Jung Y, Kong J, Sheng B *et al*. A transfer function design for medical volume data using a knowledge database based on deep image and primitive intensity profile features retrieval. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 39(2): 320-335 Mar. 2024. DOI: 10.1007/s11390-024-3419-7

A Transfer Function Design for Medical Volume Data Using a Knowledge Database Based on Deep Image and Primitive Intensity Profile Features Retrieval

Younhyun Jung[1, 2], Jim Kong[2], Bin Sheng[3, *] (盛　斌), *Member, CCF, ACM, IEEE*, and Jinman Kim[2]

[1] *School of Computing, Gachon University, Seongnam 13120, Korea*

[2] *School of Computer Science, The University of Sydney, Sydney 2006, Australia*

[3] *Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

E-mail: younhyun.jung@gachon.ac.kr; jkon2561@uni.sydney.edu.au; shengbin@sjtu.edu.cn; jinman.kim@sydney.edu.au

Received May 21, 2023; accepted January 18, 2024.

Abstract Direct volume rendering (DVR) is a technique that emphasizes structures of interest (SOIs) within a volume visually, while simultaneously depicting adjacent regional information, e.g., the spatial location of a structure concerning its neighbors. In DVR, transfer function (TF) plays a key role by enabling accurate identification of SOIs interactively as well as ensuring appropriate visibility of them. TF generation typically involves non-intuitive trial-and-error optimization of rendering parameters, which is time-consuming and inefficient. Attempts at mitigating this manual process have led to approaches that make use of a knowledge database consisting of pre-designed TFs by domain experts. In these approaches, a user navigates the knowledge database to find the most suitable pre-designed TF for their input volume to visualize the SOIs. Although these approaches potentially reduce the workload to generate the TFs, they, however, require manual TF navigation of the knowledge database, as well as the likely fine tuning of the selected TF to suit the input. In this work, we propose a TF design approach, CBR-TF, where we introduce a new content-based retrieval (CBR) method to automatically navigate the knowledge database. Instead of pre-designed TFs, our knowledge database contains volumes with SOI labels. Given an input volume, our CBR-TF approach retrieves relevant volumes (with SOI labels) from the knowledge database; the retrieved labels are then used to generate and optimize TFs of the input. This approach largely reduces manual TF navigation and fine tuning. For our CBR-TF approach, we introduce a novel volumetric image feature which includes both a local primitive intensity profile along the SOIs and regional spatial semantics available from the co-planar images to the profile. For the regional spatial semantics, we adopt a convolutional neural network to obtain high-level image feature representations. For the intensity profile, we extend the dynamic time warping technique to address subtle alignment differences between similar profiles (SOIs). Finally, we propose a two-stage CBR scheme to enable the use of these two different feature representations in a complementary manner, thereby improving SOI retrieval performance. We demonstrate the capabilities of our CBR-TF approach with comparison with a conventional approach in visualization, where an intensity profile matching algorithm is used, and also with potential use-cases in medical volume visualization.

Keywords data retrieval, direct volume rendering, ray analysis, transfer function

1 Introduction

Modern imaging modalities, such as computed tomography (CT) or magnetic resonance (MR), is volumetric in nature, e.g., the shape, size, and location of structures can be natively described in 3D (three-di-

Regular Paper

Special Section of CGI 2023

This work was supported by the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute under Grant No. HI22C1651, the National Research Foundation of Korea (NRF) under Grant No. 2021R1F1A1059554, and the Culture, Sports and Tourism Research and Development Program through the Korea Creative Content Agency Grant funded by the Ministry of Culture, Sports and Tourism of Korea under Grant No. RS-2023-00227648.

*Corresponding Author

mensional) space. They are represented as a stack of 2D (two-dimensional) image slices that collectively show volumetric information. In conventional interpretation approaches, users manually navigate through an entire volume, slice-by-slice, and then mentally reconstruct the volumetric information derived from 2D image slices. In an attempt to complement the 2D visualization, 3D rendering, such as direct volume rendering (DVR), is used to provide volumetric visualization of the data.

Fig.1 shows the advantages of a 3D DVR for a human CT volume compared with its 2D counterpart cross-sectional views, where the DVR provides an overview of the entire data, the shape of the kidneys, and its spatial relationship with the neighboring spine and rib cage. Such visualization demonstrates the complementary view that is offered by the 3D DVR to the 2D cross-sectional views.

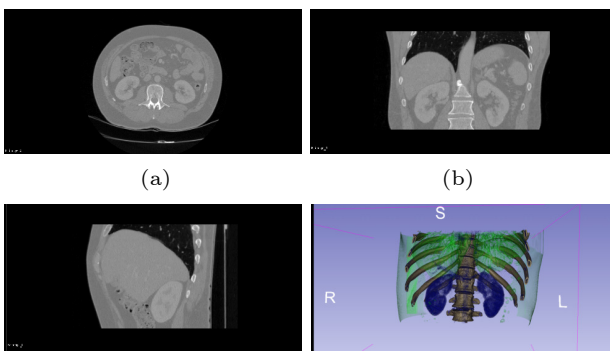

Fig.1. Quarter-view visualization example of complementary 2D and 3D for a human CT volume. The three 2D cross-sectional views are shown in (a) axial, (b) coronal, and (c) sagittal views, respectively. 3D DVR of the same volume is shown in (d). We use 3D Slicer software[1] to generate the visualizations.

There are many challenges for DVR to be more broadly employed for volumetric imaging. A major technical issue lies with the data-specific visualization requirements, i.e., to identify and emphasize different structures of interest (SOIs) in a volume. Transfer functions (TFs) play a key role in this[2]; for instance, in a conventional one-dimensional (1D) TFs that associate data density (intensity) with rendering parameters (color and opacity), a user must specify the intensity ranges (to identify SOIs) and then assign color and opacity values to the selected range (to emphasize the SOIs). As such, it involves non-intuitive trial-and-error adjustments of the TF parameters until the user "discovers" the desired visualizations, which is time-consuming and inefficient[2]. There have been several researches that have been directed at techniques to improve TF design such as dominant approaches including multi-dimensional TFs[3–6], image-centric[7, 8], and automated parameter optimization approaches[9–12].

An alternative technique to TF design is with the use of "knowledge" from similar cases[13, 14], when compared with the other dominant approaches solely relying on input data. Pioneering work by Marks *et al.*[13] and Guo *et al.*[14] introduced knowledge databases that consist of a collection of pre-designed TFs by domain experts. In the above work, instead of generating effective TFs from scratch, a user manually navigates the pre-designed TF cases in the database, with the visual aid of paired DVRs, and selects the most suitable one for the input volume. Although these researches demonstrated the feasibility of knowledge databases in TF design, they still require manual navigation of pre-designed TFs in the knowledge databases. In addition, pre-defined TFs from the knowledge database may not be always optimal for the input volume.

One solution to improve knowledge databases for TF design is to automate the manual searching process. Content-based retrieval (CBR) is a technique that enables automated searching of similar cases against an input query by using content matching. The term "content" refers to any features that can be derived from data itself for use in representing the data. For example, with image data, conventional low-level features include intensity, texture, and color, with recent work reliant on deep learned features to better describe complex patterns and high-level semantics[15]. CBRs have shown great success in general images, as well as with medical images, for their ability to effectively identify and retrieve similar cases from large databases[16].

There has been a paucity of work on the use of CBR for TF design. As a pioneering and only relevant research, Kohlmann *et al.*[17] made use of an intensity profile, as content, to represent an SOI along a viewing ray. In their quarter-view visualization approach, a user-defined ray in 3D DVR is used to link a slice position of 2D cross-sectional views; the slice position is computed by automatically matching the intensity profile of the input ray query to pre-calculated profile templates for different SOIs in the knowledge database. As such, their use of CBR is highly optimized to provide complimentary 2D cross-sectional views for a single SOI, and is not suitable for the TF design that associates with multiple SOIs. The

proposed intensity profile matching only relies on low-level primitive features and may not be sufficient to identify and retrieve multiple SOIs. A conventional Euclidean distance (ED) measure used, in addition, only considers voxel pairs that are exactly in the same position and cannot match subtle differences (e.g., the two same SOIs but with different lengths).

In this work, we propose a new TF design approach, CBR-TF, using a knowledge database that is searched using a new CBR method, which we refer to as CBR-TF. Instead of pre-defined TFs[17], our knowledge database consists of a collection of volumes with SOI labels. Given an input volume, CBR-TF automatically retrieves similar cases by matching the content pair based on a volumetric image feature. The retrieved cases enable us to identify SOIs of the input, which are then used to generate a component-based 1D TF[7]. The main technical contributions of the paper can be summarized as follows.

• We propose a "triplet input query" (TIQ) to formulate a volumetric image feature from an input volume. A TIQ comprises a user-defined ray of SOIs and two co-planar images to the ray in 3D coordinates.

