X-ray Dissectography Enables Stereotography

Chuang Niu[a] and Ge Wang[a]

[a]Department of Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY USA.

ABSTRACT

X-ray imaging is the most popular medical imaging technology. While x-ray radiography is rather cost-effective, tissue structures are superimposed along the x-ray paths. On the other hand, computed tomography (CT) reconstructs internal structures but CT increases radiation dose, is complicated and expensive. Here we propose "x-ray dissectography" to extract a target organ digitally from few radiographic projections for stereographic and tomographic analysis in the deep learning framework. As an exemplary embodiment, we propose a general X-ray dissectography network, a dedicated X-ray stereotography network, and the X-ray imaging systems to implement these functionalities. Our experiments show that x-ray stereography can be achieved of an isolated organ such as the lungs in this case, suggesting the feasibility of transforming conventional radiographic reading to the stereographic examination of the isolated organ, which potentially allows higher sensitivity and specificity, and even tomographic visualization of the target. With further improvements, x-ray dissectography promises to be a new x-ray imaging modality for CT-grade diagnosis at radiation dose and system cost comparable to that of radiographic or tomosynthetic imaging.

Keywords: X-ray radiography, digital tomosynthesis, computed tomography, x-ray dissectography, x-ray stereography, artificial intelligence, deep learning

1. INTRODUCTION

X-ray imaging is the first and still most popular modern medical imaging approach, which is performed by various kinds of systems. In the low-end, x-ray radiography takes a two-dimensional projective image through a patient, which is called a radiogram or radiograph. In the high end, many x-ray projections are first collected and then reconstructed into tomographic images transversely or volumetrically. Between these two extremes, digital tomosynthesis takes a limited number of projections over a relatively short scanning trajectory and infers three-dimensional features inside a patient. These x-ray imaging modes have their strengths and weaknesses. X-ray radiography is cost-effective but it produces a single projection, and has multiple oranges and tissues superimposed along x-ray paths, compromising the diagnostic performance. On the other hand, x-ray CT unravels structures overlapped in the projection domain into tomographic images in a 3D coordinate system but CT uses a much higher radiation dose, is complicated and expensive. Digital tomosynthesis is a balance between x-ray radiography and CT in terms of the number of needed projections, the information in resultant images, and the cost to build and operate the imaging system.

Reducing radiation dose and improving imaging quality and speed are the main tasks for the development of x-ray imaging technologies. As x-ray radiography has the lowest radiation dose, the fastest imaging speed, and the lowest price, researchers have been focusing on improving the radiogram quality. Currently, there are mainly two ways for this purpose: 1) suppressing interfered structures[1–4] or enhancing related structures,[1,5] and 2) generating 3D volumes.[6–8] It is well known that superimposed anatomical organs in 2D radiographs significantly complicate signal detection, such as for diagnosis of lung diseases. In early studies[9,10] model-based methods were developed to suppress ribs in chest radiographs, some of which require manually annotated bone masks. In recent years, deep learning methods[1–4] were proposed for suppression of ribs by leveraging 3D CT prior. Instead of suppressing the ribs in CXR images, Gozes and Greenspan proposed to enhance lung structures by extracting the extracted lungs first and then adding the result back with a scaling factor.[5] Generating a 3D

---

volume from a single or a few radiographs is another way to improve radiography. Ying et al. proposed X2CT that generates a 3D CT volume from a pair of orthogonal radiographs using a CycleGAN framework.[6] Recently, some methods[7,8] proposed to generate a 3D volume from single or few radiograhps.

Among the above-surveyed methods, those methods that are for suppressing/enhancing specific structures are mainly intended to improve the performance of classification[1-4] and detection[5] from a single radiograph without providing 3D information. On the other hand, although the rest of the existing methods that map 2D radiograms to 3D CT volumes achieved remarkable results, they cannot reconstruct structures accurately and reliably, since their clinical utilities have not been demonstrated so far. Particularly, it can be seen that almost all the above methods depend on the GAN framework and/or unpaired learning for 2D/3D image generation. A major potential problem is that GAN-based models tend to generate fake structures, which is a major concern in the medical imaging field.

