CHAPTER 4

# Extracting heterogeneous vessels in X-ray coronary angiography via machine learning

**Binjie Qin**[1], **Mingxin Jin**[1] **and Song Ding**[2]
[1]Biomedical Engineering School, Shanghai Jiao Tong University, Shanghai, China
[2]Department of Cardiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

## 4.1 Introduction

Cardiovascular disease remains the leading cause of death in the world. Percutaneous coronary intervention with the guidance of X-ray coronary angiography (XCA) has been routinely applied into the clinic since XCA is an imaging modality capable of observing internal structures and functions with superior spatiotemporal resolution. Apart from interventional guidance, XCA is assumed to be an important reference for better detection and treatment of impaired myocardial perfusion. Current XCA-guided coronary artery revascularization via stent implantation solely reconstructs the perfusion of main epicardial arteries. However, whether this artery revascularization could improve the distal microcirculation and accordingly lead to desired therapeutic effect on the acute myocardial ischemic heart diseases is not easily evaluated in preoperative and postoperative assessment. This is because the distal microcirculation on the most important parts of the coronary tree is poorly assessed by traditional XCA analysis. To solve this challenging problem, this chapter introduces machine-learning-based XCA analysis methods to extract and visualize the heterogeneous XCA vessels at pixel -level resolution.

Although XCA images are widely used in the clinic, the visibility of vessels (especially distal vessels) in XCA images is very poor due to complex spatiotemporal patterns of disturbances with the following reasons. XCA represents complex 2D/3D (2D + time) low-contrast structures of contrast-filled vessels by 2D projection of low dose X-ray imaging, which is badly overlapped by various background structures (e.g., catheters, bones, diaphragms, and lungs) and be simultaneously degraded by tissue-dependent noises and different motion patterns (i.e., respiratory and cardiac motions, as well as other tissue deformations during intervention). Furthermore, due to the bad effect of fastly evanesced contrast agents that are rapidly diffused inside the vessel networks, XCA vessels have fast change of inhomogeneous intensities that are disturbed with different vessel-like noisy structures and motion patterns.

*Cardiovascular and Coronary Artery Imaging*
DOI: https://doi.org/10.1016/B978-0-12-822706-0.00008-1

Recently, extracting contrast-filled vessels from XCA data has gradually attracted attention and achieved some developments along with some advancements in medical imaging and machine learning [1]. Specifically, robust subspace learning [2] via decomposition into low-rank plus sparse matrices [3] has proven to effectively separate moving objects from the background. Based on the fact that XCA sequence can be modeled as a sum of low-rank background structures and sparsely distributed foreground contrast-filled vessels, robust principal component analysis (RPCA) [4–8] is effectively exploited to extract the moving contrast agents of vessel networks from the noisy complex backgrounds. However, various RPCA-based methods [5–11] still cannot perfectly extract the contrast-filled vessels from the dynamic and complex backgrounds that have nonlocally coupled properties of low rankness, motion patterns, and vessel-like noisy artifacts.

With the popularity of data-driven deep learning in medical imaging including cardiac image segmentation [12], recent patch-based [13,14] or image-based [15,16] deep learning works have designed convolutional neural networks (CNN) [13,17] and encoder-decoder architectures [14–16,18] to segment entire vessel networks from the XCA sequences. By exploring various prior knowledges with image enhancement [13,16], attention mechanism [15], and transfer learning strategy [16,17], these deep learning-based methods have indeed improved the vessel segmentation performances. But there are still challenges or limitations in the noisy annotated datasets [19,20] and class imbalance problems [15].

By embracing the challenge in highlighting the minority class of foreground vessels from the majority class of complex backgrounds in XCA, we were inspired to enjoy the advantage of having taken the opposite tack by masking the initially segmented vessels as missing entries of backgrounds, such that our recently published work [21] completed the whole backgrounds of XCA via tensor completion of background layer and then subtracted the completed background layer from whole XCA sequence to well recover the heterogeneous vessels' shape and intensity. This chapter provides an overview of these related works in Section 4.2, and introduces our recent machine learning-based works on the RPCA-based vessel extraction in Section 4.3, encoder–decoder-based deep vessel segmentation networks with XCA-dedicated architecture design in Section 4.4, and heterogeneous vessel recovery with tensor completion of background layer in Section 4.5. The final discussion and conclusion are given in Section 4.6.

## 4.2 Related works

Most blood vessel extraction or vessel segmentation methodologies were reviewed in the papers [22,23] and recently updated in the works [24–27]. These review papers have mostly focused on various medical imaging modalities such as magnetic resonance angiography, computed-tomography angiography but rarely on XCA. Currently, there are few works dealing with vessel extraction for XCA, except for the

review [28] on reconstruction of high-contrast coronary arteries from X-ray angiography by focusing on the theoretical features in model-based tomographic reconstruction of coronary arteries. Generally, XCA vessel extraction works can be classified as three types of methodologies. First, earlier works extract blood vessels via digital subtraction angiography (DSA) that subtracts a precontrast mask image from later contrast images to clearly visualize the contrast-enhanced blood vessels and remove the interfering backgrounds. However, this simple technique has difficulty in handling vessels with a lot of motion artifacts since the vessels' surrounding backgrounds always have local tissue motions, noisy intensity variations, and interventional changes during DSA imaging. To reduce the motion artifact, image registration method is required to match the locally deformed mask images to the live images for minimizing their dissimilarity. Coupling with RPCA-based vessel enhancement method, the authors in the work [29] have generated a vascular roadmap that enables visualization of an entire blood vessel by combining multiple enhanced images with image registration. However, the registration-based vessel extraction methods have an unsolved challenging problem of nonrigid image registration with local large deformations and noncorresponding outlier features [30,31], therefore they cannot ensure the efficiency in reducing the sensitivity to the outlier features and motion artifacts of XCA imaging with the locally deformed backgrounds.

Second, most XCA vessel extraction methods generally belong to the machine learning-based algorithms by combining various model-based methods with various vessel filtering techniques including nonlinear and nonlocal filtering methods [24–27,32,33]. These integrated vessel extraction approaches usually involve two steps: a filtering step for enhancing and extracting vessel-like features and a mode-based classification step for highlighting the featured vessel pixels from the backgrounds. After the XCA images were denoised and enhanced, the model-based methods usually have highlighted the target vessels for further vessel refinements by formulating it as an energy optimization problem [1], which have included motion layer separation such as low-rank plus sparse decomposition modeling [4] (see Section 4.3), deformable models such as active contour [34], graph-based modeling [35], minimal path optimization [36–38], and tube-like or tree-like structure tracking model [39]. Among these model-based methods, the active contour models with level set evolution are developed and classified into region-based and edge-based models with various improvements for image segmentation. It has recently been proven that the hybrid variants combining both region and edge information [34] can improve the vessel segmentation performance. But these methods are usually sensitive to model initialization and have limited performances in extracting the real vessels with intensity inhomogeneity as well as varied vessel shapes and topologies from many vessel-like blurry edges in complex backgrounds. Furthermore, the tube-like structure tracking models [36–38] have attracted some research attentions recently. However, due to the low-contrast intensity inhomogeneities, complex noise distributions, background

overlaps, pathology and surgical changes, vessel tracking methods always result in early termination of vessel evolution when the matching between the vessel model itself and the image data in the current model neighborhood is not easily computed by both the deterministic and statistical tracking approaches.

For the filtering methods, more advanced vessel filtering methods are proposed to lift original low-dimensional image data to a high-dimensional space exploring the lifted data's multidimensional and multiscale information about the local and nonlocal curvilinear features such as orientations, phases, intensity profiles, and high-level topological and geometric features [28,33]. The nonlocal similarity-based filtering [40,41] has proven to perform well in aggregating similar patches for X-ray image denoising. In addition, new nonlocal computation schemes including the multiscale superpixels handling different vessel scales [42], nonlocal weighted fuzzy energy term in active contour models [43], and the fractional calculus [44,45] enabling long-range interaction have proven to well enhance the edge- and ridge-like features from the noisy backgrounds in medical imaging. Furthermore, inspired by the phase congruency model [46] having robustness to noise and being invariant to changes in contrast, some image enhancement methods have exploited the phase congruency based feature indicator called phase symmetry and phase asymmetry [44,47,48] to detect edge- and ridge-like features such as 2D/3D vessels. The high-level vessel feature representations have benefited subsequent classification of enhanced vessel features from complex and dynamic backgrounds. However, this vessel-feature enhancement filtering simultaneously enhances the background structures with similar vessel-like features to introduce more difficulty in subsequent vessel classification or vessel tracking. The vessel feature representations are also very important in the model-based vessel tracking methods after model initialization, where the current model's neighboring topological and geometric features are exploited to steer the evolution of model-based vessel tracking.

Third, as medical imaging data and computer power become increasingly available in our life, various data-driven deep learning network architectures have learned the end-to-end mapping between training images and their corresponding manually annotated binary ground truth maps for medical image segmentation [49,50]. By assigning a vessel or nonvessel label to each pixel, typical supervised vessel segmentation generally consists of training and testing. In the training stage, the method simultaneously learns the features and classification parameters for classifying the pixels from known labels (ground truth). After that, in test stage, the trained classifier is examined on previously unseen pixels. Generally, most current deep learning-based vessel segmentation methods are built from the encoder−decoder architectures including fully convolutional networks and U-Net [15,16,18], which learn to map input data of arbitrary sizes into a high-level feature representation with contextual information by an encoding function and then predict the output with the same size from the feature representation and recover the pixel-wise spatial detail back to the original domain. On the one hand, fusing prior knowledge about target vessels into neural network architecture

such as channel attention and loss function [15], as well as preprocessing and postprocessing modules (such as conditional random field module) will certainly improve the vessel segmentation performance. On the other hand, incorporating graphical models [14] and classical model-based machine learning such as active contour models [50] into multiscale data-driven deep neural networks will enable a full exploration of contextual information about geometrical and topological features to effectively refine the vessel segmentation. Unfortunately, most existing vessel segmentation methods are solely developed for the purpose of vessel shape recovery. To the best of our knowledge, automated vessel extraction with inhomogeneous intensity recovery from XCA images is currently an unexplored task. This chapter therefore for the first time presents the complete procedures for heterogeneous vessel extraction with intensity recovery. Fig. 4.1 displays the whole procedure of this chapter: we first present motion coherency regularized RPCA called MCR-RPCA (https://github.com/Binjie-Qin/MCR-RPCA) [5] for extracting foreground vessels from the noisy backgrounds in Section 4.3. Our sequential vessel segmentation network called SVS-net [15] is presented in Section 4.4 to demonstrate its segmentation of entire XCA vessel networks. In Section 4.5, we introduce the work of vessel region background completion with t-TNN tensor completion (VRBC-t-TNN) for accurately recovering the heterogeneous vessel network's shapes and intensities. This heterogeneous-vessel extraction is ready for further microcirculation analysis in the clinic.

