT and CT GT, so that deeply reinforces the reconstructed biological CT anatomical texture. By penalizing the inter-segmentations divergence, it can converge the reconstructed CT to the CT with the real CT consistent segmentation. Through beneficial interaction, RSC Learning enables X-CTRSNet the consistently and precisely coupled tasks reconstruction–segmentation.

Given the segmentations $\hat{y}_{Seg}, \hat{y}_{GTCT-Seg}$ of the reconstructed CT and CT GT, the loss function is defined as:

$$\mathcal{L}_{RSC}(R,S) = dice(\hat{y}_{GTCT-Seg}, y_{Seg}) + dice(\hat{y}_{Seg}, \hat{y}_{GTCT-Seg}) \tag{16}$$

3. Experiments and analysis

3.1. Materials and configurations

Clinical data from 69 patients were used for the evaluation. The segmentation GT of C-vertebras (ordered as C1,C2,C3,C4,C5, C6 and C7) is labeled by two radiologists with cross-check. Specifically, Instance Normalization and Group Normalization are used in SpaDRNet of reconstruction and MulSISNet of segmentation, respectively. The network is implemented using Tensorflow with the Adam optimizer. The initial learning rate is set as 10^{-3}. Ten-folder cross validation is adopted in the performance evaluation and comparison. The dataset is divided into 10 groups. In the first nine groups, there were 7 patients in each group. And the last group contains 6 patients. In each validation, nine groups are used to train the network, and the last group is used for test. The procedure was repeated 10 times, until all the subjects have been processed.

Structural similarity index (SSIM) [32] and peak signal to noise ratio (PSNR) are employed to evaluate the reconstruction performance, as well as Dice coefficient (Dice) [33] is used for segmentation assessment. SSIM is defined as

$$SSIM(y_{CT}, \hat{y}_{CT}) = \frac{(2\mu_{y_{CT}}\mu_{\hat{y}_{CT}} + c_1)(2cov(y_{CT}, \hat{y}_{CT}) + c_2)}{(\mu_{y_{CT}}^2 + \mu_{\hat{y}_{CT}}^2 + c_1)(\sigma_{y_{CT}}^2 + \sigma_{\hat{y}_{CT}}^2 + c_2)}, \tag{17}$$

where μ means average, σ is standard deviation, $cov(\cdot)$ denotes covariance, c represents variables to stabilize the division with weak denominator.

PSNR is calculated as

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_{y_{CT}}^2}{\frac{1}{LWH}\sum_{m=1,n=1,p=1}^{L,W,H}[\hat{y}_{CT}^{m,n,p} - y_{CT}^{m,n,p}]^2}\right), \tag{18}$$

And Dice has definition of

$$Dice = \frac{2\left|y_{Seg} \cap \hat{y}_{Seg}\right|}{\left|y_{Seg}\right| + \left|\hat{y}_{Seg}\right|} \cdot 100\% \tag{19}$$

3.2. Results and analysis

3.2.1. Overall performance

As the last row in Table 1 shows, the proposed X-CTRSNet successfully achieves high-performance 3D C-vertebra CT reconstruction and segmentation directly from the 2D X-ray images. It gains a high SSIM of 0.749 and a high PSNR of 24.58 dB for the reconstructed CT, as well as an average Dice up to 80.44% for the seven segmented C-vertebras (C1,C2,C3,C4,C5,C6 and C7). So that it anatomically enables a 3D cervical vertebra physiological structure illustration, as shown in Fig. 4, with "once is enough" fast 2D imaging to speed up the diagnostic procedure and reduce the repetitive radiation.