• For the two co-planar images, we propose the use of a pre-trained convolutional neural network (CNN)[18] to obtain image feature representations that carry high-level of regional spatial semantics. We further make use of an intensity profile along the SOIs for locating them, using dynamic time warping (DTW) technique[19] to address subtle alignment differences between similar profiles.

• We propose a two-stage CBR method to enable the use of the two different types of CNN and DTW in a complementary manner through a rank-based sequential combination of the two individual retrievals.

• We adopt a stroke-based image-centric approach[7, 8] for simple and intuitive user TIQ interactions. Using a typical quarter-view visualization (as in Fig.1), in our user interface, a user only needs to draw a single ray (on 2D cross-sectional views) or select a single point (on 3D DVR) to derive the TIQ.

• For situations where particular SOIs are required to be emphasized, a visibility-based TF parameter optimization[9, 10] can be applied to TF design.

• We demonstrate the application of our CBR-TF approach using a public medical imaging data repository of 3D Image Reconstruction for Comparison of Algorithm Database (3D-IRCADb-01)[20].

2 Related Work

2.1 Multi-Dimensional TF Designs

TF design has been advanced from traditional intensity-based 1D TFs toward multi-dimensional TFs to facilitate the identification of SOIs. A pioneer study by Kindlmann *et al.*[3] proposed 2D TFs using intensity with its first or second order derivatives. Projecting these additional derived features as secondary dimensions on the TFs improves SOIs exploration; the new dimension typically acts as additional "indicators" to guide users during TF design to help identify gradient information along the intensity (representing SOI) and visually emphasize the boundaries of SOIs. Similarly, Correa *et al.*[4] identified SOIs according to the local size. Some work used more than two dimensions (features), for example, the use of 20 local texture features by Caban *et al.*[5], to enable greater differentiation among SOIs. As the number of TF dimensions increases, there is a greater need for manual optimizations in a complex multi-dimensional space[2]. The aim of this work is to substantially reduce the manual optimization by introducing automatic identification of SOIs based on a CBR process.

2.2 Image-Centric TF Designs

Image-centric approaches make TF design intuitive by allowing users to identify and optimize SOIs directly on an initial DVR visualization through manual gestures. Ropinski *et al.*[7] enabled intuitive selection of SOIs by users drawing one or more strokes directly onto the DVR visualization near the silhouette of the SOIs. Then based on the stokes(s), an intensity histogram analysis is done to identify the intended SOIs in the intensity 1D TF widget. Guo *et al.*[8] manipulated the appearance of the intended SOIs through high-level painting metaphors, e.g., eraser, contrast, and peeling. They used a 2D graph-cut algorithm to identify SOIs. Users, however, often have difficulty in precisely inferring SOIs from their interaction, especially, when the SOIs are fuzzy, semi-transparent, and multi-layered in the initial visualization. As such, these image-centric approaches may require manual and repetitive user interactions for desired visualizations. In this work, our approach requires a user to indicate SOIs in an image slice, not the visualization where a stack of image slices is rendered, thereby making the inference intuitive and accurate.

2.3 Automated Parameter Optimization for TF Designs

Some investigators[9–11] focused on using parameter optimization algorithms to lessen the need for extensive user TF interaction. They assume if SOIs in a volume were known, the initial TF parameters could be automatically adjusted. These SOI-based TF optimizations ensure that the visibility of the SOIs is maintained by reducing the opacity parameters of other structures/voxels that are occluding the SOIs. They generally rely on greedy implementation that searches for locally optimal solutions, due to the unavailability of exhaustive search of huge TF parameters space. Such greedy solutions are sensitive to the initial parameters, which are often not properly defined by the users. A suboptimal initialization may result in the undesirable visualization of a local minimum. The CBR-TF approach proposed in this work aims to generate a TF that properly represents SOIs (in our case, along a viewing ray). CBR-TF can be complementary with the automated optimization in a way that the TF result can be used as the proper initial parameters, if further optimization is required.

2.4 Knowledge Databases and CBR in DVR Visualization

Marks *et al.*[13] firstly generated a knowledge database with a collection of pre-designed TFs by domain experts for TF design. This database is comprised of pairs of a DVR and its corresponding TF. It is then used manually by the users to search through the TF cases that are most suitable to the input volume, with aid of paired DVRs. Guo *et al.*[14] structured their knowledge database into a 2D search representation of TF cases via multi-dimensional reduction (MDS) for effective manual knowledge database navigation. In this search space, similar TF cases are grouped into clusters via the density field of their MDS. Although these researches demonstrate the potential of how to use knowledge databases in TF design, they only partially reduce the workload to generate TFs as they require manual exhaustive navigations of the TF cases from the knowledge database, as well as the potential need for additional adjustment of the selected TFs to be best mapped to the input volume. Apart from TF design, Kohlmann *et al.*[17] investigated the use of a knowledge database to provide complementary visualization in 2D cross-sectional views and 3D DVR. They introduced CBR concept to automate search through the knowledge database, where the intensity profiles are used to represent the "content", such as an SOI along the profile. This is the only work that demonstrates the use of the CBR-based knowledge database for DVR visualization. Our motivation for this work is to introduce the CBR concept to TF design.

3 Proposed CBR-TF Approach

3.1 Overview

An overview of our CBR-TF approach is shown in Fig.2 exemplified with an input volume of the human upper-abdomen. We construct, as an offline process, a knowledge database using a set of labeled volumes (Subsection 3.2). In the quarter-view visualization of the input volume with 2D cross-sectional views and a 3D DVR, a user draws a line (a ray) to visualize certain SOIs in any of the 2D views. With the user-selected ray, we generate a TIQ comprising a pair of co-planar images to the ray and its intensity profile. The TIQ is then used to query the knowledge database (Subsection 3.3). We use the SOI labels provided from the retrieved results to generate an intensity-based 1D TF for the SOIs (Subsection 3.4), which can be further optimized with the input volume to make sure that particular SOIs prioritize the visibility in the final 3D DVR (Subsection 3.5).

3.2 Knowledge Database Construction

We use a set of volumes where every voxel has two information: intensity value and SOI label (name of structures in our setting) it represents. A primary element in our knowledge database is a ray. Ray extraction from a volume is illustrated in Fig.3. We choose a point on a plane surface of the cube (volume) aligned with two of the three primary axes (x, y, z) and cast a ray down to the third axis. We derive ray representations comprising four items from the extracted ray: 1) an image pair co-planar to the ray according to the primary axes; 2) a profile of intensity values along the ray; 3) the corresponding profile of the SOI labels; and 4) location coordinates of entry- and exit-point of the ray. We uniformly extract multiple rays along a dimension (axis), with the consistent number of rays from all the three axes. Empty backgrounds of a volume are excluded during the

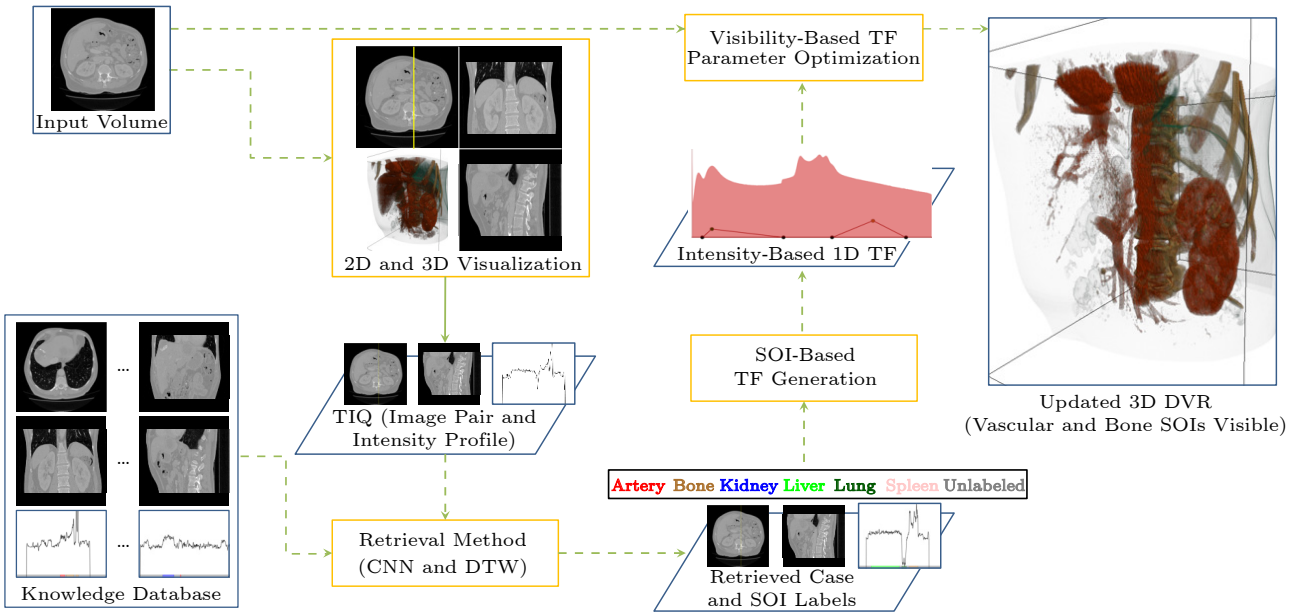

Fig.2. Overview of our CBR-TF approach.