## 4.3 MCR-RPCA: motion coherency regularized RPCA for vessel extraction

### 4.3.1 Motivation and problems

We viewed the XCA images as high dimensional video frames that have moving contrast agents continuously flowing inside vascular networks, which are distributed and overlapped within the complex backgrounds and artifacts, resulting in a big challenge for XCA analysis [5]. Fortunately, we have observed that XCA data have two specific structural characteristics, for example, the sparsity of moving contrast agent, and the similarity of background tissue structures. Then we naturally asked ourselves the question: How to simultaneously leverage the sparsity of foreground contrast agents and the similarity of the sequential backgrounds? Low-rank plus sparse modeling [2] is an effective tool to detect sparse outliers from the observed background images that are approximated with low rankness. Specifically, RPCA has been proposed in the classic work of Candès et al. [3] by decomposing the whole video data into low-rank plus sparse matrices with a convex optimization problem. Thereafter various RPCA models are exploited to outperform the state-of-the-art algorithms in many applications [4].

To directly extract the contrast-filled blood flow in vessel trajectories, the spatio-temporal sparseness and consistency of vessel trajectories within RPCA framework have been exploited in our work MCR-RPCA [5]. Specifically, based on the sparse

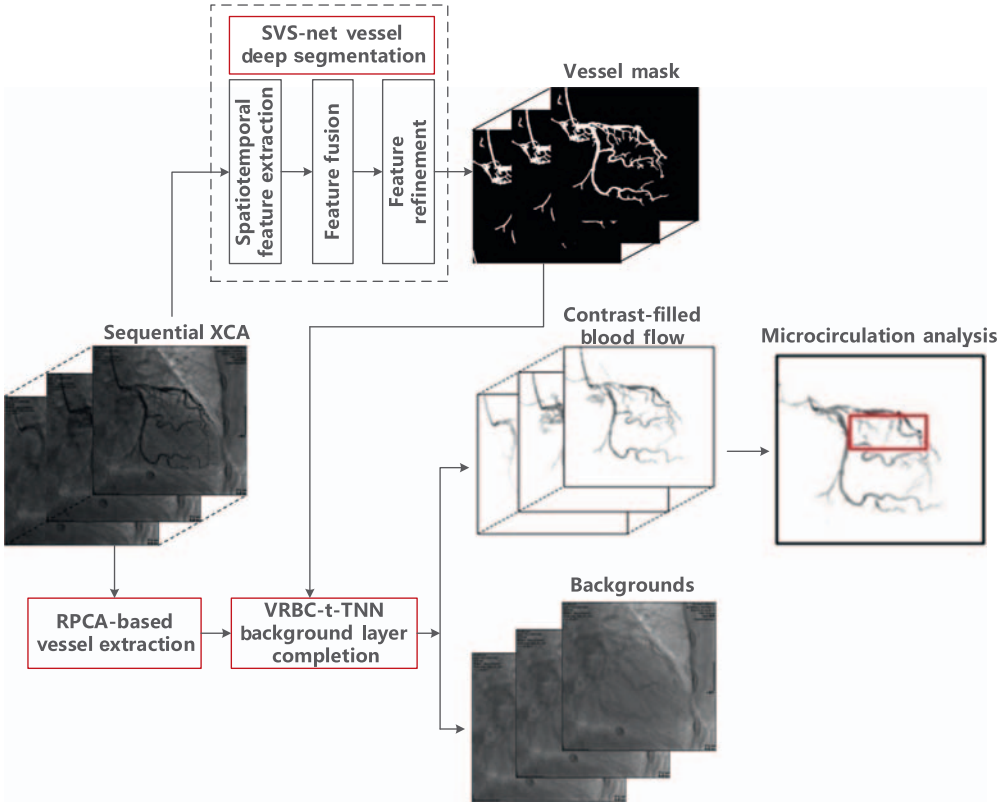


**Figure 4.1** The whole procedure of extracting heterogeneous vessels in XCA for microcirculation analysis. (1) The MCR-RPCA directly separated the heterogeneous foreground vessels from the backgrounds. (2) The deep network SVS-net segmented the vessel masks from the sequential XCA images. (3) The VRBC-t-TNN method was used to complete the background regions of vessel masks for subsequently subtracting the completed background layers from the overall XCA data to recover the heterogeneous vessel network's shapes and intensities. *XCA*, X-ray coronary angiography; *MCR-RPCA*, motion coherency regularized RPCA.

and low-rank modeling of foreground/background in XCA imaging, MCR-RPCA used a statistically structured RPCA first to identify all candidate foreground vessels. Then, we imposed total variation minimization on the foreground vessel trajectories in modeling the spatiotemporal contiguity and smoothness within RPCA modeling to eliminate the backgrounds remained in the candidates.

### 4.3.2 Candidate contrast-filled vessel detection via statistically structured MoG-RPCA

Observing that the candidate foreground vessel regions are sparsely clustered and correlated within curvilinear structures, we introduced structured sparsity constraint in RPCA

for extracting all candidate contrast-filled vessels. However, the real structural pattern in foreground regions is full of the overlapping structures of vessels and other tissues via 2D X-ray projection. To sparsely regularized the complex contrast-filled vessels in the sparse matrix $\mathbf{S}$, we associated the matrix $\mathbf{S}$ with noisy disturbances to model the moving contrast agents as a mixture of statistical structures and complex noise. Specifically, we applied mixture of Gaussian (MoG) model to statistically construct MoG-based RPCA [51] for separating the candidate contrast-filled vessels from the complex backgrounds. In fact, this MoG-RPCA has proven to achieve high detection rate of foreground moving object in some rigorous experimental evaluations using both synthetic and real videos [52].

Assuming $\mathbf{D} = \mathbf{L} + \mathbf{S}$ where $\mathbf{D}, \mathbf{L}, \mathbf{S} \in \mathbb{R}^{m \times n}$ are the original XCA sequence, low-rank backgrounds, and sparsely distributed vessels, respectively, we set each $s_{ij}$ in sparse matrix $\mathbf{S}$ with the following MoG distribution:

$$p\left(s_{ij}\middle|\mu,\tau,\mathbf{z}_{ij}\right) = \prod_{k=1}^{K}\mathcal{N}\left(s_{ij}\middle|\mu_k,\tau_k^{-1}\right)^{z_{ijk}} p\left(\mathbf{z}_{ij}\middle|\pi\right) = \prod_{k=1}^{K}\pi_k^{z_{ijk}} \tag{4.1}$$

where $\mathbf{z}_{ij}$ is associated with $s_{ij}$ to have a 1-of-$K$ indicator vector $z_{ijk} \in \{0,1\}, \sum_{k=1}^{K} z_{ijk} = 1$. $\pi = (\pi_1, \ldots \pi_K)$ is a mixing coefficient vector with $\pi_k \in \{0,1\}$ representing the existence probability of the $k$th Gaussian component having $\sum_{k=1}^{K}\pi_k = 1$. $\mu = \left(\mu_1, \ldots, \mu_K\right)$ and $\tau = (\tau_1, \ldots, \tau_K)$ with $\mu_k$ and $\tau_k$ being the mean and precision (inverse variances) of the $k$-th Gaussian component, respectively. Here $\mathbf{Z}$ is an $m \times n \times K$ array with the $(i,j,k)$th element denoted by $z_{ijk}$ to facilitate the following description.

The Gaussian-Gamma distribution was used to model the parameters $\mu_k$ and $\tau_k$ with the Dirichlet distribution modeling the mixing coefficient $\pi$

$$p\left(\mu_k,\tau_k\right) = \mathcal{N}\left(\mu_k\middle|\mu_{0k},\beta_0^{-1}\tau_k^{-1}\right)\cdot \mathrm{Gam}\left(\tau_k\middle|\alpha_0,\chi_0\right)\ \ p(\pi) = \mathrm{Dir}\left(\pi\middle|\eta_0\right) \tag{4.2}$$

where $\mu_{0k}$ is the mean of the $k$thGaussian component, $\mathrm{Gam}(\cdot)$ is the Gamma distribution with hyperparameters $\alpha_0$, $\beta_0$, and $\chi_0$ being set to be small deterministic value (e.g., $10^{-5}$) to obtain broad hyperpriors. $\mathrm{Dir}\left(\pi\middle|\eta_0\right)$ denotes the Dirichlet distribution parameterized by $\eta_0 = \left(\eta_{01}, \ldots, \eta_{0K}\right)$.

For low-rank background modeling in MoG-RPCA, the $\mathbf{L} \in \mathbb{R}^{m \times n}$ with rank $l \le \min(m,n)$ was set to be the product of $\mathbf{U} \in \mathbb{R}^{m \times R}$ and $\mathbf{V} \in \mathbb{R}^{n \times R}$ with $R > l$:

$$\mathbf{L} = \mathbf{U}\mathbf{V}^T = \sum\nolimits_{r=1}^{R}\mathbf{u}_{\cdot r}\mathbf{v}_{\cdot r}^T \tag{4.3}$$

where $\mathbf{u}_{\cdot r}(\mathbf{v}_{\cdot r})$ is the $r$th column of $\mathbf{U}(\mathbf{V})$. The low-rank nature of $\mathbf{L}$ is to achieve column sparsity in $\mathbf{U}$ and $\mathbf{V}$ by setting the following priors on $\mathbf{U}$ and $\mathbf{V}$:

$$p(\mathbf{u}_{\cdot r}) = \mathcal{N}\left(\mathbf{u}_{\cdot r}\middle|\mathbf{0},\gamma_r^{-1}\mathbf{I}_m\right), p(\mathbf{v}_{\cdot r}) = \mathcal{N}\left(\mathbf{v}_{\cdot r}\middle|\mathbf{0},\gamma_r^{-1}\mathbf{I}_n\right) \tag{4.4}$$

where $\mathbf{I}$ denotes identity matrix. The conjugate prior on each precision variable is

$$p(\gamma_r) = \mathrm{Gam}(\gamma_r|a_0, b_0) \propto {\gamma_r}^{a_0-1}\exp(-b_0\gamma_r) \tag{4.5}$$

where $a_0$ and $b_0$ are treated as small deterministic parameters to obtain broad hyperpriors. The common precision variable $\gamma_r$ leading to large precision values of some $\gamma_r$s can result in a good low-rank estimate of $\mathbf{L}$ [53]. After the above-mentioned combination of equations, we modeled the MoG-RPCA to achieve the candidate contrast-filled vessel detection, which turned to infer the posterior $p(\mathbf{U}, \mathbf{V}, \mathbf{Z}, \mu, \tau, \pi, \gamma|\mathbf{D})$ of all involved variables, where $\mathbf{Z} = \{\mathbf{z}_{ij}\}$, $\mu = (\mu_1, \ldots, \mu_K)$, $\tau = (\tau_1, \ldots, \tau_K)$, and $\gamma = (\gamma_1, \ldots, \gamma_R)$.