3.2.2. Ablation study

As shown in Table 1, the innovative components designed for X-CTRSNet, including SpaDRNet, $loss_{MV-vgg}$, MulSISNet and RSC Learning, enable robust improvements. By using **SpaDRNet**, the anatomical structure from the overlapped 2D X-ray images are effectively decomposed layer-by-layer as shown in Fig. 5, thanks to its progressive 2D–3D conversion with spatial correspondence. By using $\boldsymbol{loss_{MV-vgg}}$, the performance of reconstruction gains 2.11% improvement in SSIM and 0.31 dB improvement in PSNR. It is beneficial from the guidance of $loss_{MV-vgg}$ for content consistency among voxels in 3D scenes to express biological anatomy. By using **MulSISNet**, an accurate segmentation is further enabled for 3D morphology extraction, and meanwhile enhance 3D CT reconstruction. It is contributed by the multi-scale 3D extraction and interoperation in MulSISNet, as well as utilizing the reciprocal relation inter tasks. By using **RSC Learning**, the best performance in both reconstruction and segmentation is further achieved, as RSC Learning promotes the reconstruction–segmentation consistency of the interactive tasks with CT GT-to-segmentation learning.

3.2.3. Comparison experiments

As shown in Figs. 2 & 3, X-CTRSNet achieves superior accuracy in both tasks compared to the state-of-the-art methods: **(1)** SIT [22], PSR [23] and X2CT-GAN [24] for reconstruction, as well as **(2)** 3D UNet [31], DSN [34], DenseBiasNet [35], CS^2-Net [36] and ConResNet [37] for after-reconstruction 3D segmentation.

In the reconstruction comparison (Table 2), X-CTRSNet improved the SSIM by 13.78% on average for accurate anatomical structures, and increased the PSNR by 2.27 dB for clearly readable imaging. As shown in Fig. 6, it visually enables clear and robust full-view biologic structures without overlapping that have legible distribution, shape and explicitly readable anatomical texture, and further promotes the precise 3D segmentation of the C-vertebras morphology on the reconstructed stereo CT data. So that X-CTRSNet distinctly provides the detailed 3D imaging diagnosis basis from the 2D X-ray imaging characterized by low radiation risk. But the compared ones behave poorly, which causes rough shape of C-vertebras groups, and fails on each detailed C-vertebra and the interlock relation among C-vertebras.

In the segmentation comparison, as shown in Table 3, our proposed method still effectively improves the average Dice with 3.25%, and comprehensive promotes the more precise segmentation for all C-vertebras C1 to C7. Visually in Fig. 7, it is robust to segment the multi-scale structure components distributed among C-vertebras. As the multi-scale structure components (circled by dotted line in Fig. 7) including foramen transversarium, spinous process, and vertebral body which cause difficulties to the compared methods. Our X-CTRSNet still precisely segments them for the morphology extraction, thanks to the multi-scale path and interoperation in its sub module MulSISNet.

Furthermore, besides the above accuracy comparison in both tasks, the comparison of model complexities is also made. As