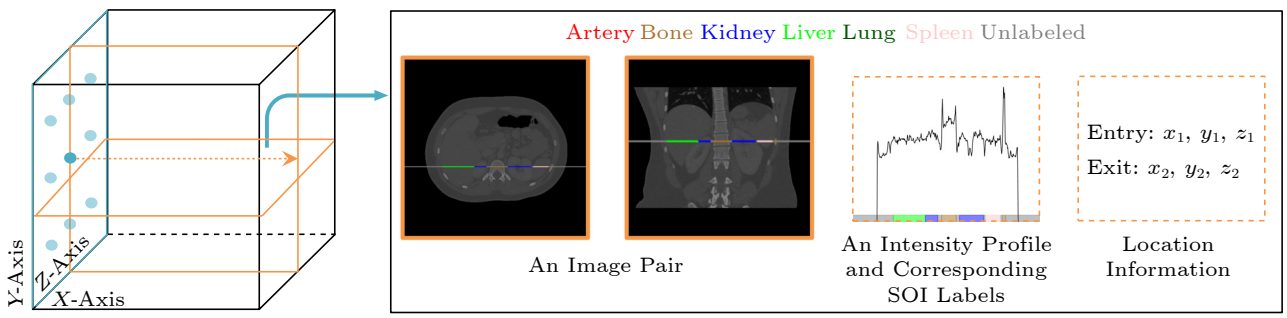

Fig.3. Ray extraction example and the ray representations (four items) used for our knowledge database construction. A point (cyan circle) is selected on a plane surface (cyan lines) of a volume (black line cube) in the two primary axes (*Y*- and *Z*-axes). A ray (orange arrow dot line) is determined by casting it down to the third axis (*X*-axis). The four components were obtained as an image pair co-planar to the ray in *X*-*Z* axes and *X*-*Y* axes (two rectangles with orange lines), and an intensity and SOI label profile along the ray with its location coordinates.

ray extraction. A volume consists of 192 rays in total and 64 rays from each of the three axes. The extraction interval, i.e., the number of rays, is experimentally determined to reduce the likelihood of extracting highly similar rays as well as lowering the computational complexity of our CBR process.

3.3 Two-Stage CBR Method

For a TIQ from a user-selected ray q: a pair of images q_i and an intensity profile q_p, we retrieve the most similar counterpart c from the knowledge database by calculating the similarity distances between the triplet pairs (q and c) in a two-stage manner. We first match q_i with c_i. Among the top N retrieval results based on the image matching, as the second stage, we re-rank them according to the similarity of intensity profiles (q_p and c_p).

CNN-Based Image Matching. We use a well-established CNN, AlexNet[18], which is pre-trained using 1.2 million natural images from the ImageNet database[18]. The architecture of AlexNet is shown in Fig.4. It consists of total seven trainable layers with 60 million parameters; the first five are convolutional layers, and the remaining two are fully-connected layers. The output of the last fully-connected layer is a vector of 4 096 dimensions.

We use it as a feature vector for the image matching. We measure the image similarity distances by calculating Euclidean distances (EDs) between CNN image features extracted from q_i and c_i. Each image from q_i is only compared with the corresponding image from c_i aligned with the same primary axes. We use the image similarity distance between q_i and c_i:

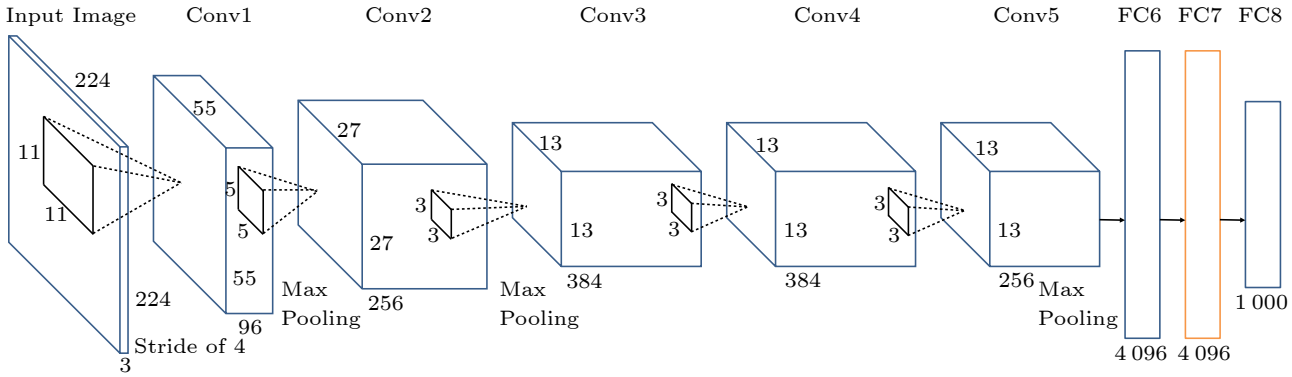

Fig.4. Illustration of AlexNet architecture[18]. The 1st convolutional (Conv1) layer filters the 224×224 input image with 96 kernels of size 11×11 with a stride of 4 pixels. The 2nd convolutional (Conv2) layer takes as input the (max pooled) output of Conv1 and filters it with 256 kernels of size 5×5. The 3rd, 4th, and 5th convolutional layers are connected to one another without any pooling. The 5th layer is max pooled to the fully-connected (FC) 6th layer. We use FC7 as a feature extractor for image matching.

$$\begin{aligned} &\textit{Image Similarity Distance}(q_i, c_i) \\ &= \sum_{j=1}^{2} \sum_{b=1}^{4\,096} \sqrt{(F^b_{q_{i_j}} - F^b_{c_{i_j}})^2}, \end{aligned}$$

where q_{i_j} is the j-th image from the input image pair, c_{i_j} is the j-th image from the image pair from the knowledge database, and F^b is the b-th CNN image feature.

DTW-Based Intensity Profile Matching. We measure the intensity profile similarity distance between q_p and c_p by calculating the cost of an optimal warping path using DTW. An $L_{q_p} \times L_{c_p}$ distance matrix is created which presets all pairwise distances between each element in q_p and c_p, where L_{q_p} is the length of q_p and L_{c_p} is to c_p. A warping path P is then defined to pass through the contiguous low-cost parts of distance matrix to create a sequence of points $= (P_1, P_2, \ldots, P_K)$ as shown in Fig.5.

The optimal warping path P^* is calculated as the path with the minimum cost:

$$\begin{aligned} &\textit{Profile Similarity Distance}(q_p, c_p) \\ &= \min \left(\sqrt{\sum_{k=1}^{K} p_k / K} \right), \end{aligned}$$

where K is the length of a sequence, and there are three constraints: 1) boundary condition: P starts and ends in the diagonally opposite corners of the distance matrix; 2) continuity: each step in P proceeds from, and moves to, an adjacent cell; and 3) monotonicity: P does not take a step backwards in spatial locations.

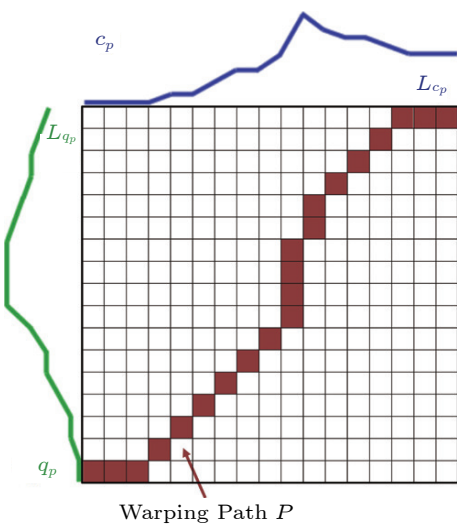

Fig.5. Warping path between an intensity profile pair example (q_p and c_p) in DTW.