To get variational inference of the posterior of MoG-RPCA, we computed the true posterior $p(\mathbf{x}|\mathcal{D})$ with an approximation distribution $q(\mathbf{x})$ by minimizing the Kullback–Leibler (KL) divergence between $p(\mathbf{x}|\mathcal{D})$ and $q(\mathbf{x})$, that is, $\min_{q\in\mathcal{C}} \mathrm{KL}(q\|p) = -\int q(\mathbf{x})\ln\{p(\mathrm{x}|D)/q(\mathbf{x})\}\,\mathrm{d}\mathbf{x}$, where $\mathcal{C}$ denotes the set of probability densities with certain constraints to make the minimization tractable. If we partition the elements of $\mathbf{x}$ into disjoint groups $\{\mathbf{x}_i\}$, $q(\mathbf{x})$ is then generally assumed to be factorized as $q(\mathbf{x}) = \prod_i q_i(\mathbf{x}_i)$. The solution for each group $\{\mathbf{x}_j\}$, with the others fixed, can then be obtained by

$$q_j^*(\mathbf{x}_j) = \frac{\exp\left(\mathbb{E}_{i\neq j}[\ln p(\mathbf{x}, \mathcal{D})]\right)}{\int \exp\left(\mathbb{E}_{i\neq j}[\ln p(\mathbf{x}, \mathcal{D})]\right)\mathrm{d}\mathbf{x}_j} \tag{4.6}$$

where $p(\mathbf{x}, \mathcal{D})$ is the joint distribution of parameter $\mathbf{x}$ and observations $\mathcal{D}$, and $\mathbb{E}_{i\neq j}[\cdot]$ denotes the expectation with respect to $\mathbf{x}_i$s except $\mathbf{x}_j$. The solution for each group $\{\mathbf{x}_j\}$can then be approached by alternatively optimizing each $q_j(\mathbf{x}_j)$ by Eq. (4.6). The posterior distribution of MoG-RPCA is factorized as the following form:

$$q(\mathbf{U}, \mathbf{V}, \mathbf{Z}, \mu, \tau, \pi, \gamma) = \prod_i q(\mathbf{u}_{i\cdot})\prod_j q(\mathbf{v}_{j\cdot})\prod_{ij} q(\mathbf{z}_{ij})\prod_k q(\mu_k, \tau_k)\, q(\pi)\prod_r q(\gamma_r) \tag{4.7}$$

where $\mathbf{u}_{i\cdot}(\mathbf{v}_{j\cdot})$ is the $i$th ($j$th) row of $\mathbf{U}(\mathbf{V})$. The estimation of candidate foreground and low-rank background components via MoG-RPCA can then be derived by estimating all the factorized distributions involved in the above equation as follows.

#### 4.3.2.1 Estimation of candidate foreground component

To estimate the parameters $\mu, \tau, \mathbf{Z}$ and $\pi$ involved in the foreground vessel detection, we got the following update equation for each parameter using the prior imposed in Eq. (4.2) and its conjugate property:

$$q(\mu_k, \tau_k) = \mathcal{N}\left(\mu_k|m_k, (\beta_k\tau_k)^{-1}\right)\cdot\mathrm{Gam}(\tau_k|\alpha_k, \chi_k) \tag{4.8}$$

where

$$\beta_k = \beta_0 + \sum_{ij}\mathbb{E}\left[z_{ijk}\right],$$

$$m_k = \frac{1}{\beta_k}\left(\beta_0\mu_{0k} + \sum_{ij}\mathbb{E}\left[z_{ijk}\right]\left(d_{ij} - \mathbb{E}[\mathbf{u}_{i\cdot}]\mathbb{E}\left[\mathbf{v}_{j\cdot}\right]^T\right)\right), \alpha_k = \alpha_0 + \frac{1}{2}\sum_{ij}\mathbb{E}\left[z_{ijk}\right]$$

$$\chi_k = \chi_0 + \frac{1}{2}\left\{\sum_{ij}\mathbb{E}\left[z_{ijk}\right]\mathbb{E}\left[\left(d_{ij} - \mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T\right)^2\right] + \beta_0\mu_{0k}^2 - \frac{1}{\beta_k}\left(\sum_{ij}\mathbb{E}\left[z_{ijk}\right]\left(d_{ij} - \mathbb{E}[\mathbf{u}_{i\cdot}]\mathbb{E}\left[\mathbf{v}_{j\cdot}\right]^T\right) + \beta_0\mu_0\right)^2\right\}$$

The posterior mixing coefficients $\pi$ is similarly updated using the equation $q(\pi) = \mathrm{Dir}(\pi|\eta)$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K), \eta_k = \eta_{0k} + \sum_{ij}\mathbb{E}\left[z_{ijk}\right]$. The variational posterior for the indicators $\mathbf{Z}$ is computed by $q(\mathbf{z}_{ij}) = \prod_k r_{ijk}^{z_{ijk}}$, where $r_{ijk} = \delta_{ijk}/\sum_k \delta_{ijk}$,

$$\delta_{ijk} = 0.5\mathbb{E}[\ln\tau_k] - 0.5\ln 2\pi - 0.5\mathbb{E}[\tau_k]\mathbb{E}\left[\left(d_{ij} - \mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T - \mu_k\right)^2\right] + \mathbb{E}[\ln\pi_k] \tag{4.9}$$

#### 4.3.2.2 Estimation of low-rank background component

The posterior distribution for each row $\mathbf{u}_{i\cdot}$ of $\mathbf{U}$ (and $\mathbf{v}_{j\cdot}$ of $\mathbf{V}$) involved in low-rank background component can be approximated by

$$q(\mathbf{u}_{i\cdot}) = \mathcal{N}\left(\mathbf{u}_{i\cdot}\left|\boldsymbol{\mu}_{\mathbf{u}_{i\cdot}}, \sum\nolimits_{\mathbf{u}_{i\cdot}}\right.\right), q(\mathbf{v}_{j\cdot}) = \mathcal{N}\left(\mathbf{v}_{j\cdot}\left|\boldsymbol{\mu}_{\mathbf{v}_{j\cdot}}, \sum\nolimits_{\mathbf{v}_{j\cdot}}\right.\right) \tag{4.10}$$

where

$$\boldsymbol{\mu}_{\mathbf{u}_{i\cdot}}^T = \sum\nolimits_{\mathbf{u}_{i\cdot}}\left\{\sum\nolimits_k \mathbb{E}[\tau_k]\sum\nolimits_j \mathbb{E}\left[z_{ijk}\right]\left(d_{ij} - \mathbb{E}\left[\mu_k\right]\right)\mathbb{E}\left[\mathbf{v}_{j\cdot}\right]\right\}^T$$

$$\sum\nolimits_{\mathbf{u}_{i\cdot}} = \left\{\sum\nolimits_k \mathbb{E}[\tau_k]\sum\nolimits_j \mathbb{E}\left[z_{ijk}\right]\mathbb{E}\left[\mathbf{v}_{j\cdot}^T\mathbf{v}_{j\cdot}\right] + \Gamma\right\}^{-1}$$

$$\boldsymbol{\mu}_{\mathbf{v}_{j\cdot}}^T = \sum\nolimits_{\mathbf{v}_{j\cdot}}\left\{\sum\nolimits_k \mathbb{E}[\tau_k]\sum\nolimits_i \mathbb{E}\left[z_{ijk}\right]\left(d_{ij} - \mathbb{E}\left[\mu_k\right]\right)\mathbb{E}\left[\mathbf{v}_{j\cdot}\right]\right\}^T$$

$$\sum\nolimits_{\mathbf{v}_{j\cdot}} = \left\{\sum\nolimits_k \mathbb{E}[\tau_k]\sum\nolimits_i \mathbb{E}\left[z_{ijk}\right]\mathbb{E}\left[\mathbf{u}_{i\cdot}^T\mathbf{u}_{i\cdot}\right] + \Gamma\right\}^{-1}, \quad \Gamma = \mathrm{diag}(\mathbb{E}[\gamma])$$

We displayed the experimental results for the candidate vessel extraction (see the middle column of Fig. 4.2), which leave some noisy artifacts in the vessel layers for the subsequent vessel refinement via trajectory decomposition in MCR-RPCA.

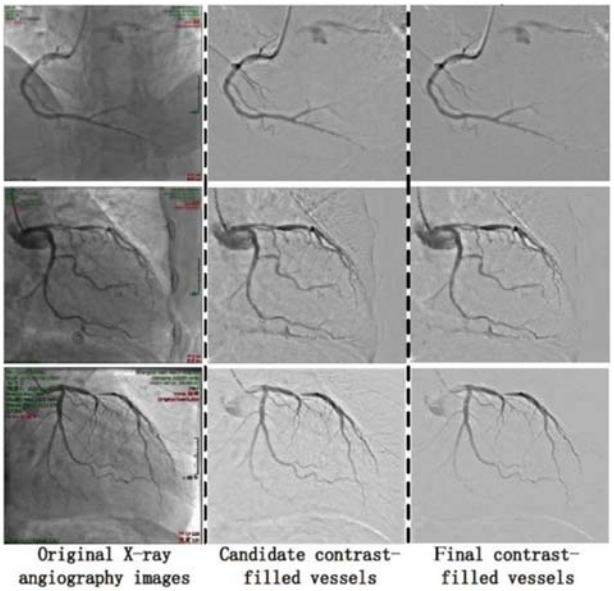


**Figure 4.2** Vessel extraction examples via MCR-RPCA [5]. From left to right: the original images, the candidate foreground vessels via the statistically structured MoG-RPCA, the final vessels via the motion coherency regularized RPCA.