In this study, we propose x-ray dissectography (XDT) in general and specialize it for x-ray stereotography to improve image quality and diagnostic performance. The essential idea is that we digitally extract a target organ from the original radiograph or radiogram which contains superimposed organs via deep learning, facilitating both visual inspection and quantitative analysis. Considering that radiograhs from different views contain complementary information, we design a physics-based XDT network to extract the multi-view features and transform them into a 3D space. In this way, the target organ can be synergistically analyzed in isolation from different projection angles. As a special yet important application of our XDT approach, we propose X-ray stereography that allows a reader immersively perceive the target organ from two dissected radiographs in 3D, synergizing machine intelligence and human intelligence, similar to what CT does. Biologically, stereo perception is based on binocular vision for the brain to reconstruct a 3D scene, and can be applied to see through dissected radiograms and form a 3D rendering in a radiologist in mind. In this work, we design an X-ray imaging system dedicated to this scenario. Different from our daily visual information processing, which senses surroundings with reflected light signals, radiograms are projective through an object to allow 3D concept of x-ray semi-transparent features.

To avoid fake structures, we optimize XDT neural networks in a supervised 2D-to-2D learning paradigm without using a GAN-like model. To obtain a 2D radiograph of a target organ without surrounding tissues, we can manually or automatically segment the organ in the associated CT volume first and then compute the ground-truth radiograph through projecting the dissected organ according to the system parameters. In other words, radiographs and CT images are obtained from the same patient and the same imaging system to avoid unpaired learning. We utilize a cutting-edge simulation platform, such as the popular academic[11] and industrial? simulators, for training XDT networks. These simulators can take either a clinical CT volume or a digital 3D phantom to compute a conventional x-ray radiograph, and then extracts a target organ digitally to produce the ground-truth radiograph of the organ. Our initial experimental results have shown that XDT indeed separates the lungs with faithful texture and structures, and we can perceive the extracted lungs via stereoscopic viewing with a pair of 3D glasses. Potentially, this approach can improve the diagnostic performance in lung cancer screening, COVID-19 follow-up, and other applications.

2. METHODOLOGY
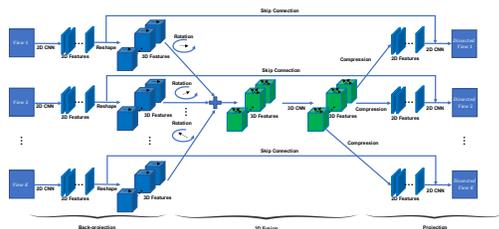
Figure 1. X-ray dissectography network (XDT-Net).
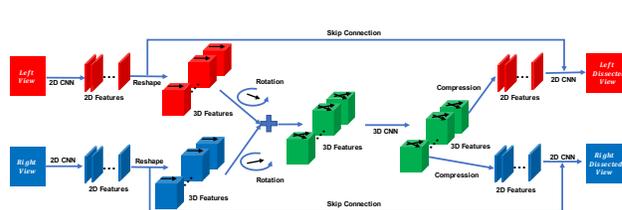
Figure 2. X-ray stereoraphic imaging network (XST-Net).

2.1 General Workflow of X-ray Dissectography

X-ray dissectography (XDT) is dedicated to transform a conventional radiogram $\boldsymbol{x} = \sum_{i=1}^{B} \boldsymbol{y}_i + \boldsymbol{y}_t$ to a projection image $\boldsymbol{y}_t$, where $\boldsymbol{y}_t$ is a projection of only a target organ, and $\sum_{i=1}^{B} \boldsymbol{y}_i$ represents the superimposed image of the B anatomical components involved in the conventional radiogram. In fact, it is an extremely ill-posed problem as we can only observe the radiograph $\boldsymbol{x}$, and impossible to obtain any analytic solution in a general setting. Fortunately, a specific organ in the human body has a fixed relative location, a strong prior on material composition, and similar patterns (such as shapes, textures, and other properties). Given this kind of knowledge, a radiologist can identify different organs in the conventional radiogram. However, the superimposed organs challenge the human in visual inspection for the target one. Considering the great progresses in deep imaging,[12–14] deep neural networks (DNNs) is used to learn such priors and extract purified radiographs as if x-rays go only through the target organ. Such DNNs can be trained with an individual or a specific population for quantitative accuracy and clinical utilities.