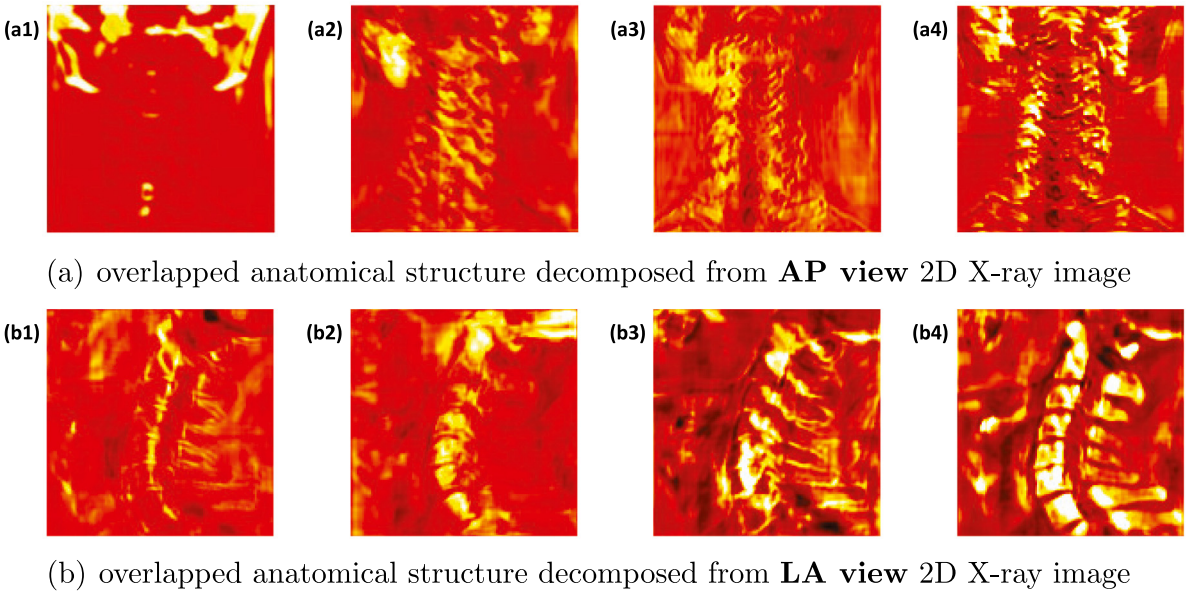

Fig. 5. SpaDRNet effectively decomposes the anatomical structure from the overlapped 2D X-ray images, by progressive 2D–3D conversion with spatial correspondence.

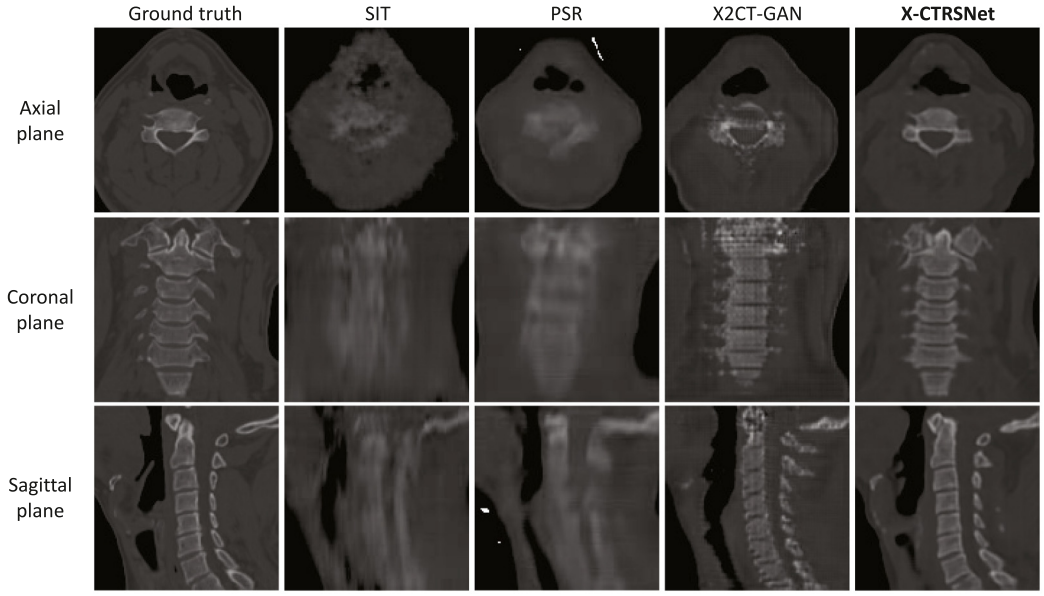

Fig. 6. X-CTRSNet achieves the reconstruction of the detailed CT anatomy of legible distribution, shape and explicitly readable anatomical texture, and thus further promotes the precise 3D structure segmentation.

Table 1
X-CTRSNet successfully achieves the accurate reconstruction and segmentation, contributed to its innovative components.

SpaDRNet	$loss_{MV-vgg}$	MulSISNet	RSC Learning	Reconstruction		Segmentation(Dice %)							
				SSIM	PSNR(dB)	C1	C2	C3	C4	C5	C6	C7	Mean
✓				0.712 ± 0.047	23.54 ± 1.90	/	/	/	/	/	/	/	/
✓	✓			0.727 ± 0.048	23.85 ± 1.88	/	/	/	/	/	/	/	/
✓	✓	✓		0.736 ± 0.045	24.27 ± 1.63	72.41 ± 10.59	81.81 ± 9.13	80.72 ± 6.64	79.78 ± 8.79	77.77 ± 10.43	80.71 ± 6.55	83.22 ± 4.22	79.49 ± 8.05
✓	✓	✓	✓	**0.749 ±0.043**	**24.58 ±1.56**	**74.71 ±8.39**	**83.25 ±7.21**	**81.89 ±5.83**	**80.15 ±8.42**	**78.37 ±9.75**	**80.91 ±6.45**	**83.80 ±4.88**	**80.44 ±7.27**

shown in the last 3 columns in Figs. 2 & 3, compared with the known methods, our proposed method gains acceptable performance in model complexities including the number of trainable parameters, floating point operations (FLOPs), and inference speed, especially when combined with our high-quality reconstruction and segmentation results. (1) Significantly, during the inference, our method only needs 0.592 s to conduct both the 3D CT reconstruction and segmentation from two 2D X-ray images