### 4.3.3 Motion coherency regularized RPCA for trajectory decomposition

Since the statistically structured RPCA is unable to encode the long-term temporal coherency of XCA sequence, there still have some parts of noisy backgrounds in the extracted candidate foregrounds. Inspired by the trajectory decomposition works [54] that have preserved temporal consistency by assigning similar lables to the pixels belonging to the same trajectories, we exploited the trajectory properties of blood vessels by directly representing the motion coherency as spatiotemporal contiguity and smoothness of the image pixels to refine the vessel extraction in the subsequent trajectory decomposition.

The sparsely distributed sinuous courses of blood flow in vessels make the foreground trajectories occupy a set of contiguous regions throughout image sequences. Therefore, being different from the traditional $L_{2,1}$ norm-based group sparsity constraint $\|\mathbf{S}\|_{2,1}$ [2] that treats each column of matrix as a whole, the $L_1$-norm-based pixel-wise sparsity constraint is more appropriate for the foreground trajectory extraction task. Then, the total variation minimization is further exploited as a motion coherency constraint for the contrast-filled vessel extraction. Based on a difference operator, the weighted TV regularization of matrix $\mathbf{S}$ can be defined as $\|\mathbf{S}\|_{\mathrm{TV}} = w_x\|\nabla_x\mathbf{S}\|_1 + w_y\|\nabla_y\mathbf{S}\|_1 + w_z\|\nabla_z\mathbf{S}\|_1$ with default setting $w_x = w_y = w_z = 1$, where $\nabla_x$, $\nabla_y$, and $\nabla_z$ correspond to difference operations performed along $x$ (or

row), $y$ (or column), and $z$ (or frame) directions, respectively. The combination of TV and $L_1$-norm of matrix $\mathbf{S}$ will encourage the proximate trajectories with similar appearances to be grouped into the same foreground vessels. The motion coherency regularized RPCA model is carried out as follows:

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{S}\|_{\mathrm{TV}} \quad \text{s.t. } \mathbf{D} = \mathbf{L} + \mathbf{S} \tag{4.11}$$

where $\mathbf{D}$ represents the candidate foreground matrix, $\mathbf{L}$ and $\mathbf{S}$ represent new low-rank background and sparse foreground trajectories, respectively. The $\lambda_1$ and $\lambda_2$ are regularizing parameters. We used alternating direction method (ADM) to solve the minimization problem in Eq. (4.11) with the solver Inexact Augmented Lagrangian Multiplier Method (IALM) [53] by rewriting the Eq. (4.11) into another form:

$$\min_{\mathbf{L},\mathbf{S},\mathbf{T}} \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{T}\|_{\mathrm{TV}}, \quad \text{s.t. } \mathbf{D} = \mathbf{L} + \mathbf{S} \quad \mathbf{T} = \mathbf{S} \tag{4.12}$$

The augmented Lagrangian function of Eq. (4.12) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{L},\mathbf{S},\mathbf{T},\mathbf{X},\mathbf{Y};\mu) = \quad & \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{T}\|_{\mathrm{TV}} + \frac{\mu}{2} \|\mathbf{D}-\mathbf{L}-\mathbf{S}\|_F^2 \\ & + \langle \mathbf{X}, \mathbf{D}-\mathbf{L}-\mathbf{S} \rangle + \frac{\mu}{2} \|\mathbf{T}-\mathbf{S}\|_F^2 + \langle \mathbf{Y}, \mathbf{T}-\mathbf{S} \rangle \end{aligned} \tag{4.13}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the Lagrangian multipliers, and $\mu$ is a positive penalty scalar. We summarized the solutions of the subproblems based on the ADM strategy in Algorithm 1.

**Algorithm 1** ADM for motion coherency regularized RPCA

**Input:** Matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$, $\lambda_1 > 0$, $\lambda_2 > 0$

1: Initializing: $\mathbf{L}$, $\mathbf{S}$ and $\mathbf{X}$, $\mathbf{Y}$

2: **while** not converged **do**

3: **L sub-problem:**

$$\begin{aligned}\mathbf{L}_{k+1} &= \arg\min_{\mathbf{L}} \mathcal{L}(\mathbf{L}, \mathbf{S}_k, \mathbf{T}_k, \mathbf{X}_k, \mathbf{Y}_k, \mu) \\ &= \arg\min_{\mathbf{L}} \|\mathbf{L}\|_* + \frac{\mu}{2}\|(\mathbf{D} - \mathbf{S}_k + \mu^{-1}\mathbf{X}_k) - \mathbf{L}\|_F^2\end{aligned}$$

solved by:

$$\begin{aligned}(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) &= \text{svd}(\mathbf{D} - \mathbf{S}_k + \mu^{-1}\mathbf{X}_k) \\ \mathbf{L}_{k+1} &= \mathbf{U}\mathcal{S}_{\mu^{-1}}(\mathbf{\Sigma})\mathbf{V}^T\end{aligned}$$

4: **S sub-problem:**

$$\mathbf{S}_{k+1} = \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{L}_{k+1}, \mathbf{S}, \mathbf{T}_k, \mathbf{X}_k, \mathbf{Y}_k, \mu)$$

solved by the soft shrinkage operator.

5: **T sub-problem:**

$$\begin{aligned}\mathbf{T}_{k+1} &= \arg\min_{\mathbf{T}} \mathcal{L}(\mathbf{L}_{k+1}, \mathbf{S}_{k+1}, \mathbf{T}, \mathbf{X}_k, \mathbf{Y}_k, \mu) \\ &= \arg\min_{\mathbf{T}} \lambda_2\|\mathbf{T}\|_{\text{TV}} + \frac{\mu}{2}\|\mathbf{T} - \mathbf{S}_{k+1}\| + \langle \mathbf{Y}_k, \mathbf{T} - \mathbf{S}_{k+1}\rangle \\ &= \arg\min_{\mathbf{T}} \lambda_2\|\mathbf{T}\|_{\text{TV}} + \frac{\mu}{2}\|\mathbf{T} - \mathbf{S}_{k+1} + \mu^{-1}\mathbf{Y}_k\|_F^2\end{aligned}$$

solved by split Bregman method.

6: Update Lagrange multiplier $\mathbf{X}$ and $\mathbf{Y}$:

$$\begin{aligned}\mathbf{X}_{k+1} &= \mathbf{X}_k + \mu(\mathbf{D} - \mathbf{L}_{k+1} - \mathbf{S}_{k+1}) \\ \mathbf{Y}_{k+1} &= \mathbf{Y}_k + \mu(\mathbf{T}_{k+1} - \mathbf{S}_{k+1})\end{aligned}$$

7: **end while**

**Output:** $(\mathbf{L}_k, \mathbf{S}_k)$

Using the ADM algorithm, we alternately optimized one variable ($\mathbf{L}$, or $\mathbf{S}$, or $\mathbf{T}$) with the other two variables fixed: for $\mathbf{L}$ subproblem optimization, it has closed-form solution by a soft shrinkage operator $\mathcal{S}_{\mu^{-1}}(\gamma) = \text{sgn}(\gamma)\max(|\gamma| - \mu^{-1}, 0)$ with a threshold $\mu^{-1}$ for a scalar $\gamma$, which operator is extended entry-wisely to vectors and matrices; $\mathbf{S}$ subproblem is solved by a shrinkage operator with the implementation detail at the following section; the split Bregman method [55] is adopted to solve $\mathbf{T}$ subproblem optimization. In Algorithm 1, $\lambda_1 = 0.5/\sqrt{\max(m, n)}$ and $\lambda_2 = 0.2/\sqrt{\max(m, n)}$, which can perform well on most experimental data.

For the $\mathbf{S}$ subproblem optimization in Algorithm 1, the objective function is

$$\begin{aligned}\mathbf{S}_{k+1} &= \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{L}_{k+1}, \mathbf{S}, \mathbf{T}_k, \mathbf{X}_k, \mathbf{Y}_k, \mu) \\ &= \arg\min_{\mathbf{S}} \lambda_1 \|\mathbf{S}\|_1 + \frac{\mu}{2}\|\mathbf{D}-\mathbf{L}_{k+1}-\mathbf{S}\|_F^2 + \langle \mathbf{X}_k, \mathbf{D}-\mathbf{L}_{k+1}-\mathbf{S}\rangle \\ &\quad + \frac{\mu}{2}\|\mathbf{T}_k-\mathbf{S}\|_F^2 + \langle \mathbf{Y}_k, \mathbf{T}_k-\mathbf{S}\rangle \\ &= \arg\min_{\mathbf{S}} \lambda_1 \|\mathbf{S}\|_1 + \mu\|\mathbf{Q}-\mathbf{S}\|_F^2 \end{aligned} \tag{4.14}$$

where we denote $\mathbf{Q} = \frac{\mathbf{D}-\mathbf{L}_{k+1}+\mathbf{T}_k+\mu^{-1}(\mathbf{X}_k+\mathbf{Y}_k)}{2}$. The objective function can be further expressed as

$$\mathbf{S}_{k+1} = \arg\min_{\mathbf{S}\in\mathbb{R}^{m\times n}} \sum_{i=1}^{m}\sum_{j=1}^{n}(\lambda_1|\mathbf{S}_{i,j}| + \mu|\mathbf{Q}_{i,j}-\mathbf{S}_{i,j}|^2) \tag{4.15}$$

Eq. (4.15) can be minimized for each $\mathbf{S}_{i,j}$ separately. The solution $\mathbf{S}_{k+1}$ for $\mathbf{S}$ subproblem is obtained by the soft shrinkage operator $\mathbf{S}_{k+1} = \mathcal{S}_{\lambda_1(2\mu)^{-1}}(\mathbf{Q})$, where the operator $\mathcal{S}$ is performed element-wisely.