3.4 SOI-Based TF Generation

We generate a conventional intensity-based 1D TF for SOIs of a TIQ using SOI labels retrieved from the knowledge database. We formulate a separate TF component T_m for each SOI S_m, $m = 1, \ldots, M$, where M is the number of SOI labels. Each component is a tent-shape in the TF parameter space (intensity-opacity), and is specified as follows:

$$T_j = \{(I_{\mathrm{l},m}, 0), (\overline{I}_m, \sigma_m), (I_{\mathrm{h},m}, 0)\},$$

where $I_{\mathrm{l},m}$ is the lowest intensity value in S_m, $I_{\mathrm{h},m}$ is the highest intensity value in S_m, $\overline{I}_m$ is the average intensity value in S_m, and σ_m is the initial opacity of a tent peak set to 0.3 (where σ_m=1.0 indicates full opacity). The initial TF is given by the union of the tent shapes:

$$\cup_{m=1}^{M} T_m.$$

We use the tent shapes since they reveal the isosurface of their corresponding SOIs, allowing a TF to visualize multiple SOIs in a semi-transparent manner.

This is a desirable property in the visualization of complex SOIs. Nevertheless, our TF generation is not limited to this particular shape and can be re-parameterized to other shapes depending on specific visualization and data needs[21]. We finally assign each tent apex to unique colors using ColorBrewer[22] and all tent end points to a single color (black) to improve visual color depth perception among SOIs[6]. However, there is no limitation to apply different color schemes, e.g., the tent end points and apex can have the same color.

3.5 Visibility-Based TF Parameter Optimization

Visibility-based TF parameter optimization[9, 10] is an iterative process: the initial TF and target visibilities of SOIs are entered as the inputs. Each iteration calculate an intermediate new TF and the resulting visibility of SOIs. The visibility calculation is explained later in this subsection. The iteration continues until it converges to a minimum of an energy function E of visibility tolerance:

$$min_\theta\, E(\theta) = \sum_{m=1}^{M} (V_{\mathrm{T}_m} - V_{\mathrm{R}_m})^2,$$

where M is the total number of SOIs, and θ is a set of opacity parameters of tent apexes forming the initial TF. The visibility tolerance is used to control how much the SOIs are visible in the optimized visualizations, and it can be calculated via the sum of the squared differences between target visibility V_{T} and resultant visibility V_{R} for individual SOIs. The value of V is the visibility proportion among all SOIs (where $V = 1.0$ means the corresponding SOI exclusively priorities visibility); the assignment of V_{T} is either automatically uniformly set to all SOIs or user-defined to enhance the visibility of particular SOIs.

We use a downhill simplex method[23] to solve the optimization problem. This method is effective in a variety of practical non-linear optimization problems with multiple local minima[24]. However, there is no limitation to apply other gradient-based approaches, e.g., gradient-descent.

The visibility $F(p)$ for a voxel is the front-to-back opacity composition of all the voxels of a volume starting from a view-point h to p according to [10]:

$$F(p) = e^{-\int_p^h A(g)\mathrm{d}g},$$

$$A(p) = A(p - \Delta p) + (1.0 - A(p - \Delta p))O(p),$$

where $A(p)$ is the composited opacity of the voxel coordinate p, $O(p)$ is its opacity, which is defined by a TF, and Δp is the size of the sampling step. The visibilities of all voxels, weighted as the product of their opacity, are then added to determine the visibility of SOI (intensity range) given by:

$$Visibility of SOI = \sum_{p \in SOI} O(p)F(p).$$

3.6 Evaluation Procedure

The 3D-IRCADb-01 dataset[20] is used for the evaluation. It consists of 20 CT volumes of the upper-abdomen (10 males and 10 females). Since the dataset is mainly designed for benchmarking liver segmentation algorithms, there are limited manual labels available. We select six SOIs that commonly appear at the abdomen, which consist of the artery (1 058 occurrences among all the images slices), bone (1 824), kidney (506), liver (1 645), lung (1 101), and spleen (306).

We believe that our chosen SOIs have sufficient complexity to measure the retrieval performances since the multiple SOIs, e.g., kidney, liver, and spleen, are similar in image features such as intensity and shape, and accurate retrieval among the SOIs is challenging. In total, our knowledge database is made of 3 840 ray elements derived from the 20 volumes, with each having 192 rays.

We compare our CBR-TF approach with the approach published in the visualization community by Kohlmann *et al.*[17]. This baseline approach relies on intensity profile matching using ED for measuring the similarity between the input and the cases from the knowledge database. For this comparison, only the second stage of our CBR-TF approach, which is the intensity profile matching using DTW, is applied in the knowledge database retrieval. We evaluate the first stage of our CBR-TF approach, which is image matching using pre-trained AlexNet[18], by comparing with two established pre-trained CNNs of GoogleNet[25] and ResNet[26] by measuring the retrieval accuracy. The pre-trained AlexNet is also compared with the fine-tuned counterpart using a fine-tuning method[27]. We note that there are no comparable work that uses image retrieval in TF design as well as DVR visualization.

The most common retrieval evaluation metrics of recall and precision are used. Recall refers to the number of times the SOI is correctly identified out of the number of times the SOI occurs. Precision refers

to the number of times the SOI is correctly identified out of the number of times the SOI is identified by a retrieval approach. The values of both recall and precision range from 0.0 to 1.0, where a value of 1.0 indicates that all instances of an SOI type are identified by a retrieval approach. In our evaluation, recall is considered the most representative in evaluating retrieval performances; if false-positive occurs, the user can manually eliminate it in the subsequent TF design, but such manual optimization can be much more challenging in missing SOIs (i.e., false-negative). We, therefore, discuss the recall results and include the precision results in the online resource (see Table A1 in the supplementary file[①]). We conduct 10-fold cross validation for the evaluation, where for each fold, 18 CT volumes are acted as the knowledge database, and the other two CT volumes are used as the queries. We rotate this process nine times to cover all 20 CT volumes.

For all these experiments, we use a consumer PC with an Intel i7 CPU and an Nvidia GTX 980 Ti GPU, running Windows 7 (x64), and implement our CBR-TF approach using volume rendering engine (Voreen)[28].

4 Results

4.1 CBR Analysis

We show the recall metrics for retrieval among six different SOIs between the ED baseline[17] and our two-stage CBR method in Table 1. We also present results from using individual compartments of our method DTW and AlexNet to analyze their individual performances and their contributions to the combined results. Compared with the ED baseline method[17], our two-stage method outperform in every SOI retrieval by 0.201 on average and kidney retrieval by 0.304 as the largest improvement.

Our two-stage CBR method outperform both the individual compartments in all the SOIs. Within the two compartments, the AlexNet method is better than the DTW method for artery, bone, and kidney. These SOIs tend to globally exhibit in an image, and regional spatial semantics available from the AlexNet method plays an important role in identifying them. In contrast, the DTW can provide the primitive features that better represent the local SOIs of the liver, lung, and spleen. The DTW method improves upon the ED baseline counterpart across all the SOIs.

Our image retrieval compartment method is not limited to the particular CNN backbone of AlexNet[18] and can be applied to other CNNs. In Table 2, we show the retrieval accuracies of pre-trained CNNs including AlexNet[18] and two other CNNs with deeper layers, GoogleNet[25] and ResNet[26]. Our results show that the pre-trained AlexNet, our default, has the highest average recall across all the SOIs except for lung where the pre-trained GoogleNet results in marginal improvement with 0.007. Regardless of the type of CNNs, AlexNet, GoogleNet, and ResNet outperform the ED baseline counterpart[17] across every SOI. The comparison result in Table 3 shows that the fine-tuned AlexNet is not always able to achieve higher recall accuracy compared with the pre-trained AlexNet; only spleen produces some meaningful improvement from the fine-tuned AlexNet by 0.044. This means that the fine-tuning method used[27] might

Table 1. Recall for Retrieval Among Six Different SOIs for the ED Baseline[17] and Our Two-Stage CBR Method with Two Individual Compartments

Method	Artery	Bone	Kidney	Liver	Lung	Spleen	All
ED baseline	0.453	0.491	0.344	0.649	0.552	0.363	0.518
Our two-stage CBR method	**0.612**	**0.692**	**0.648**	**0.778**	**0.829**	**0.644**	**0.719**
DTW (individual compartment)	0.530	0.645	0.542	0.760	0.808	0.569	0.672
AlexNet (individual compartment)	0.549	0.681	0.615	0.745	0.795	0.454	0.679

Note: The bold numbers are the highest.

Table 2. Recall for Retrieval Among Six Different SOIs for Our Two-Stage CBR Method with Three Different Types of Pre-Trained CNNs[18, 25, 26] in Comparison with the ED Baseline[17]

Method	Artery	Bone	Kidney	Liver	Lung	Spleen	All
With pre-trained AlexNet	**0.612**	**0.692**	**0.648**	**0.778**	0.829	**0.644**	**0.719**
With pre-trained GoogleNet	0.591	0.681	0.616	0.770	**0.836**	0.608	0.707
With pre-trained ResNet	0.543	0.624	0.547	0.682	0.668	0.487	0.620
ED baseline	0.453	0.491	0.344	0.649	0.552	0.363	0.518

Note: The bold numbers are the highest.