Table 2
X-CTRSNet gains superior accuracy in reconstruction compared to the state-of-the-art method, with the acceptable model complexity.

	SSIM	PSNR (dB)	Parameters number	FLOPs	Inference speed
SIT [22]	0.631 ± 0.058	22.00 ± 1.82	**8,354,112**	**36,970,987,528**	**0.040 s/patient**
PSR [23]	0.653 ± 0.049	21.90 ± 1.77	519,432,132	43,572,466,660	0.177 s/patient
X2CT-GAN [24]	0.691 ± 0.051	23.04 ± 1.70	23,564,691	227,531,788,978	0.210 s/patient
X-CTRSNet	**0.749 ± 0.043**	**24.58 ± 1.56**	22,992,322	216,830,517,469	0.182 s/patient

Table 3
X-CTRSNet gains superior accuracy in after-reconstruction 3D segmentation compared to the state-of-the-art method, with the acceptable model complexity.

	Dice (%)								Parameters number	FLOPs	Inference speed
	C1	C2	C3	C4	C5	C6	C7	Mean			
3D UNet [31]	68.65 ± 12.51	72.50 ± 10.30	73.29 ± 10.18	72.83 ± 12.14	68.08 ± 14.48	71.88 ± 11.85	76.67 ± 8.98	71.98 ± 11.49	1,755,416	601,140,127,338	0.350 s/patient
DSN [34]	71.82 ± 11.16	82.33 ± 9.26	77.44 ± 8.67	72.97 ± 11.45	73.98 ± 13.24	80.24 ± 7.32	82.61 ± 6.03	77.34 ± 9.59	7,028,344	**513,911,708,845**	**0.269 s/patient**
DenseBiasNet [35]	72.75 ± 10.14	81.24 ± 9.88	81.35 ± 6.15	79.44 ± 9.38	76.59 ± 10.45	79.54 ± 7.83	81.25 ± 6.97	78.88 ± 8.69	**1,568,820**	756,871,554,273	0.446 s/patient
CS^2-Net [36]	72.94 ± 9.65	80.94 ± 9.23	77.97 ± 14.59	77.61 ± 13.99	78.72 ± 10.21	79.96 ± 6.80	82.56 ± 5.60	78.67 ± 10.01	5,831,628	594,944,273,005	0.329 s/patient
ConResNet [37]	71.62 ± 10.98	81.20 ± 9.24	79.84 ± 8.79	79.71 ± 9.86	78.11 ± 10.70	79.93 ± 7.59	83.02 ± 5.38	79.06 ± 8.93	19,175,488	1,583,526,133,185	0.780 s/patient
X-CTRSNet	**74.71 ± 8.39**	**83.25 ± 7.21**	**81.89 ± 5.83**	**80.15 ± 8.42**	**78.37 ± 9.75**	**80.91 ± 6.45**	**83.80 ± 4.88**	**80.44 ± 7.27**	4,410,136	658,383,183,305	0.410 s/patient