As for the $\mathbf{T}$ subproblem optimization in Algorithm 1, the objective function is as follows:

$$\begin{aligned}\mathbf{T}_{k+1} &= \arg\min_{\mathbf{T}} \mathcal{L}(\mathbf{L}_{k+1}, \mathbf{S}_{k+1}, \mathbf{T}, \mathbf{X}_k, \mathbf{Y}_k, \mu) \\ &= \arg\min_{\mathbf{T}} \lambda_2||\mathbf{T}||_{\mathrm{TV}} + \frac{\mu}{2}||\mathbf{T}-\mathbf{S}_{k+1}|| + \langle \mathbf{Y}_k, \mathbf{T}-\mathbf{S}_{k+1}\rangle \\ &= \arg\min_{\mathbf{T}} \lambda_2||\mathbf{T}||_{\mathrm{TV}} + \frac{\mu}{2}||\mathbf{T}-(\mathbf{S}_{k+1}-\mu^{-1}\mathbf{Y}_k)||_F^2 \\ &= \arg\min_{\mathbf{T}} \lambda_2(||\nabla_x\mathbf{T}||_1 + ||\nabla_y\mathbf{T}||_1 + ||\nabla_z\mathbf{T}||_1 + \frac{\mu}{2}||\mathbf{T}-(\mathbf{S}_{k+1}-\mu^{-1}\mathbf{Y}_k)||_F^2 \end{aligned} \tag{4.16}$$

To apply Bregman splitting method, we first replaced $\nabla_x\mathbf{T}$ by $d_x$, $\nabla_y\mathbf{T}$ by $d_y$, and $\nabla_z\mathbf{T}$ by $d_z$ to yield the following constrained problem:

$$\begin{aligned}\mathbf{T}_{k+1} &= \arg\min_{\mathbf{T}} \lambda_2\left(||d_x||_1 + ||d_y||_1 + ||d_z||_1\right) + \frac{\mu}{2}||\mathbf{T}-\mathbf{K}||_F^2 \quad \text{s.t.} d_x = \nabla_x\mathbf{T}, d_y \\ &= \nabla_y\mathbf{T}, d_z = \nabla_z\mathbf{T} \end{aligned} \tag{4.17}$$

where $\mathbf{K} = \mathbf{S}_{k+1} - \mu^{-1}\mathbf{Y}_k$. To weakly enforce the constraints in this formulation, we added penalty function terms to obtain the following objective function:

$$\mathbf{T}_{k+1} = \arg\min_{d_x,d_y,d_z,\mathbf{T}} \lambda_2(||d_x||_1 + ||d_y||_1 + ||d_z||_1) + \frac{\mu}{2}||\mathbf{T} - \mathbf{K}||_F^2 + \frac{\lambda_t}{2}||d_x - \nabla_x\mathbf{T}||_F^2 + \frac{\lambda_t}{2}||d_y - \nabla_y\mathbf{T}||_F^2 + \frac{\lambda_t}{2}||d_z - \nabla_z\mathbf{T}||_F^2 \quad (4.18)$$

where $\lambda_t$ represents the weight of penalty function. Finally, we strictly enforced the constraints by applying the Bregman iteration [55] to get

$$\mathbf{T}_{k+1} = \arg\min_{d_x,d_y,d_z,\mathbf{T}} \lambda_2(||d_x||_1 + ||d_y||_1 + ||d_z||_1) + \frac{\mu}{2}||\mathbf{T} - \mathbf{K}||_F^2 + \frac{\lambda_t}{2}||d_x - \nabla_x\mathbf{T} - b_x^k||_F^2 + \frac{\lambda_t}{2}||d_y - \nabla_y\mathbf{T} - b_y^k||_F^2 + \frac{\lambda_t}{2}||d_z - \nabla_z\mathbf{T} - b_z^k||_F^2 \quad (4.19)$$

where the proper values of $b_x^k$, $b_y^k$, and $b_z^k$ are chosen through Bregman iteration. We can efficiently solve the minimization problem in Eq. (4.19) by iteratively minimizing it with respect to $d$ and $\mathbf{T}$ separately. For the solution of $\mathbf{T}$, we have the following subproblem:

$$\mathbf{T}_{k+1} = \arg\min_{\mathbf{T}} \frac{\mu}{2}||\mathbf{T} - \mathbf{K}||_F^2 + \frac{\lambda_t}{2}||d_x^k - \nabla_x\mathbf{T} - b_x^k||_F^2 + \frac{\lambda_t}{2}||d_y^k - \nabla_y\mathbf{T} - b_y^k||_F^2 + \frac{\lambda_t}{2}||d_z^k - \nabla_z\mathbf{T} - b_z^k||_F^2 \quad (4.20)$$

with an optimality condition

$$(\mu\mathbf{I} - \lambda_t\Delta)\mathbf{T}_{k+1} = \mu\mathbf{K} + \lambda_t\nabla_x\mathbf{T}(d_x^k - b_x^k) + \lambda_t\nabla_y\mathbf{T}(d_y^k - b_y^k) + \lambda_t\nabla_z\mathbf{T}(d_z^k - b_z^k). \quad (4.21)$$

In order to achieve optimal efficiency, we used a fast iterative algorithm to get an approximate solution to the problem in Eq. (4.20). Due to the problem being strictly diagonally dominant, a natural component-wise solution of Gauss−Seidel method for this problem should be represented as $\mathbf{T}_{i,j}^{k+1} = \mathbf{G}_{i,j}^k$

$$\begin{aligned}\mathbf{G}_{i,j}^k = \frac{\lambda_t}{\mu + 6\lambda_t}(&\mathbf{T}_{i+1,j,l}^k + \mathbf{T}_{i-1,j,l}^k + \mathbf{T}_{i,j+1,l}^k + \mathbf{T}_{i,j-1,l}^k + \mathbf{T}_{i,j,l+1}^k + \mathbf{T}_{i,j,l-1}^k \\ &+ d_{x,i-1,j,l}^k - d_{x,i,j,l}^k + d_{y,i,j-1,l}^k - d_{y,i,j,l}^k + d_{z,i,j,l-1}^k - d_{z,i,j,l}^k \\ &- b_{x,i-1,j,l}^k + b_{x,i,j,l}^k - b_{y,i,j-1,l}^k + b_{y,i,j,l}^k - b_{z,i,j,l-1}^k + b_{z,i,j,l}^k) + \frac{\mu}{\mu + 6\lambda_t}\mathbf{K}_{i,j,l}\end{aligned} \quad (4.22)$$

where $\mathbf{T}^{k+1}$ has the same meaning as $\mathbf{T}_{k+1}$ in Eqs. (4.17) and (4.21), here is for the sake of convenience.

Table 4.1 demonstrates the performance comparison of vessel extraction via our MCR-RPCA and other classical RPCA-based methods. The performance is evaluated in terms of detection rate, precision, F-measure, and contrast-to-noise ratio (CNR) that measure the contrast between the foreground and background pixel intensities in relation to the standard deviation of the background pixel intensities. A larger CNR value implies a better contrast and thus a better performance of foreground/background separation. We used two different versions of CRN by defining two different background masks in XCA images. We refer the interested readers to Ref. [5] and the references therein.

## 4.4 SVS-net: sequential vessel segmentation via channel attention network

Recently, deep learning has become the most widely used approach for cardiac image segmentation [12] which covers common imaging modalities including MRI, CT, and ultrasound. However, due to the overlapping structures, low contrast, and the presence of complex and dynamic background artifacts in XCA images, accurately segmenting contrast-filled vessels from the XCA image sequence is particularly challenging for both classical modal-driven and modern data-driven deep learning methods. Being different from current several works that solely focused on segmenting main vessel from XCA [63–65], we have developed a novel encoder–decoder deep sequential vessel segmentation network (SVS-net, https://github.com/Binjie-Qin/SVS-net) [15] for entire XCA vessel network segmentation, which exploits several contextual frames of 2D + t sequential

**Table 4.1** Performance comparison of eight RPCA methods on the X-ray coronary angiography sequences.

| Method | Detection rate | Precision | *F*-measure | CNR1 | CNR2 | Time |
|---|---|---|---|---|---|---|
| MCR-PCA[1] | 0.71258 | 0.83955 | 0.76976 | 2.88208 | 3.27754 | 919.34 s |
| Block-RPCA [56][2] | 0.70313 | 0.83132 | 0.7591 | 2.57224 | 2.93497 | 4 h |
| DECOLOR [57][3] | 0.68435 | 0.84038 | 0.75024 | 2.62923 | 2.69954 | 148.5 s |
| IALM-BLWS [58][4] | 0.68957 | 0.84053 | 0.75618 | 2.58344 | 2.95259 | 74.84 s |
| FPCP [59][5] | 0.67088 | 0.76977 | 0.71303 | 2.39507 | 2.22859 | 0.98 s |
| GoDec [60][6] | 0.656308 | 0.77700 | 0.70826 | 2.34807 | 2.12574 | 1.41 s |
| TFOCS-EC [61][7] | 0.63397 | 0.79817 | 0.70273 | 2.34603 | 2.11375 | 304.37 s |
| PRMF [62][8] | 0.68573 | 077546 | 0.72556 | 2.56356 | 2.45889 | 42.06 s |

[1]https://github.com/Binjie-Qin/MCR-RPCA.
[2]https://www.ece.nus.edu.sg/stfpage/eleclf/.
[3]http://bioinformatics.ust.hk/decolor/decolor.html.
[4]https://www.ece.nus.edu.sg/stfpage/eleclf/.
[5]https://sites.google.com/a/istec.net/prodrig/Home/en/pubs.
[6]https://github.com/andrewssobral/godec/.
[7]https://github.com/cvxr/TFOCS.
[8]http://winsty.net/prmf.html.

images in a sliding window centered at current frame to continuously segment 2D vessel masks from the current frame in the XCA sequence.

### 4.4.1 Architecture of sequential vessel segmentation-network

Being modified from classical U-net architecture [66,67], the SVS-net architecture has three main modules that are equipped with spatiotemporal (2D + t) feature extraction in encoder module, 3D feature fusion operation (FFO) in skip connections, and channel attention block (CAB) in decoder module (See Fig. 4.3). Specifically, (1) A sequence of 3D convolutional layers hierarchically extracted spatiotemporal 3D