[①] https://github.com/JCST-supplementary/Paper-Supplementary/blob/main/CGI-2023-ID44-OnlineResource.pdf, Mar. 2024.

Table 3. Recall for Retrieval Among Six Different SOIs for the Pre-Trained AlexNet[18] and Its Fine-Tuning Counterpart[27]

Method	Artery	Bone	Kidney	Liver	Lung	Spleen	All
With pre-trained AlexNet	**0.612**	0.692	**0.648**	**0.778**	**0.829**	0.644	**0.719**
With fine-tuned AlexNet	0.609	**0.700**	0.617	0.765	0.816	**0.688**	0.714

Note: The bold numbers are the highest.

not be optimal for our application.

4.2 Qualitative Analysis of Our Two-Stage CBR Method

Fig.6 exemplifies the capability of our two-stage method for SOI retrieval with comparison with the ED baseline method using only the intensity profile[17]. The results show that the ED baseline method incorrectly labels the arteries and kidney along the TIQ. In contrast, all the SOIs along the TIQ are correctly retrieves when our two-stage CBR method is applied.

In Fig.7, we show the SOIs retrieval results from our two-stage method in a comparison with the two individual compartments. The results demonstrate that none of the compartment methods are sufficient for the adequate retrieval of SOIs present within the TIQ. In contrast, the two-stage method produces the same SOI labels as in the TIQ.

In the online resource (see Fig.A1 in the supplementary file[②]), we include an intensity profile re-

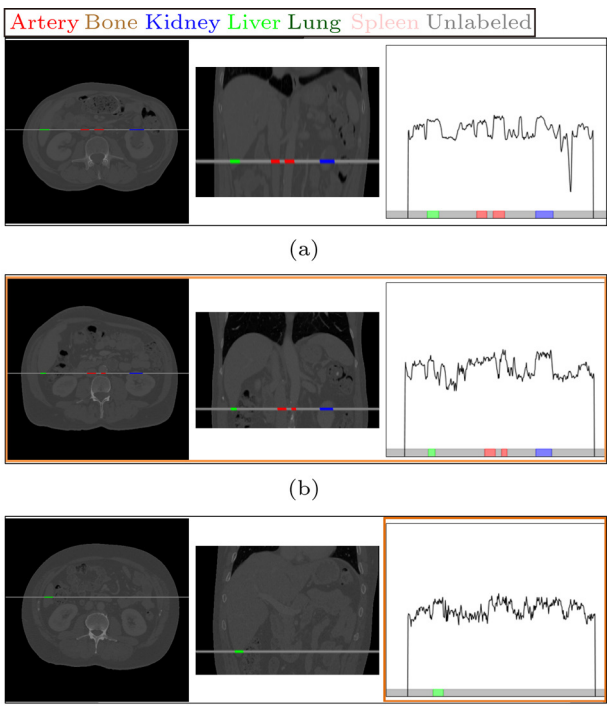

Fig.6. Comparison of our two-stage method with the ED baseline method[17] using (a) a TIQ that encounters three SOIs: the liver (bright green color), arteries (red color), and the kidney (blue color). The right column shows the intensity profiles with the SOI labels, and the other two left columns show the associated image pair, respectively. The bounding boxes with orange lines indicate the compartments used for the retrieval computation. (b) Retrieval from our two-stage method (an image pair + an intensity profile). (c) Retrieval from ED (an intensity profile).

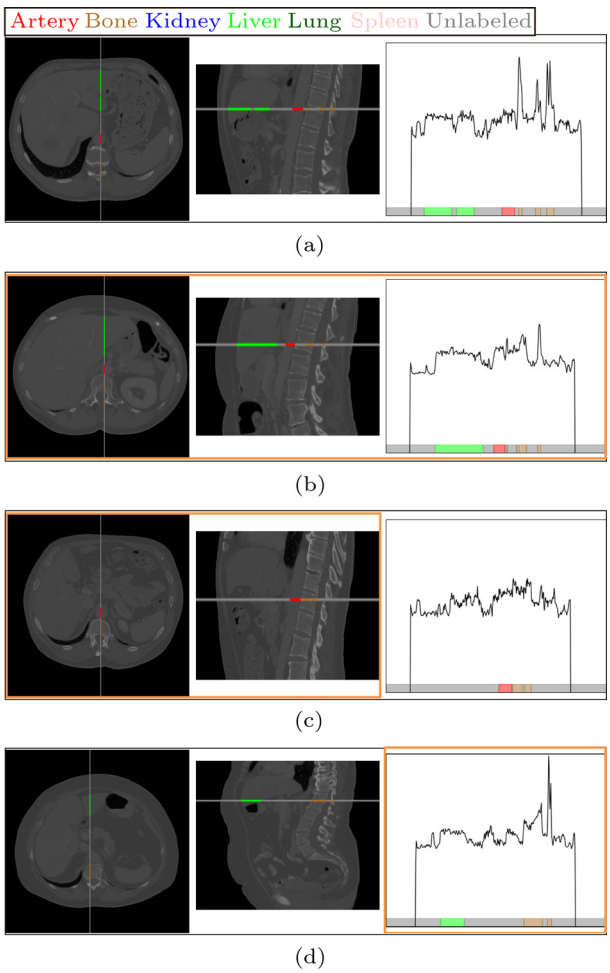

Fig.7. Comparison of our two-stage method with the two individual compartments: the image based AlexNet method and the intensity profile based DTW method, with (a) a TIQ that passes through three SOIs: the liver (bright green color), the artery (red color), and bones (brown color). The bounding boxes with orange lines indicate the compartments used for the retrieval computation. (b) Retrieval from our two-stage method (an image pair + an intensity profile). (c) Retrieval from AlexNet compartment (an image pair). (d) Retrieval from DTW compartment (an intensity profile).

[②]https://github.com/JCST-supplementary/Paper-Supplementary/blob/main/CGI-2023-ID44-OnlineResource.pdf, Mar. 2024.

trieval comparison between our DTW compartment and the ED baseline method[17] using another volume, where our DTW compartment method correctly identifies all three SOIs while only one SOI is achieved in the ED baseline method.

We illustrate the characteristics of our two-stage method, which searches across the entire image knowledge database, and then refines the retrieved results using the intensity profile matching. In Fig.8, we show the four most similar image retrieval results.

We can see that these images are similar to the image query pair, with greater similarity (i.e., consistency in the same SOIs) in the coronal views (the 2nd column). With the inclusion of the intensity profile matching results to the top 4 ranked image retrieval results, we see that the best intensity profile matching is with the 4th best image retrieval, among the only two results that are able to match the spleen (with the other result being the 3rd best image retrieval). Thus, in the combined result, the 4th image retrieval result will become the 1st ranked final retrieval result and has the overall best matching.

4.3 Automated TF Design Using Our CBR-TF Approach

In Fig.9, a use-case of our CBR-TF approach is presented to effectively explore a medical volume and generate an SOI-based DVR visualization. The user is only required to draw a line (a ray) onto any of the 2D cross-sectional views for our CBR-TF approach to identify and visualize SOIs. In this example, a line is drawn to contain the kidneys and bones in the axial view. An image pair and intensity profile associated with the ray are then used as the TIQ to retrieve annotations (labels) from the knowledge database to generate 3D DVR for the SOIs. The resultant DVR visualization provides the user with 3D information about the SOIs, including volumetric shapes and spatial relationships among the SOIs, which complements the 2D cross-sectional views.

In Fig.10, we present another use-case of our CBR-TF approach. Instead of drawing a line in 2D cross-sectional views, the user can select a single point directly on an initial DVR visualization. From the selected point, a viewing ray perpendicular to a camera origin is constructed and used to generate a TIQ of an image pair and an intensity profile. This example illustrates how different SOIs can be identified based on the simple point selection, and how the TIQs can be used to automatically design the TF parameters such as to emphasize the local SOIs: the liver, the kidney, and the bone.

We compare the TF generation results of our CBR-TF approach and the ED baseline counterpart[17], in Fig.11. The results show that our CBR-TF approach produces the enhanced DVR outcomes, i.e., the correct identification of the SOIs, in comparison with the ED baseline counterpart. With the ED baseline, the bone fails to be identified, thus resulting in the missing TF component.

4.4 Computational Performance

Computational time among the ED baseline[17] and our two-stage CBR method for retrieval from the knowledge database is presented in Table 4. Our DTW compartment is marginally slower (30 ms) than the ED baseline counterpart. As expected, the greater computation is required for the retrieval of images (AlexNet) compared with intensity profiles (DTW). In addition, the TF generation computation time to obtain final DVR visualizations is 1 s in all the figures in this work.

5 Discussion and Future Work

The results demonstrate that with few user interactions, our CBR-TF is able to identify SOIs and generate TFs for an input volume. We also show simple and intuitive usability of our CBR-TF.