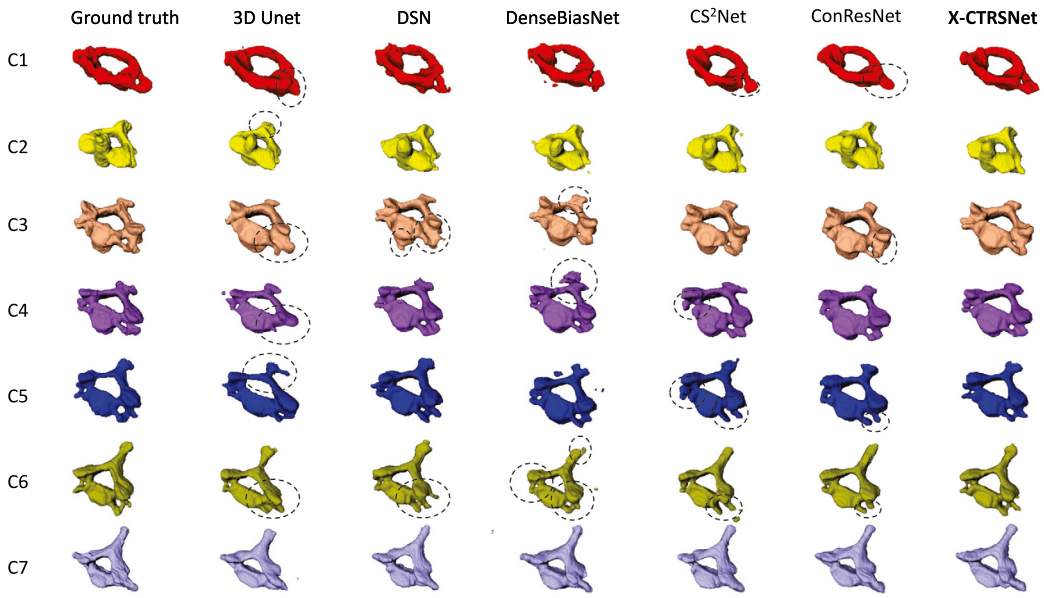

Fig. 7. X-CTRSNet shows superiority to precisely segment C-vertebra from the interactively reconstructed 3D CT. For the cervical vertebra that has multi-scale structure components, it still makes robust segmentation. As the foramen transversarium, spinous process, and vertebral body of the multi-scale structure components (circled by dotted line) cause difficulties to the compared ones, X-CTRSNet precisely segment them for the morphology extraction.

for one patient on a laptop with one Nvidia RTX 3080 GPU and an Intel i9 CPU. As can be seen, our method just takes less than a second for processing, so that remarkably saves time in clinical 3D CT imaging and analysis, as well as reduces unwanted repetitive radiation of excessive examination, especially for the triage of emergency department. (2) Besides, in clinical applications, the accuracy of the method is a more important priority [38]. The results of the accuracy comparison show that the performance of our method is significantly better than those of other known methods, gaining 13.78% improvements in SSIM for accurate anatomical structure, and increasing PSNR by 2.27*dB* for clearly readable imaging, as well as improving the average Dice with 3.25% for precise segmentation. Especially compared with SIT and DSN which have the lowest model complexities for the reconstruction and the 3D segmentation, respectively, our method achieves 18.7% higher SSIM, 2.58*dB* higher PSNR and 3.10% higher Dice to achieve the best model accuracy. Combining both the accuracy and complexities of our method, our method has great potential to effectively and quickly make 3D CT reconstruction and segmentation directly from 2D X-ray images in clinical.

4. Conclusion

In this paper, we propose X-CTRSNet, the first powerful work to simultaneously and accurately enable 3D C-vertebra CT reconstruction and segmentation directly from 2D X-ray images. The method is innovatively achieved by the following components: **(1)** SpaDRNet for the overlapped anatomy decomposition and reconstructing into the pathological information detailed 3D CT; **(2)** MulSISNet for the multi-scale stereo structure extraction and the further segmentation on the reconstructed CT, where the shape constrains are interpretively fed back; and **(3)** RSC Learning for the reconstruction–segmentation consistency in the interactive multi-tasks. Extensive experiments on reconstruction and segmentation reveal "once is enough" with X-CTRSNet to improve the diagnosis efficiency of 2D X-ray imaging and avoid the repetitive radiation of overtreatment in clinical.