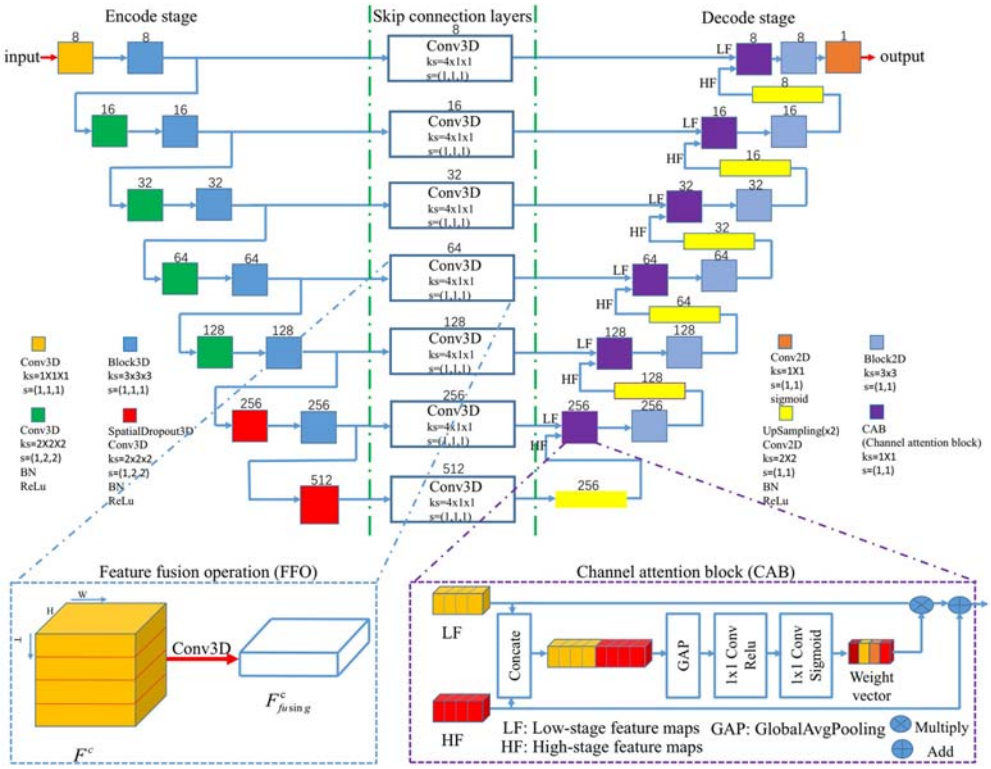


**Figure 4.3** The proposed sequential vessel segmentation-network architecture, which is based on U-Net with the encoder network extracting 3D feature from the input sequence and the decoder network learning the salient feature via upsampling and operation of CAB, between the encoder and decoder network is the skip connection layer with FFO. The numbers 8, 16, 32,... above each block denote the number of output channels for that block. Each block has different convolutional kernel sizes and strides (s: strides). In the FFO and CAB in the bottom, the $F \in \mathbb{R}^{C \times T \times H \times W}$ denotes the temporal-spatial feature maps, $C$ denotes channel axis, $T$ denotes temporal axis, $H$ denotes height axis, $W$ denotes width axis, $F^c \in \mathbb{R}^{C \times H \times W}$: the $c$th channel of temporal-spatial feature maps. $F^c_{fusing} \in \mathbb{R}^{H \times W}$ denotes the $c$th channel of fused temporal-spatial feature map through Conv3D with kernel size $4 \times 1 \times 1$ and strides (1,1,1) [15].

features in the encoder module. We introduced 3D residual blocks to extract multi-scale spatiotemporal features while easing network optimization in the encoder module; (2) Skip connections that connect the encoder and decoder subsequently fused the spatiotemporal feature maps along temporal axis and delivered the fused 2D spatial feature maps to the corresponding decoder stages. Through the 3D feature fusion in the left bottom of Fig. 4.3, the feature maps' dimension mismatch problems between the 3D encoder stage and the 2D decoder stage was solved with computation cost being simultaneously reduced; (3) The decoder module utilized CAB to refine the intermediate feature maps from skip connections by learning discriminative feature representation and suppressing the complex and noisy artifacts in the XCA images; (4) Dice loss function was further implemented to train the SVS-net in order to solve the class imbalance problem in the XCA data causing by the imbalanced ratio between backgrounds and foreground vessels.

#### *4.4.1.1 Modification of U-net*

The U-Net was chosen as a fundamental architecture for SVS-net due to its being a classical and powerful segmentation network architecture widely used for biomedical imaging, which can improve the spatial accuracy of a deep neural network for final high-resolution segmentation results by effectively exploring underlying both high-resolution and low-resolution information in biomedical images and transmitting multiscale information from encoder network to decoder network via skip connections. Specifically, to accurately segment vessels from XCA with low contrast and fuzzy boundary, the skip connection mechanism in U-Net allowed high-resolution information delivery to the decoder network for detail recovery. Furthermore, the low-resolution semantic information about the internal tissues and their topologies in XCA images can be simply provided by multiple downsampling operations in U-Net's encoder network for the target vessel recognition.

To achieve accurate XCA vessel segmentation, we have designed a U-Net-based deep learning architecture with several notable features: (1) The encoder network was modified to extract 3D spatiotemporal contexts through 3D convolutions followed by 3D residual convolutional blocks except the last convolutional operation. There were totally seven stages of 3D convolutions with the first six convolutional stages followed by 3D residual convolutional block, which were utilized to extract rich spatiotemporal feature contexts for subsequent vessel mask extraction in the decoder network. The output of each 3D residual convolutional block was linked to the next 3D convolutional stage and skip connection layer, respectively. The spatial dropout (0.5) at the last two 3D convolutional stages was employed before executing convolution to avoid overfitting. The encoder network was capable of effectively extracting 3D spatiotemporal features from the context of sequential XCA images which contain long-range vessel details of different scales and different vessel types that may not appear in a single image.

(2) In the skip connection layer, the spatiotemporal feature representations were fused by mapping from 3D space to 2D space via $4 \times 1 \times 1$ convolutional kernels in FFO block, where the first dimension of convolutional kernel indicates four channels (frames) in the temporal axis. The temporal domain features from these four frames were then fused together by temporal axis convolution. The computation of FFO can be formulated as follows:

$$X_{F_l} = \text{Squeeze}(X_F \otimes W) \tag{4.23}$$

where $X_F \in \mathbb{R}^{C \times T \times H \times W}$ is the spatiotemporal feature map outputted from each 3D convolutional stage in the encoder network with the $C, T, H, W$ being features' channel dimension, temporal dimension, height, and width, respectively. $X_{F_L} \in \mathbb{R}^{C \times H \times W}$ fuses the spatiotemporal feature map. $W$ is a $4 \times 1 \times 1$ convolutional kernel with $\otimes$ representing convolution operation. Squeeze denotes dimension compress.

(3) In the decoder network, the parameter-free bilinear upsampling strategy rather than transposed convolutional operations was adopted to gradually recover the feature maps' spatial resolution, which can reduce the number of trainable parameters without degrading the segmentation performance. A sequence of CAB modules that follow upsampling layer was designed to gradually fuse low-stage and high-stage XCA feature maps, whose output was further linked to 2D residual convolutional blocks (Block2D, see Fig. 4.3). The two outputs of low-stage feature maps from the skip connection layer and the high-stage feature maps from the upsampling layer were feed simultaneously in CAB (as illustrated in the right bottom of Fig. 4.3) to learn the most discriminative features from noisy and complex background artifacts (see details in Section 4.4.1.4). After the last 2D residual convolutional block in the end of sequential decoder networks, we employed $1 \times 1$ convolution followed by sigmoid activation function to yield final high-resolution vessel mask.

#### 4.4.1.2 3D spatiotemporal feature encoder

For a set of frames $(F_1, F_2, \ldots F_n)$ in the XCA image sequence, the contextual information such as the low contrast intensity distribution and the similar appearances embedded in the successive frames are helpful to infer whether one pixel of each frame belongs to either the foreground or background. Specifically, in these successive frames, blood flow in the contrast-filled vessel regions moves fast and consistently through the contiguous frames and the noisy backgrounds fluctuate synchronously along with human breathing and heart beating. Therefore these consistent contexts can provide spatiotemporal information to accurately identify vessels from backgrounds. The SVS-net accepts inputs from the successive four frames (i.e., $F_{i-2}, F_{i-1}, F_i, F_{i+1}$) to predict a binary probability map (i.e., $P_i$) where the foreground vessel pixel values equal 1 while other background pixel values equal 0.

To verify the choice of four frames as an appropriate input configuration, we chose different frames, that is, 2, 3, 4, 5 frames, as the input of network to investigate the network's different convergence performances. We observed some subtle differences of convergence results on Dice loss function in our experiment results [15]. When inputs from four frames were fed in the network, the loss function converged at about −0.86, which is the smallest compared with those of other input strategies. The choice of four frames is thus appropriate to be served as the input configuration for accurate vessel segmentation.

#### 4.4.1.3 2D and 3D residual convolutional blocks

Increasing the depth of a neural network can improve network generalization capacity, but it introduces the difficulty in promoting gradient backpropagation for the training of deep neural network. To ease the training of deep neural network, He et al. [68] developed deep residual network to facilitate gradient backpropagation by explicitly reformulating the network layers as learning residual functions with reference to the layer inputs instead of learning unreferenced functions. As demonstrated in the work of wide residual networks [69], two stacked convolutional layers in single residual block are optimal architecture compared with other settings. Inspired by these works, we employed 3D residual blocks and 2D residual blocks in encoder and decoder networks respectively. Between the encoder and decoder networks, the skip connection layer with FIFO was added to increase the depth and improve the accuracy of deep CNNs.

#### 4.4.1.4 Channel attention mechanism

Aiming to highlight the most dominant target regions in an image, visual attention mechanism has been widely developed from traditional model-driven image computation to modern data-driven deep learning for visual saliency detection in both cognitive and computational sciences. The visual saliency computations for single image, image pair, and video sequence have been widely exploited in numerous applications to improve their performance for object detection [70], image retrieval [71], image registration [72–74], image cosegmentation [75,76], and video analysis [77]. Attention mechanism in deep learning can be viewed as a generalization of estimating weights for a weighted average in a regression model, where weights encode the relevance of the training instance to the query. Visual attention has become popular in many deep learning-based computer vision tasks to focus on relevant regions within the image and capture structural long-range dependencies between the images of video sequences. Furthermore, attention modeling is particularly important for the interpretability of deep learning architectures since the magnitude of attention is assumed to be a particular weight that correlates with how relevant a specific region of input is for the prediction of output at each position in a spatiotemporal sequence.