5.1 CBR

We attribute our superior retrieval performance to our two-stage method, which utilizes a rich volumetric image feature derived from our TIQ: 1) DTW characterized local primary intensity profiles from the ray of a TIQ, and 2) regional high-level spatial semantics extracted by a CNN from the two co-planar images along the ray, in a complementary manner. The results demonstrate that none of the individual compartments can outperform our combined use (see Fig.7). Our two-stage method is particularly effective in the cases where individual compartments produce poor performance, such as artery, kidney, and spleen SOIs (see Table 1). This is consistent with the findings from existing work such as [29], which demonstrates the advantages in retrieval performance when leveraging two different yet complementary feature sets.

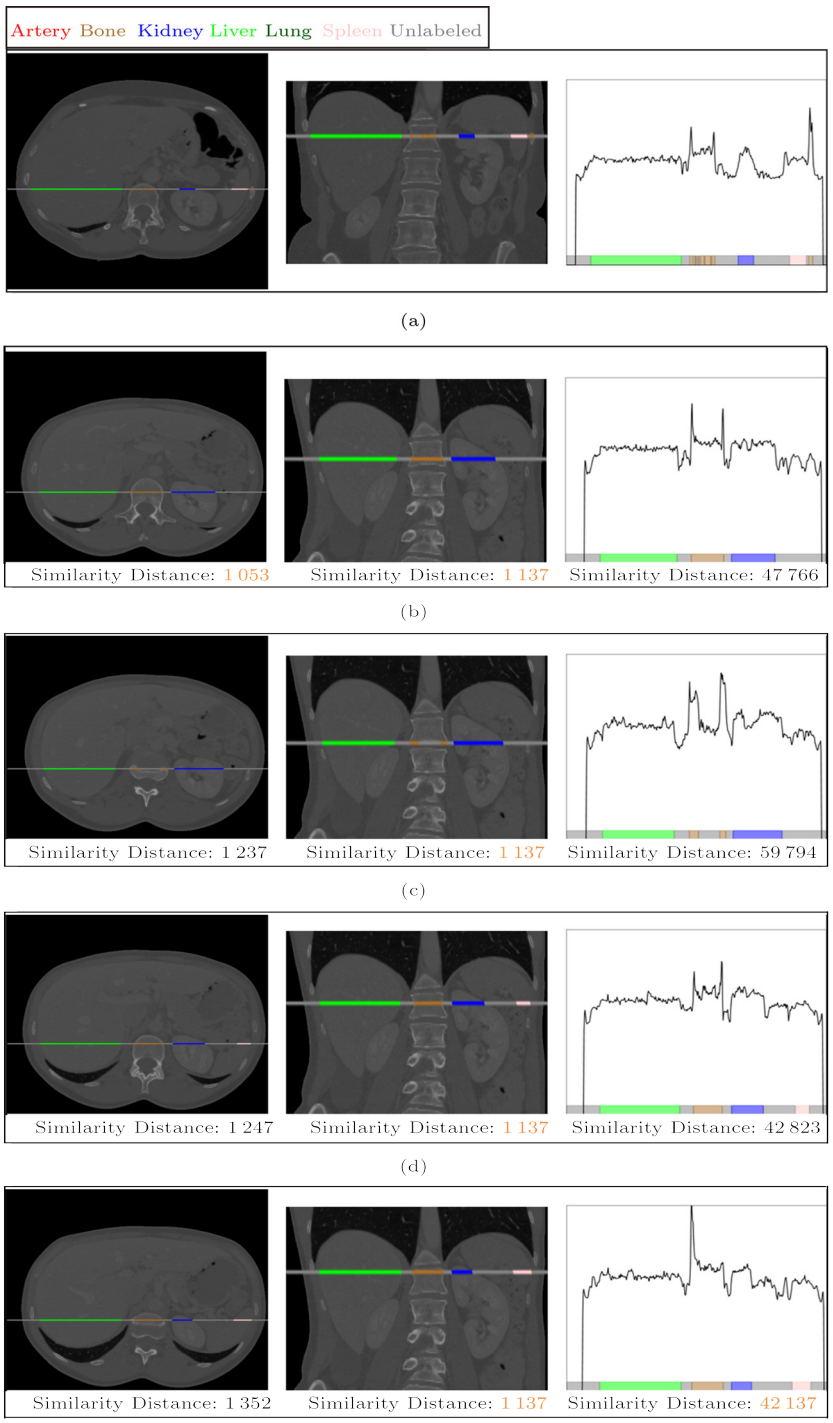

Fig.8. (a) Complex TIQ that passes through four SOIs: the liver (bright green color), bones (brown color), the kidney (blue color), and the spleen (peach color). (b)–(e) The four most similar retrieval results from our two-stage method. The smallest similarity distance (the best matching) is colored in orange for each compartment (each column). (e) The overall best result from our two-stage method, denoted by "*".

Our two-stage method compliments the intensity profile retrieval[17], with co-planar image retrieval. One of the benefits of introducing image retrieval is that it can provide regional image semantics, e.g., not only shapes or textures of SOIs as well as the spatial relationship among the SOIs. This contributes towards differentiating the SOIs that have similar intensity values, and this differentiation is not able to be with the intensity profile retrieval only (see Fig.6 with the SOIs of the liver, artery, and kidney).

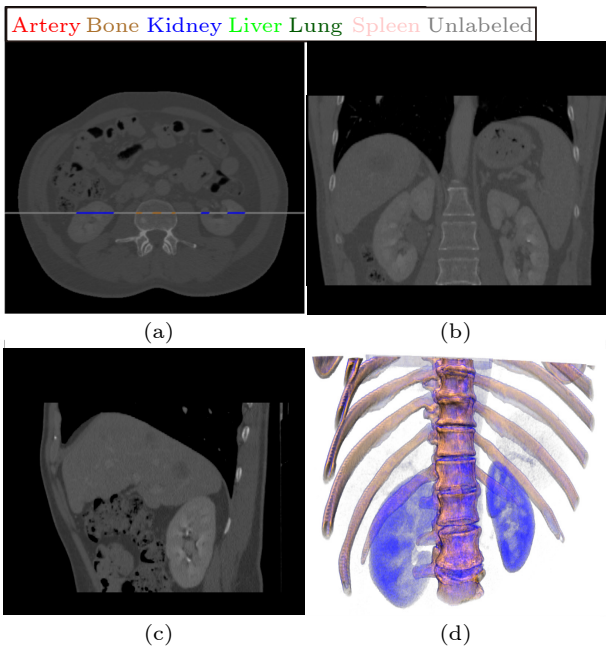

Fig.9. Example of a use-case for our CBR-TF approach, where the typical quarter-view visualization of a medical volume is used: 2D cross-sectional (a) axial, (b) coronal, and (c) sagittal view, together with (d) a 3D DVR. A user specifies SOIs such as the kidneys (blue color) and the spine (brown color) by drawing a line containing them onto (a) the axial view (a ray with gray color) and our CBR-TF approach updates the DVR to emphasize them in 3D.

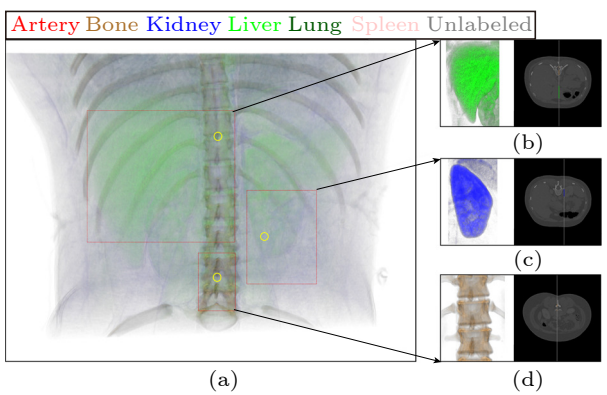

Fig.10. Example of a use-case for our CBR-TF approach for the visual enhancement of local SOIs directly from DVR visualization: (b) the liver (bright green color), (c) the kidney (blue color), and (d) the bone (brown color), within (a) an upper-abdomen DVR. Each of the SOIs is selected by a single mouse click (yellow circles) on (a) the initial DVR where we construct a viewing ray perpendicular to the camera origin which contains each SOI.

Our results from DTW-based intensity profile retrieval outperform the ED baseline counterpart[17]. The ED baseline only considers voxel pairs that are exactly in the same position and cannot match subtle differences. Our DTW-based method addresses timing differences by warping the primary axis of one, so

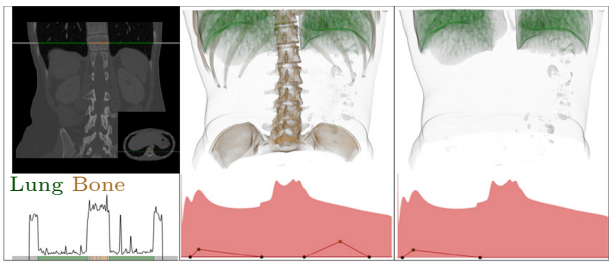

Fig.11. Comparison of TF and DVR visualization results from (b) our CBR-TF approach and (c) the ED baseline approach[17] using (a) TIQ that consists of two SOIs of the lungs (dark green color) and bones (brown color).