In this study, we propose a physics-based XDT network (XDT-Net) for separating a target organ from more than one views, as shown in Fig. 1. Note that various organs may be separated using this framework in different combinations, depending on specific applications. The XDT-Net consists of the three modules: 1) a back-projection module, 2) a 3D fusion module, and 3) a projection module. The back-projection module maps 2D radiographs to 3D features, like a tomographic back-projection process. It consists of k 2D convolutional neural networks (CNNs) followed by reshape operators, where k is the number of input views, each CNN is applied to a specific view, and different CNNs have the same architecture but trainable parameters may be optimized differently. The fusion module integrates the information from all views in the 3D feature space. It first aligns the 3D features of different views by rotation according to their projection angles, and then combines them by a 3D CNN. The projection module predicts each radiograph containing only the target organ. It first squeezes each 3D feature volume to a 2D feature map along a given angle, and then the 2D CNN takes the 2D features from both the squeeze operator and the back-projection module to predict the radiograph of the target organ.

2.2 Specific Embodiment for X-ray Stereography

We perceive the world in 3D thanks to binocular vision. Given binocular disparity, the human brain is capable of sensing the depth in the scene. Inspired by this amazing fact, here we investigate X-ray stereography (XST) with two radiograms of an isolated organ. When inspecting the human body with x-rays, organs with large linear attenuation coefficients will overwhelm the ones with small attenuation coefficients in radiograms. As a result, it is difficult to discern subtle changes in internal organs due to the superimposition of multiple organs, significantly compromising stereopsis. With our proposed XDT-Net and XST-Net, we can integrate machine intelligence for target organ dissection and human intelligence for stereographic perception so that a radiologist can perceive a target organ in 3D with details much more vivid than in 2D, potentially improving diagnostic performance.

To enable XST of a specific organ, we adapt the XDT-Net to the XST-Net as shown in Fig 2. The XST-Net also consists of the same three modules: the backprojection module, the 3D fusion module, and the projection module. Each module of the XST-Net shares the same network architecture as that of the XDT-Net but needs to be adapted for stereo viewing. First, the backprojection module of the XST-Net takes two radiographs as inputs, which are two images into our eyes. Second, in reference to the view angles of two eyes, the 3D fusion module uses a different rotation center to align 3D features from two branches appropriately. Our proposed XST imaging system is described in Subsection 2.3. Third, the projection module translates the merged 3D feature first and then squeezes it to 2D feature maps according to the human reader's viewing angles. Finally, two dissected radiographs are respectively sent to the left and right eyes through a pair of 3D glasses for stereoscopy.

2.3 Design on X-ray Dissectography and Stereography

We assume that radiographs from sufficiently many different angles can be obtained in the network training stage such that image volumes can be reconstructed. The traditional cone-beam CT system, as shown in Fig. 3 (a), serves this purpose. Then, many pairs of conventional radiographs and the counterparts of the target-only radiographs can be obtained from a reconstructed CT volume and a segmented organ in the reconstructed CT volume respectively. In the testing stage, the same XDT system only needs to generate few radiographs at
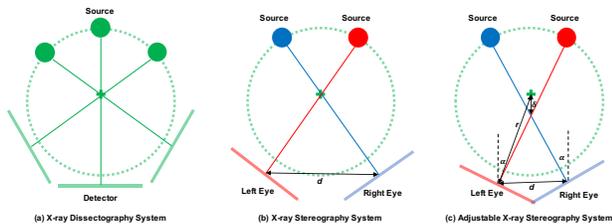
Figure 3. X-ray imaging system configuration to facilitate radiographic, stereographic and tomographic analysis on a digitally isolated organ or tissue.
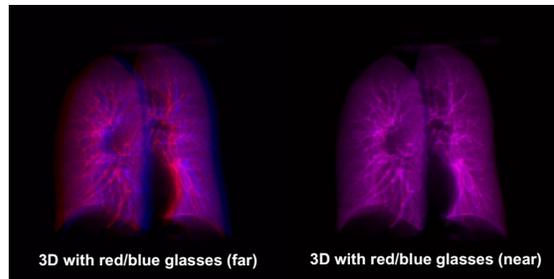
Figure 4. XST image pair and rendering for stereo-viewing.