To represent rich multiscale features for extracting target vessels from complex and dynamic backgrounds, the SVS-net first exploits encoder network to extract multiple types of features from the whole XCA images by multiple convolutional kernels in every convolutional stage of the encoder module (see Fig. 4.3). Through the attention mechanism of CAB assigning different weights to the extracted features, the SVS-net can adaptively highlight some channel information meanwhile suppress the trivial channel information containing noisy background disturbances. Hence, the predicted probability map is gradually improved. Specifically, the feature maps of the low-stage output from the skip connection layer are weighted in CAB and then combined with the corresponding high-stage feature maps output from the upsampling layer. Since high-stage output feature maps provide more advanced global semantic information while low-stage feature maps contain more detailed yet noisy information, the high-stage features can then be used as guide clues to filter useful information from low-stage feature maps and achieve more pure feature representation for target vessel. The core modules of CAB is presented as follows (see the CAB in the bottom right of Fig. 4.3):

Global average pooling (GAP) can be considered as a simplest version of channel-wise attention, which computes the channel-wise statistic $G \in \mathbb{R}^{C \times 1 \times 1}$ of the given input feature maps by averaging across each $H \times W$ spatial dimension. The $c$th channel statistic $s_c$ of $s$ is computed as follows:

$$G_c(X) = \frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} X_c(h, w) \tag{4.24}$$

where $X_c \in \mathbb{R}^{H \times W}, \forall c \in \{1, \ldots, C\}$ refers to the $c$th feature of the input feature map. In our work, the low-stage feature maps $X_{F_l} \in \mathbb{R}^{C \times H \times W}$ and the corresponding high-stage feature maps $X_{F_h} \in \mathbb{R}^{C \times H \times W}$ are concatenated together to make feature maps $X_F \in \mathbb{R}^{2C \times H \times W}$. Then a GAP is performed on the concatenated feature maps to generate a weighted vector $W^{X_F} \in \mathbb{R}^{2C \times 1 \times 1}$. Two $1 \times 1$ convolutional operations, which are followed by the rectified linear unit function and sigmoid function, respectively, are performed on $W_{X_F} \in \mathbb{R}^{2C \times 1 \times 1}$ to learn interchannel relationship and the final channel attention weight vector $W_{X_F} \in \mathbb{R}^{C \times 1 \times 1}$ is achieved. The obtained attention vector multiplies low-stage feature maps in channel-wise manner, then the weighted feature maps from low stage are added with the corresponding high-stage feature maps to be subsequently passed to the next layer. The whole process of generating attention weights can be expressed as:

$$W_{X_F} = \phi(\varphi(G(X_F))) \tag{4.25}$$

where $G$ means the operation of GAP, $\varphi$ denotes $1 \times 1$ convolution followed by rectified linear unit and $\phi$ indicates $1 \times 1$ convolution followed by sigmoid activation. An intuitive display of CAB is shown in the left bottom of Fig. 4.4. Under the guidance of high-stage features, the attention weights are learned and used to obtain discriminative

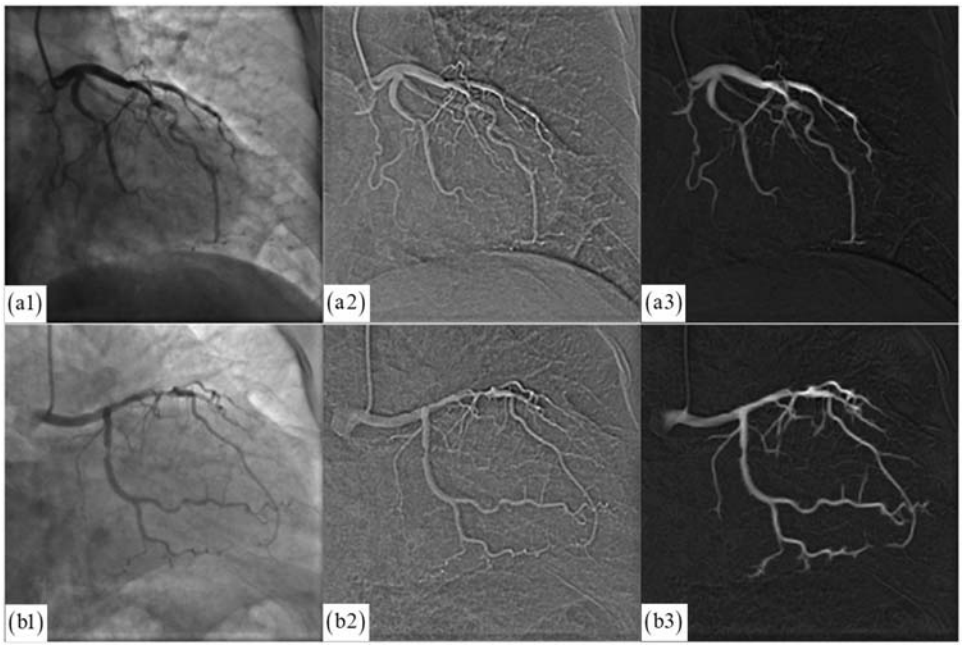


**Figure 4.4** Illustrating the CAB's effects on suppressing the noises in the background areas while highlighting the foreground vessel feature. From left to right, each row displays the original XCA image; the second channel of fused spatial feature maps in the output of the second skip connection layer (Fig. 4.3) before inputting to the CAB, which contains noise pollution from the background areas; the second channel of refined feature maps from the output of CAB in the decoder stage (Fig. 4.3).

salient features for target vessels. As displayed in Fig. 4.4(a3)(b3), CAB has successfully refined the low-stage feature map from the output of skip connection. The background disturbances in Fig. 4.4(a2)(b2) are greatly suppressed while the foreground vessel features are highlighted.

#### *4.4.1.5 Data augmentation*

Since there are several limitations in annotated data for supervised segmentation deep network, including scarce annotations where only limited labeled data are available for training, as well as weak and noisy annotations where the training data have sparse and erroneous annotations. Especially, for vessel annotation in XCA images, the contrast-filled vessel networks are distributed and overlapped in the complex noisy backgrounds with similar appearances to the foreground vessels, which make it very difficult and time-consuming for manually labeling vessel networks (especially with distal vessels) from backgrounds. To improve the network's generalization and accuracy based on small labeled datasets, data augmentation techniques increasing both the size and diversity of labeled training datasets are

required to generate new labeled data points for training via transformations of existing image.

SVS-net used various sample transformations to augment data from existing data, which include rotations by the angles ranging in $[-10^{\circ}, 10^{\circ}]$, flipping both horizontally and vertically, scaling by a factor of 0.2 random crop, and affine transformations. For the images in our dataset, there is a 50% probability to perform each of the above transformations to generate new samples in real-time during the training process.

#### 4.4.1.6 Loss function for class imbalance problem

Class imbalance is a common problem in which a dataset consists of examples primarily from one specific class. This could manifest itself in XCA vessel segmentation problems such that there is a clear majority−minority class distinction between the backgrounds and foreground vessels due to the contrast-filled vessels being sparsely visualized in the spatio-temporal sequence. Specifically, the class imbalance problem is typically represented as two aspects in XCA vessel segmentation: first, the number of negative pixels (being 0 for background) is much greater than the number of positive pixels (being 1 for the vessel pixels); second, the ratio between the two classes varies a lot in terms of both interframe differences in the same XCA sequence and intersequence differences in the different XCA sequences.

In supervised deep learning, imbalanced datasets are harmful because they bias deep learning models toward majority class predictions and render accuracy as a deceitful performance metric. Many data-level and algorithm-level solutions [78] are proposed to deal with class imbalance in big data in deep learning. The above-mentioned data augmentation falls under a data-level solution to class imbalance.

Among the algorithm-level solutions, the design of loss function [79] is very important in the deep learning to evaluate how well the predicted segmentation matches the ground truth. Currently, there is no widely accepted common view about the best loss function for image segmentation. However, there are some insights from existing literature. For example, the Dice loss or generalized Dice loss has proven to perform well for the mildly imbalanced problems [79]. Being different from the cross entropy highlighting each pixel equally to the CE loss, the Dice loss function measures the overlap ratio between ground truth mask and the predicted vessel mask such that it is appropriate to be adopted in SVS-net to guide parameters learning. Dice loss is defined as follows:

$$L_{\text{DiceCoef}} = -\frac{2\sum_{1}^{N} p_i \gamma_i + \epsilon}{\sum_{1}^{N} p_i + \sum_{1}^{N} \gamma_i + \epsilon} \tag{4.26}$$

where $\gamma_i \in \{0, 1\}$ is the ground truth label and $p_i \in [0, 1]$ is a predicted value for location $i$. $N$ is the total number of pixels, $\epsilon$ is a very small constant used to keep value stable.

From Eq. (4.26) we can find that the Dice loss is applied to the whole mask and it measures the overall loss for that mask rather than the average loss across all the pixels.

Recently, Kervadec et al. [80] have demonstrated that combining boundary loss with generalized Dice loss can address highly imbalanced segmentation tasks, we assume that not only using compound loss functions but also highlighting the boundary information with the overlap ratio between ground truth mask and the predicted vessel mask may be a more better choice for segmenting the dynamically and sparsely distributed vessel networks from the complex background.

## 4.4.2 Segmentation experimental results

### *4.4.2.1 Materials*

All the experiments performed in this work were approved by our institutional review board. Our experiments were implemented on GPU (i.e., NVIDIA 1080 Ti, 11GB) and acquired 120 sequences of real clinical X-ray coronary angiography images from Ren Ji Hospital of Shanghai Jiao Tong University. Each sequence has a length of 30−140 frames. Three experts manually labeled the images from 120 sequences to constitute the ground truth. To eliminate the differences in the resolution, the noise distribution, and the pixels' intensity range of different frames in these XCA sequences that are acquired from different machines (i.e., 800 mAh digital silhouette angiography X-ray machine from Siemens, medical angiography X-ray system from Philips), we resized the images from the XCA sequence into $512 \times 512$ resolution with 8 bits per pixel, and employed Poisson denoising methods to smooth the noise and normalize the pixels' intensity range into 0−1. The dataset is stored in mat array format according to the corresponding filenames and will be available on the website https://github.com/Binjie-Qin/SVS-net.

Due to the varieties of XCA images with different directions and angles of X-ray penetration as well as different patient sources with different dosages of contrast agents, the visibility of different vessel sequences in the clinic is quite changeable such that designing a robust vessel segmentation algorithm is necessary for the XCA data with this poor visual quality. In addition, proper selection of frames from each sequence for experiment is crucial especially when both the background and foreground are dynamic and contain many artifacts. From the 120 annotated sequences, we selected the XCA images that contain most vessel structures with good visual quality as a total of 332 experiment samples. The dataset is randomly divided into training dataset, validation dataset, and test data at approximately 0.5, 0.25, and 0.25 ratios, respectively.

We validated the SVS-net's performance on the above-mentioned dataset by plotting the loss curves for both the training and validation set in the training process [15]. For both the training and validation set, the loss reduced quickly at the beginning stage of training process and gradually converged, indicating the SVS-Net having no

problem of over-fitting or under-fitting state. Meanwhile, the size of our dataset was assumed to be properly matched into the size of our model.

#### 4.4.2.2 Performance comparison

We compared SVS-net with three traditional vessel segmentation algorithms, that is, Coye's method [81], Jin's spatially adaptively filtering method (Jin's) [11], Kerkeni's multiscale region growing method (Kerkeni's) [82], and four deep learning-based methods that include Retinal-net [67,83], bridge-style U-Net with salient mechanism (S-UNet) [84], X-ray net [85], and short connected deep supervised net (BTS-DSN) [86].