Table 4. Average Time to Calculate the Retrieval Results from the Knowledge Database Among the Four Different CBR Methods

CBR Method	Time (s)
ED baseline	0.84
Our two-stage method	15.93
DTW (individual compartment)	1.14
AlexNet (individual compartment)	15.70

that the maximum coincidence is attached with the other. This characteristic is important for our medical imaging application where the extents of the same SOIs (the intensity profiles) are widely varied among the patient volumes. In such a scenario, a retrieval method must be able to accommodate these variations. A representative example is shown in the online resource (see Fig.A1 in the supplementary file[③]).

For image retrieval, we choose the high-level CNN features learned from the large-scale knowledge database rather than relying on the low-level features of an input volume manually engineered by a user. Our choice is attributed to the findings that although requiring the increased cost of computational complexity and the long period of a training process, the high-level CNN features better represent SOIs and more accurately perform the CBR retrieval task[30]. It may be useful to compare our application (i.e., the quality of ROI visualizations) using the high-level CNN features with that using the low-level features. The comparison would be an interesting addition to our current work. We use a pre-trained AlexNet[18] due to its highest average recall of 0.719 when compared with the two pre-trained deeper CNNs, GoogleNet[25] and ResNet[26] (see Table 2). This can be because the deeper networks learn more data-specific features that are relevant to general images, i.e., less generalizable to medical images.

[③]https://github.com/JCST-supplementary/Paper-Supplementary/blob/main/CGI-2023-ID44-OnlineResource.pdf, Mar. 2024.

Our two-stage method, i.e., the rank-based sequential combination of the two individual compartments, is experimentally designed based on the observation that the two individual compartments can be complimentary, but this advantage is not properly encoded when two individual compartments are used at the same time (i.e., a single-stage retrieval). In the first stage of image retrieval, we use the top 40 retrieval results (i.e., $N = 40$), with the second stage of intensity profile retrieval only executed within the N retrievals. This is empirically derived with variations in N having a minor impact on the overall retrieval performances. Full details of this experiment are included in the online resource (see Table A2 in the supplementary file④).

In this work, we use a well-established recall metric to evaluate the retrieval performances. As a complementary metric, we also include precision results in the online resource (see Table A1 in the supplementary file④); we note that, as with recall, our two-stage method perform better in precision compared with the ED baseline[17].

5.2 TF Design

Our results show that our image-centric interaction for TF design is intuitive and effective. Our CBR-TF approach requires a user to simply draw a line in 2D cross-sectional views (see Fig.9) or a point in 3D DVRs (see Fig.10) to select SOIs. The usability of such image-centric interaction to directly select SOIs in a visualization space, not in an indirect TF widget space, is investigated by the early work[7], where positive informal feedback from clinicians who conducted interaction using a medical CT volume was observed.

The use-case with the quarter-view visualization (see Fig.9) can be effective in medical image visualization where the users, i.e., clinicians, generally have prerequisite knowledge of their data; they can readily define a line (a ray) with SOIs on the 2D cross-sectional views. The fact that they are novice users for TF design makes our CBR-TF approach more valuable. They do not need to rely on the TF widgets which are complicated to and unfamiliar with novice users. Instead, the SOIs can be simply defined in the 2D cross-sectional views where they are usually working on in the clinical routine.

With another use-case (see Fig.10) where the single point selection is directly done in an initial 3D DVR for the SOI identification, we suggest that our CBR-TF approach enables a user to obtain a more detailed visualization of specific SOIs intuitively and automatically, thereby easily inspecting information that might be hidden in the initial DVR.

Our CBR-TF approach focuses on generating opacity parameters of TFs, and color parameters are not much investigated. We note that static opacity settings, even one that has been carefully pre-defined by domain experts, cannot be universally applied to all cases, and manual opacity manipulation must be involved to some extent for new volumes and visualization scenarios. In contrast, color settings can be relatively consistent in multiple cases, e.g., the same blueish color can be applied to lung tissues in different volumes. This observation is the basis to develop our TF usage scenarios. Our CBR-TF approach enables the user to generate an opacity TF that guarantees the visibility of SOIs, and then applies to pre-defined color settings to add similar visual perceptions to the same SOIs. As such, we consider all the opacity and color parameters in the TF generation.

In our CBR-TF approach, we adopt an automated visibility-based TF parameter optimization to ensure that some of the identified SOIs prioritize visibility in resulting DVRs. The user, therefore, can avoid additional manual TF fine tuning that may be required for the cases where the TFs from knowledge databases are not optimal for the particular SOIs of the input volume. A TF parameter optimization application example is shown in Fig.11, where the visibility of the bones is prioritized over the lungs by setting V_{T} of the bones to 0.7 and V_{T} of the lungs to the rest (0.3).

In our CBR-TF approach, TFs are comprised of individual tent-shape components to represent each SOI. In the cases when the false-positive from our SOI retrieval occurs, we note that the user needs to manually eliminate the TF component of the false-positive SOI. Such manual TF optimization can be more challenging in missing structures (i.e., false-negative). However, our CBR-TF approach is still effective because the user does not need to start TF design from scratch, and it is easier to add new SOIs to the TF that already contains other SOIs.

In terms of the generation of knowledge databases, existing approaches[14, 17] collect pre-designed TFs

④https://github.com/JCST-supplementary/Paper-Supplementary/blob/main/CGI-2023-ID44-OnlineResource.pdf, Mar. 2024.

by domain experts, whereas our CBR-TF approach relies on SOI labels of raw volumes. In terms of the perspective of experts in knowledge database generation, generating TFs may take much effort compared with generating SOI labels.

5.3 Future Work

The image retrieval results (Table 2) demonstrate the adaptability of our CBR-TF approach to different CNNs. As such, our CBR-TF approach is not constrained to a particular CNN backbone. We note that recent advances in CNNs, e.g., large visual models for segmentation or classification tasks[31] and self-supervised learning with limited or unlabeled data[32], demonstrate the capabilities to better describe image features to represent SOIs[33]. These backbones can be leveraged for our CBR-TF approach, but optimal adoption should be investigated to improve SOI retrieval performance (see Table 3). Similarly, our ranking-based sequential combination for the SOI retrieval still has room for improvement. We experimentally determine the adoption of our ranking method because it outperforms other established ad-hoc counterparts, e.g., ranking based on the sum of the normalized distances from CNN and DTW compartments. We suggest that it can be further explored to develop a learning-based ranking, e.g., by using ensemble learning[34]. We consider these challenging while interesting investigations as our future work.

Our CBR-TF approach does not require any dataset-specific parameter settings for the knowledge database construction. Furthermore, we automate the entire construction procedure including ray extraction, representation, and storage. Different types of imaging modalities can be easily integrated for different application domains. We limit the scope of this work to medical imaging data, but we suggest that our CBR-TF approach can be applied to non-medical imaging datasets.

We demonstrate our CBR-TF approach with intensity-based 1D TF design, because it is commonly used in a wide range of visualization applications and has fewer variables that influence visualization experiments. Our CBR-TF approach, however, is not restricted to 1D TFs and can be adapted to multi-dimensional TFs[3, 4] for greater differentiation of SOIs. Huang *et al.*[35] proposed an approach to generate 2D TFs when the identification of SOIs in the 2D cross-sectional views are available. We regard this as an important new avenue of research.

6 Conclusions

We present CBR-TF, a new transfer function design approach by proposing an automated two-stage content-based retrieval method for a knowledge database, and the use of its retrieved results in generating transfer functions. Our experimental results demonstrate that our CBR-TF is capable of identifying and visualizing structures of interests (SOIs) from the combination of the complementary image semantics with the intensity profile matching.

Conflict of Interest The authors declare that they have no conflict of interest.

References

[1] Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller J V, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 2012, 30(9): 1323–1341. DOI: 10.1016/j.mri.2012.05.001.

[2] Ljung P, Krüger J, Groller E, Hadwiger M, Hansen C D, Ynnerman A. State of the art in transfer functions for direct volume rendering. *Computer Graphics Forum*, 2016, 35(3): 669–691. DOI: 10.1111/cgf.12934.

[3] Kniss J, Kindlmann G, Hansen C. Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. In *Proc. the 2001 IEEE Visualization*, Oct. 2001, pp.255–562. DOI: 10.1109/VISUAL.2001.964519.

[4] Correa C, Ma K L. Size-based transfer functions: A new volume exploration technique. *IEEE Trans. Visualization and Computer Graphics*, 2008, 14(6): 1380–1387. DOI: 10.1109/TVCG.2008.162.