any angles for the trained XDT-Net to extract the corresponding radiograms of the target organ alone, without surround tissues, for much-improved visual inspection and quantitative analysis. To achieve x-ray stereopsis, we design an XST imaging system, as shown in Fig. 3 (b) and (c), where each source is regarded as an eye while the projection through the body is recorded on the opposite detector. In a simple setting, we directly take two radiograms from the XDT system so that the distance between eyes is d, as shown in Fig. 3 (b). In this case, the center X-rays from the source positions intersect at the object center. For the adaption to different applications and readers, we further design an adjustable XST system in Fig. 3 (c). There are two parameters of the system to control the offset between the two eyes and the viewing angle from a pre-specified principal direction. In Fig. 3 (c), red and blue dots denote the left and right eyes, the red and blue plates are the corresponding detectors, and green cross is viewed as the object center. The distance between two eyes is d, the distance between the source and the object center is r, and the angle between center X-ray and the pre-specified reference direction is α for both eyes. Thus, the intersection point of two center X-rays is translated from the object center along the vertical direction. The distance offset δ can be computed as $\delta = \frac{d}{2\tan(\alpha)} - \sqrt{r^2 - (d/2)^2}$, which is used to adjust the rotation center for XST-Net as introduced in Subsection 2.2.

In practice, the XST system for inspecting different organs may require different geometric parameters. Both XDT and XST systems can be implemented in various ways such as with robotic arms[15] so that the geometric parameters can be easily set to match a reader's preference.

3. EXPERIMENTAL RESULTS

3.1 XDT and XST Simulation

Here we used a clinical CT dataset, denoted as CT-lung, and simulated radiograms in cone-beam geometry. Specifically, 50 reconstructed CT volume of patients were selected, including 10 patients from[16] and 40 patients from.[17] The data from[16] provide the 3D lung masks. Hence, we can simulate the paired radiograms with and without lung masks.[18] Note that before performing the cone-beam projection, the patient bed in the CT volume was first masked out in a semi-automatic manner. Since the lung masks provided in[17] are not consistent with those provided in,[16] we trained a semantic segmentation UNet with the annotated data in[16] to identify the body region for removal of the patient bed and segment the lung region slice-by-slice for the data in[17] consistently. When 2D radiograms were synthesized, we rotated the patient CT volume from $0°$ to $180°$, where the angle of $0°$ is the frontal view. Totally, we obtained 9,000 pairs of radiograms, and the image size is 320×320.

3.2 XDT Results

We first evaluated the effectiveness of the proposed XDT-Net on the CT-lung dataset. In the current experiments, we focused on simultaneously dissecting two radiograms at orthogonal angles. To be more specific, the region of interest was first segmented from 2D radiograms before forwarding to the XDT-Net. In this way, the task of XDT-Net is purified to improve dissection results. For this purpose, we trained a segmentation model to identify the region of interest. The target mask for training this segmentation model can be easily obtained by thresholding the radiogram of isolated lungs, where the threshold was empirically set to 0.01 in the unit of linear attenuation coefficient.
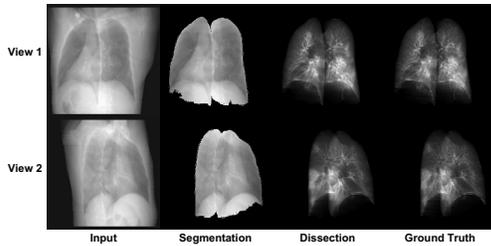
Figure 5. XDT testing results from two new orthogonal views of the same patient.
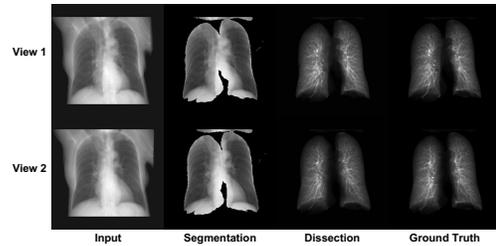
Figure 6. XST testing results from two new stereo views of the same patient.

The testing results are shown in Fig. 5, where radiographs of the same patient were collected from different angles. From the first to the fourth columns are the input radiograms, the segmentation results, the dissection results, and the ground-truth respectively. The first and the second rows present two orthogonal projections respectively. The visual inspection shows that the dissected radiograms are very close to the ground-truth in terms of detailed structures despite being slightly smoother. The blurring effect may be due to noise reduction.[19] Compared to normal radiograms, the dissected radiograms remove irrelevant surrounding structures, highlight the target organ, and potentially improve the diagnosis performance.