CRediT authorship contribution statement

Rongjun Ge: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Yuting He:** Formal analysis, Investigation. **Cong Xia:** Resources, Data curation. **Chenchu Xu:** Methodology. **Weiya Sun:** Validation. **Guanyu Yang:** Formal analysis. **Junru Li:** Validation. **Zhihua Wang:** Validation. **Hailing Yu:** Validation. **Daoqiang Zhang:** Supervision, Project administration. **Yang Chen:** Project administration, Funding acquisition. **Limin Luo:** Supervision. **Shuo Li:** Supervision, Conceptualization. **Yinsu Zhu:** Resources, Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the Fundamental Research Funds for the Central University, China (No. NS2021067); the National Natural Science Foundation, China (No. 62101249, 61871117, 62171123 and 81871444); the China Postdoctoral Science Foundation (No. 2021TQ0149); the Natural Science Foundation of Jiangsu Province (No. BK20210291); the State's Key Project of Research and Development Plan (No. 2017YFC0109202, 2018YFA0704102).

References

[1] S. Ehara, G.Y. El-Khoury, C.R. Clark, Radiologic evaluation of dens fracture. Role of plain radiography and tomography, Spine 17 (5) (1992) 475–479.
[2] C. Bach, I. Steingruber, S. Peer, R. Peer-Kühberger, W. Jaschke, M. Ogon, Radiographic evaluation of cervical spine trauma, Arch. Orthop. Trauma Surg. 121 (7) (2001) 385–387.
[3] K. Ofori, S.W. Gordon, E. Akrobortu, A.A. Ampene, E.O. Darko, Estimation of adult patient doses for selected X-ray diagnostic examinations, J. Radiat. Res. Appl. Sci. 7 (4) (2014) 459–462.
[4] D.J. Brenner, E.J. Hall, Computed tomography-an increasing source of radiation exposure, N. Engl. J. Med. 357 (22) (2007) 2277–2284.
[5] G.Y. El-Khoury, M.H. Kathol, W.W. Daniel, Imaging of acute injuries of the cervical spine: value of plain radiography, CT, and MR imaging, AJR Am. J. Roentgenol. 164 (1) (1995) 43–50.
[6] K.C. Kim, H.C. Cho, T.J. Jang, J.M. Choi, J.K. Seo, Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation, Comput. Methods Programs Biomed. (2020) 105833.
[7] M.S. Kang, J.W. Lee, H.Y. Zhang, Y.E. Cho, Y.M. Park, Diagnosis of cervical OPLL in lateral radiograph and MRI: is it reliable? Korean J. Spine 9 (3) (2012) 205.
[8] M. Yamazaki, T. Akazawa, A. Okawa, M. Koda, Usefulness of three-dimensional full-scale modeling of surgery for a giant cell tumor of the cervical spine, Spinal Cord 45 (3) (2007) 250–253.
[9] S.A. Stratemann, J.C. Huang, K. Maki, D.C. Hatcher, A.J. Miller, Evaluating the mandible with cone-beam computed tomography, Am. J. Orthod. Dentofacial Orthop. 137 (4) (2010) S58–S70.
[10] T. Izumi, T. Hirano, K. Watanabe, A. Sano, T. Ito, N. Endo, Three-dimensional evaluation of volume change in ossification of the posterior longitudinal ligament of the cervical spine using computed tomography, Eur. Spine J. 22 (11) (2013) 2569–2574.
[11] C. Mouhanna-Fattal, M. Papadopoulos, J. Bouserhal, A. Tauk, N. Bassil-Nassif, A. Athanasiou, Evaluation of upper airway volume and craniofacial volumetric structures in obstructive sleep apnoea adults: a descriptive CBCT study, Int. Orthod. 17 (4) (2019) 678–686.
[12] B. Qiu, J. Guo, J. Kraeima, H.H. Glas, R.J. Borra, M.J. Witjes, P.M. van Ooijen, Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network, Phys. Med. Biol. 64 (17) (2019) 175020.
[13] I. Barngkgei, E. Joury, A. Jawad, An innovative approach in osteoporosis opportunistic screening by the dental practitioner: the use of cervical vertebrae and cone beam computed tomography with its viewer program, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 120 (5) (2015) 651–659.
[14] B. Müller, D. Evangelopoulos, K. Bias, A. Wildisen, H. Zimmermann, A.K. Exadaktylos, Can S-100B serum protein help to save cranial CT resources in a peripheral trauma centre? A study and consensus paper, Emerg. Med. J. 28 (11) (2011) 938–940.