[5] Caban J J, Rheingans P. Texture-based transfer functions for direct volume rendering. *IEEE Trans. Visualization and Computer Graphics*, 2008, 14(6): 1364–1371. DOI: 10.1109/TVCG.2008.169.

[6] Jung Y, Kim J, Kumar A, Feng D D, Fulham M. Feature of interest-based direct volume rendering using contextual saliency-driven ray profile analysis. *Computer Graphics Forum*, 2018, 37(6): 5–19. DOI: 10.1111/cgf.13308.

[7] Ropinski T, Praßni J, Steinicke F, Hinrichs K. Stroke-based transfer function design. In *Proc. the 5th Eurographics/IEEE VGTC Conference on Point-Based Graphics*, Aug. 2008, pp.41–48.

[8] Guo H Q, Mao N Y, Yuan X R. WYSIWYG (what you

see is what you get) volume visualization. *IEEE Trans. Visualization and Computer Graphics*, 2011, 17(12): 2106–2114. DOI: 10.1109/TVCG.2011.261.

[9] Correa C D, Ma K L. Visibility histograms and visibility-driven transfer functions. *IEEE Trans. Visualization and Computer Graphics*, 2011, 17(2): 192–204. DOI: 10.1109/TVCG.2010.35.

[10] Jung Y, Kim J, Eberl S, Fulham M, Feng D D. Visibility-driven PET-CT visualisation with region of interest (ROI) segmentation. *The Visual Computer*, 2013, 29(6): 805–815. DOI: 10.1007/s00371-013-0833-1.

[11] Jung Y, Kim J, Kumar A, Feng D D, Fulham M. Efficient visibility-driven medical image visualisation via adaptive binned visibility histogram. *Computerized Medical Imaging and Graphics*, 2016, 51: 40–49. DOI: 10.1016/j.compmedimag.2016.04.003.

[12] Jung Y, Kim J, Bi L, Kumar A, Feng D D, Fulham M. A direct volume rendering visualization approach for serial PET–CT scans that preserves anatomical consistency. *International Journal of Computer Assisted Radiology and Surgery*, 2019, 14(5): 733–744. DOI: 10.1007/s11548-019-01916-2.

[13] Marks J, Andalman B, Beardsley P A, Freeman W, Gibson S, Hodgins J, Kang T, Mirtich B, Pfister H, Ruml W, Ryall K, Seims J, Shieber S. Design galleries: A general approach to setting parameters for computer graphics and animation. In *Proc. the 24th Annual Conference on Computer Graphics and Interactive Techniques*, Aug. 1997, pp.389–400. DOI: 10.1145/258734.258887.

[14] Guo H Q, Li W, Yuan X R. Transfer function map. In *Proc. the 2014 IEEE Pacific Visualization Symposium*, Mar. 2014, pp.262–266. DOI: 10.1109/PacificVis.2014.24.

[15] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436–444. DOI: 10.1038/nature14539.

[16] Kumar A, Kim J, Cai W D, Fulham M, Feng D G. Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data. *Journal of Digital Imaging*, 2013, 26(6): 1025–1039. DOI: 10.1007/s10278-013-9619-2.

[17] Kohlmann P, Bruckner S, Kanitsar A, Groller M E. Contextual picking of volumetric structures. In *Proc. the 2009 IEEE Pacific Visualization Symposium*, Apr. 2009, pp.185–192. DOI: 10.1109/PACIFICVIS.2009.4906855.

[18] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90. DOI: 10.1145/3065386.

[19] Keogh E, Ratanamahatana C A. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 2005, 7(3): 358–386. DOI: 10.1007/s10115-004-0154-9.

[20] Soler L, Hostettler A, Agnus V, Charnoz A, Fasquel J B, Moreau J, Osswald A B, Bouhadjar M, Marescaux J. 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. Technical Report, IRCAD, 2010. https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/, Mar. 2024.

[21] Castro S, Königy A, Löffelmanny H, Gröllery E. Transfer function specification for the visualization of medical data. Technical Report, Vienna University of Technology, 1998. https://citeseerx.ist.psu.edu/doc_view/pid/08d0bdf2ffe297661e78568baa8f612c91d8e1c1, Mar. 2024.

[22] Harrower M, Brewer C A. ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 2003, 40(1): 27–37.

[23] Nelder J A, Mead R. A simplex method for function minimization. *The Computer Journal*, 1965, 7(4): 308–313. DOI: 10.1093/comjnl/7.4.308.

[24] Lagarias J C, Reeds J A, Wright M H, Wright P E. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 1998, 9(1): 112–147. DOI: 10.1137/S1052623496303470.

[25] Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In *Proc. the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015, pp.1–9. DOI: 10.1109/CVPR.2015.7298594.

[26] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In *Proc. the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp.770–778. DOI: 10.1109/CVPR.2016.90.

[27] Shin H C, Roth H R, Gao M C, Lu L, Xu Z Y, Nogues I, Yao J H, Mollura D, Summers R M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Medical Imaging*, 2016, 35(5): 1285–1298. DOI: 10.1109/TMI.2016.2528162.

[28] Meyer-Spradow J, Ropinski T, Mensmann J, Hinrichs K. Voreen: A rapid-prototyping environment for ray-casting-based volume visualizations. *IEEE Computer Graphics and Applications*, 2009, 29(6): 6–13. DOI: 10.1109/MCG.2009.130.

[29] Ahmed K T, Ummesafi S, Iqbal A. Content based image retrieval using image features information fusion. *Information Fusion*, 2019, 51: 76–99. DOI: 10.1016/j.inffus.2018.11.004.

[30] Vishraj R, Gupta S, Singh S. A comprehensive review of content-based image retrieval systems using deep learning and hand-crafted features in medical imaging: Research challenges and future directions. *Computers and Electrical Engineering*, 2022, 104: 108450. DOI: 10.1016/j.compeleceng.2022.108450.

[31] Wasserthal J, Breit H C, Meyer M T, Pradella M, Hinck D, Sauter A W, Heye T, Boll D T, Cyriac J, Yang S, Bach M, Segeroth M. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 2023, 5(5): e230024. DOI: 10.1148/ryai.230024.

[32] Zhang C Y, Zheng H, Gu Y. Dive into the details of self-supervised learning for medical image analysis. *Medical Image Analysis*, 2023, 89: 102879. DOI: 10.1016/j.media.

2023.102879.

[33] Chen X X, Wang X M, Zhang K, Fung K M, Thai T C, Moore K, Mannel R S, Liu H, Zheng B, Qiu Y C. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 2022, 79: 102444. DOI: 10.1016/j.media.2022.102444.

[34] Wang L, Qian X M, Zhang Y T, Shen J L, Cao X C. Enhancing sketch-based image retrieval by CNN semantic re-ranking. *IEEE Trans. Cybernetics*, 2020, 50(7): 3330–3342. DOI: 10.1109/TCYB.2019.2894498.

[35] Huang R Z, Ma K L. RGVis: Region growing based techniques for volume visualization. In *Proc. the 11th Pacific Conference on Computer Graphics and Applications*, Oct. 2003, pp.355–363. DOI: 10.1109/PCCGA.2003.1238277.

Younhyun Jung received his B.S. degree in computer science from Inha University, Incheon, in 2008, and his Ph.D. degree in computer science from The University of Sydney, Sydney, in 2016. He is currently an assistant professor in computer science with School of Computing, Gachon University, Seongnam. He was a software engineer with Samsung Electronics from 2007 to 2010. His current research interests include volume rendering and multimodal medical image visualization.

Jim Kong received his B.S. (Hons.) degree in computer science from The University of Sydney, Sydney, in 2016. He is currently a software development engineer in Amazon, Sydney.

Bin Sheng received his B.A. degree in English and his B.Eng. degree in computer science from Huazhong University of Science and Technology, Wuhan, in 2004, and his M.Sc. degree in software engineering from the University of Macau, Macau, in 2007, and his Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2011. He is currently a professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai. He is an associate editor of the IEEE Transactions on Circuits and Systems for Video Technology, and The Visual Computer Journal. His current research interests include virtual reality and computer graphics.

Jinman Kim received his B.S. (Hons.) and his Ph.D. degrees in computer science both from the University of Sydney, Sydney, in 2001 and 2006, respectively. From 2008 to 2012, he was an ARC Post-Doctoral Research Fellow, one year leave from 2009 to 2010 to join the MIRALab Research Group, Geneva, as a Marie Curie senior research fellow. Since 2013, he has been with the School of Information Technologies, The University of Sydney, Sydney, where he was a senior lecturer, and became a professor in 2022. His current research interests include medical image analysis and visualization, computeraided diagnosis, and telehealth technologies. He is an associate editor of The Visual Computer Journal.