3.3 XST Results

Then, we evaluated the feasibility of the proposed XST-Net for X-ray stereoscopic imaging on the CT-lung dataset. We first evaluated the joint dissection results from two stereo radiograms collected at two new angles of the same patient and then generalized the stereo-imaging technology to different patients. Our representative results are shown in Fig. 6, showing XST-Net achieves very promising results for stereo views. In addition, we have found that the dissection networks are quite robust to segmentation results, geometric parameters, and image noise. Finally, we generated 3D perception by overlapping the left-eye and right-eye images in red and blue channels and then viewing both through a pair of red/cyan glasses. Fig. 4 shows stereo images and two 3D images adjusted with different geometric parameters, as discussed in Subsection 2.3. Readers can enjoy watching the 3D lungs through red/cyan glasses (you may need some visual adaptation to see the 3D scene).

4. CONCLUSION

In conclusion, we have proposed the x-ray dissectography (XDT) and x-ray stereography (XST) systems and methods for improving the utilities of conventional X-ray radiography and digital tomosynthesis. The proposed XDT and XST can dissect a target organ type from X-ray radiograms with deep learning. The experimental results clearly demonstrate the feasibility and potential utility of the proposed imaging technology. In the future, we will continue improving the network model and producing clinically relevant results systematically. Hopefully, the proposed XDT and XST techniques empowered by artificial intelligence may open a new door for traditional X-ray radiography to have new impacts on healthcare.

REFERENCES

[1] Li, Z. and et al., "Encoding ct anatomy knowledge for unpaired chest x-ray image decomposition," (2019).

[2] Peng, C. and et al., "Xraysyn: Realistic view synthesis from a single radiograph through ct priors," (2020).

[3] Li, H. and et al., "High-resolution chest x-ray bone suppression using unpaired ct structural priors," *IEEE TMI* **39**(10), 3053–3063 (2020).

[4] Han, L. and et al., "Gan-based disentanglement learning for chest x-ray rib suppression," (2021).

[5] Gozes, O. and Greenspan, H., "Lung structures enhancement in chest radiographs via CT based FCNN training," *arXiv* (2018).

[6] Ying, X. and et al., "X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks," in [*CVPR*], (2019).

[7] Shen, L., Zhao, W., and Xing, L., "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature Biomedical Engineering* **3**(11), 880–888 (2019).

[8] Shen, L. and et al., "A geometry-informed deep learning framework for ultra-sparse 3d tomographic image reconstruction," (2021).

[9] Hogeweg, L. and et al., "Suppression of translucent elongated structures: Applications in chest radiography," *IEEE TMI* **32**(11), 2099–2113 (2013).

[10] Wu, D. and et al., "A learning based deformable template matching method for automatic rib centerline extraction and labeling in ct images," 980–987 (2012).

[11] Abadi, E. and et al., "Dukesim: A realistic, rapid, and scanner-specific simulation framework in computed tomography," *IEEE TMI* **38**, 1457–1465 (2019).

[12] Niu, C. and et al., "Low-dimensional manifold constrained disentanglement network for metal artifact reduction," *IEEE TRPMS* , 1–1 (2021).

[13] Niu, C., Li, M., and Wang, G., "Multi-window learning for metal artifact reduction," in [*Developments in X-Ray Tomography XIII*], **11840**, 1184015 (2021).

[14] Niu, C. and et al., "Noise entangled gan for low-dose ct simulation," *arXiv preprint arXiv:2102.09615* (2021).

[15] Fieselmann, A. and et al., "Twin robotic x-ray system for 2D radiographic and 3D cone-beam CT imaging," in [*Medical Imaging 2016: Physics of Medical Imaging*], **9783**, 128 – 133 (2016).

[16] Morozov, S. P. and et al., "Mosmeddata: Chest CT scans with COVID-19 related findings dataset," *CoRR* (2020).

[17] Bilic, P. and et al., "The liver tumor segmentation benchmark (lits)," *CoRR* (2019).

[18] van Aarle, W. and et al., "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express* **24**(22), 25129–25147 (2016).

[19] Niu, C. and Wang, G., "Noise2sim - similarity-based self-learning for image denoising," *CoRR* **abs/2011.03384** (2020).