[15] R. Ge, G. Yang, y. Chen, L. Luo, C. Feng, H. Zhang, S. Li, PV-LVNet: Direct left ventricle multitype indices estimation from 2D echocardiograms of paired apical views with deep neural networks, Med. Image Anal. 58 (2019) 101554.
[16] Z. Wu, R. Ge, M. Wen, G. Liu, Y. Chen, P. Zhang, X. He, J. Hua, L. Luo, S. Li, ELNet: Automatic classification and segmentation for esophageal lesions using convolutional neural network, Med. Image Anal. 67 (2021) 101838.
[17] Y. He, T. Li, R. Ge, J. Yang, Y. Kong, J. Zhu, H. Shu, G. Yang, S. Li, Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching, IEEE J. Biomed. Health Inf. (2021).
[18] R. Ge, G. Yang, Y. Chen, L. Luo, C. Feng, H. Ma, J. Ren, S. Li, K-Net: integrate left ventricle segmentation and direct quantification of paired echo sequence, IEEE Trans. Med. Imaging 39 (5) (2020) 1690–1702.
[19] R. Ge, T. Shen, Y. Zhou, C. Liu, L. Zhang, B. Yang, Y. Yan, J.-L. Coatrieux, S. Li, Convolutional squeeze-and-excitation network for ECG arrhythmia detection, Artif. Intell. Med. 121 (2021) 102181.
[20] D. Hu, W. Wu, M. Xu, Y. Zhang, J. Liu, R. Ge, Y. Chen, L. Luo, G. Coatrieux, SISTER: Spectral-image similarity-based tensor with enhanced-sparsity reconstruction for sparse-view multi-energy CT, IEEE Trans. Comput. Imaging 6 (2019) 477–490.
[21] G. Luo, W. Wang, C. Tam, K. Wang, S. Cao, H. Zhang, B. Chen, S. Li, Dynamically constructed network with error correction for accurate ventricle volume estimation, Med. Image Anal. 64 (2020) 101723.
[22] P. Henzler, V. Rasche, T. Ropinski, T. Ritschel, Single-image tomography: 3D volumes from 2D cranial X-Rays, in: Computer Graphics Forum, vol. 37, (2) Wiley Online Library, 2018, pp. 377–388.
[23] L. Shen, W. Zhao, L. Xing, Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning, Nat. Biomed. Eng. 3 (11) (2019) 880–888.
[24] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, Y. Zheng, X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 10619–10628.
[25] K.T. Johnson, W.N. Al-Holou, R.C. Anderson, T.J. Wilson, T. Karnati, M. Ibrahim, H.J. Garton, C.O. Maher, Morphometric analysis of the developing pediatric cervical spine, J. Neurosurg. Pediatr. 18 (3) (2016) 377–389.
[26] K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693–5703.
[27] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.
[28] R. Ge, G. Yang, C. Xu, Y. Chen, L. Luo, S. Li, Stereo-correlation and noise-distribution aware ResVoxGAN for dense slices reconstruction and noise reduction in thick low-dose CT, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 328–338.

[29] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[30] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, Springer, 2016, pp. 694–711.
[31] O. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[32] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[33] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[34] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, P.-A. Heng, 3D deeply supervised network for automated segmentation of volumetric medical images, Med. Image Anal. 41 (2017) 40–54.
[35] Y. He, G. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao, S. Zhang, et al., Dpa-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep priori anatomy, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 139–147.
[36] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A.F. Frangi, M. Akiba, J. Liu, CS^2-Net: Deep learning segmentation of curvilinear structures in medical imaging, Med. Image Anal. 67 (2021) 101874.
[37] J. Zhang, Y. Xie, Y. Wang, Y. Xia, Inter-slice context residual learning for 3D medical image segmentation, IEEE Trans. Med. Imaging 40 (2) (2021) 661–672.
[38] C. Zhang, H. Shu, G. Yang, F. Li, Y. Wen, Q. Zhang, J.-L. Dillenseger, J.-L. Coatrieux, HIFUNet: Multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning, IEEE Trans. Med. Imaging 39 (11) (2020) 3309–3320.