In Fig. 4.5, the deep learning methods surpassed traditional methods with higher performances in terms of some metrics [15]. The patch-based Retinal-net method introduced more background residuals since it lacked more global contextual information to guide the segmentation. X-ray net simply accepted the current frame image with its first three frames as the inputs to the network and cannot effectively extract the temporally consistent information. Therefore it not only increased temporal information but also introduced disturbances at the same time. BTS-DSN adopted deeply supervised strategy to achieve relative higher metrics with some false positives in the vessel regions. Being different from the above deep network methods, SVS-net not only robustly detected the vessels with entire structures and continuous branches but also effectively removed the noisy backgrounds. The continuity and integrity of the segmented vessel branches were due to the contextual information inferred in the spatiotemporal features extracted by the encoder network and feature fusion in the skip connections. The noise suppression in the segmented vessel regions was mostly derived from the

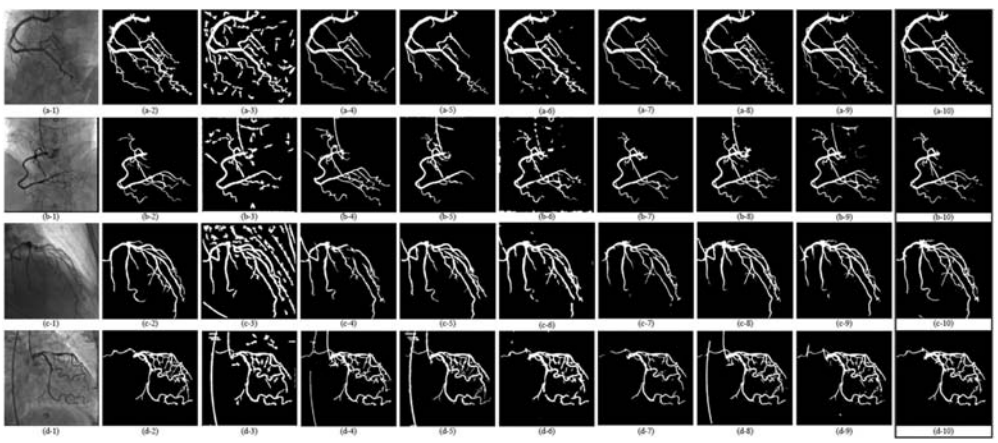

**Figure 4.5** Vessel segmentation examples for real XCA image sequence by different vessel segmentation methods [15]. From left to right, each row displays the original XCA image, the ground truth vessel segmentation, the vessel images segmented by Coye's, [71] Jin's, [11] Kerkeni's, [72] Retinal-net, [73] SU-Net, [74] X-ray net, [75] BTS-DSN, [76] and SVS-net, [15] respectively.

discriminative feature selection implemented by the channel attention mechanism. Therefore the spatiotemporal feature extraction, feature fusion, and the discriminative feature learning adopted in SVS-net were helpful to improve the segmentation performances.

However, there is a small number of thin-vessel branches unrecognized by SVS-net. A new design of a hybrid loss function combined with network architectures is desirable to efficiently differentiate the thick and thin vessels with their different weights. That is, the deep gives more weight to these thin vessels via topology-preserving loss function [87] and pays more attention to these thin vessels. Topology-preserving vessel segmentation is definitely a very promising direction for improving the clinical value of XCA images.

## 4.5 VRBC-t-TNN: accurate heterogeneous vessel extraction via tensor completion of X-ray coronary angiography backgrounds

By modeling XCA sequence with low-rank plus sparse decomposition, motion coherency regularized RPCA [5] has been successfully exploited to recover the sparsely distributed contrast agent in the vessel networks from the low-rank subspace of backgrounds [21]. However, current RPCA-based methods have the following limitations: (1) Some parts of XCA vessel have low-rank properties due to the underlying pattern of periodic heartbeat and the contrast agent's adhesion along the vessel wall, current RPCA methods therefore always have vessel residue as part of the low-rank backgrounds such that the extracted vessels suffer from severe distortion or loss of vessel intensity. This intensity loss results in an incomplete vessel recovery and makes it impossible for accurately analyzing the contrast agent concentration and the corresponding blood flow perfusion. (2) Vectorizing the XCA sequence into a matrix in RPCA ignores the 3D spatiotemporal information between the consecutive frames of the XCA sequence. For example, X-ray imaging produces a lot of dense noisy artifacts, whose positions change in gradually moving patterns in the XCA frames. The RPCA methods often mistakenly recognize these moving artifacts as foreground objects. (3) XCA sequence has serious signal-dependent noises that locally affect every entry of the data matrix and result in unsatisfying foreground vessels that are contaminated with residual artifacts. Though MCR-RPCA takes advantage of Bayesian RPCA that models data noise as MoGs, current RPCA-based methods cannot tackle the challenging problem of spatially varying noise in low-rank plus sparse decomposition.

To extract the foreground vessels that are overlapped with backgrounds, we assume that the overlap between foreground vessels and the backgrounds is present in the vessel regions. Once contrast-filled vessel regions are masked by the SVS-net vessel segmentation with high detection rates, all the other pixels in the remaining regions can be regarded as sequential background layers. Furthermore, based on the spatiotemporal consistency and low-rankness of the sequential background layers, the missing parts of

the background layers masked with the foreground vessels can be fully completed using the state-of-the-art tensor completion method. Then, the challenging problem of foreground vessel extraction can be tackled by subtracting the completed background layers from the overall XCA data (see Fig. 4.6). Therefore we exploited the low-rank tensor completion [88] to recover a low-rank tensor from noisy partial observations of its entries. Tensors [89,90] refer to multidimensional arrays, which can naturally reserve more spatiotemporal information than do the matrices. Different from RPCA using matrix representation for XCA sequence, the tensor-based completion methods infer the unknown missing pixels from the known pixels in the spatiotemporal contexts of XCA sequence. Mathematically, the low-rank tensor completion problem can be written as

$$\min_{\mathcal{X}} \text{rank}(\mathcal{X}), \quad \text{s.t. } P_{\Omega}(\mathcal{X}) = P_{\Omega}(\mathcal{M}) \tag{4.27}$$

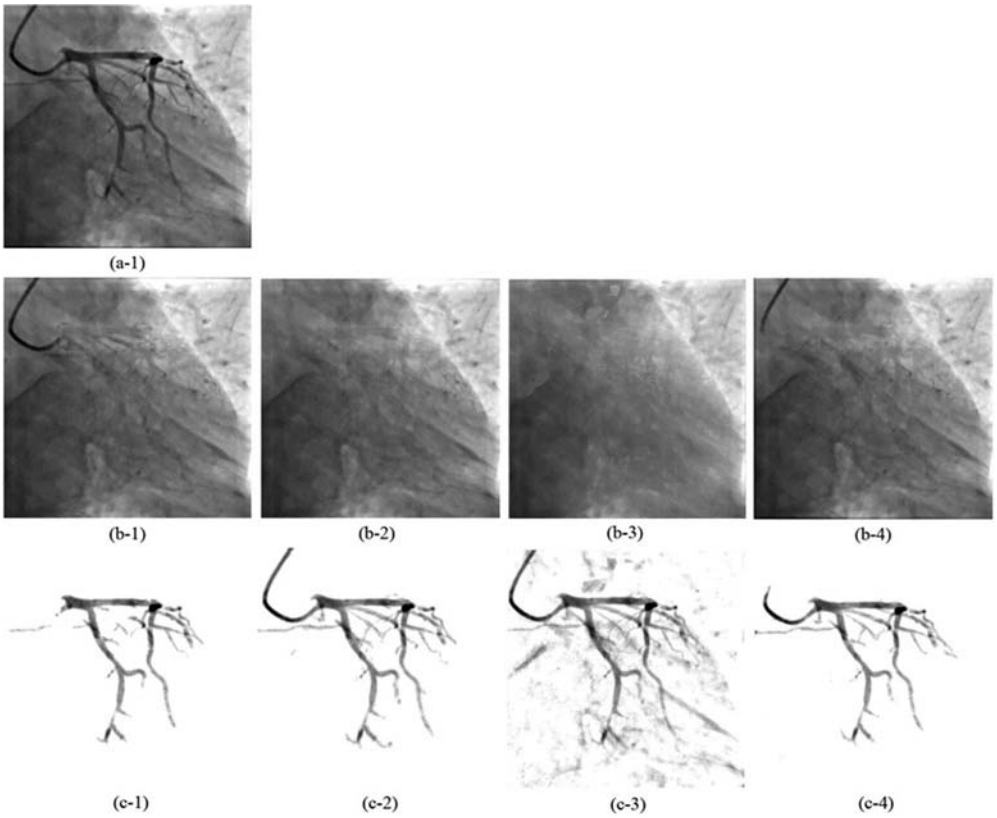


**Figure 4.6** Vessel intensity recovery [15]. From top to bottom row, each row displays the original XCA image, the background and foreground vessel images recovered from SU-Net [74], X-ray net [75], BTS-DSN [76] and SVS-net [15], respectively.

where $\mathcal{M}$ is the corrupted tensor, $\mathcal{X}$ is the underlying tensor, and $\mathcal{P}_{\Omega}(\mathcal{X})$ refers to the projection of $\mathcal{X}$ on the observed entries $\Omega$.

Based on different definitions of tensor ranks, many tensor completion models have been proposed. Typically, tensor nuclear norm (TNN) [91] that is designed for 3D tensors based on tensor singular value decomposition (t-SVD) [92] has proven effective at 3D tensor completion [88]. Hu et al. have further optimized the TNN model for the video completion task by integrating a twist operation [93].

According to Beer–Lambert Law, a given X-ray image is assumed to reflect the X-ray exponential attenuation composition (or sum) of material linear attenuation coefficients for the foreground contrast-filled vessels and background layers along the X-ray projection paths. The additive property of X-ray exponential attenuation composition along the vessel and background layers in XCA can be directly exploited to exactly decompose the whole XCA image into the vessel and background layers. Therefore this X-ray attenuation sum model justifies the above-mentioned foreground vessel extraction strategy via completion and subtraction of the background layers from the whole XCA images. Specifically, we proposed an accurate heterogeneous vessel extraction framework in a logarithmic domain, into which the raw XCA image is first mapped. We extracted the vessel mask regions via SVS-net and subsequently recovered vessel intensities in these regions by a tensor completion method called t-TNN (twist tensor nuclear norm) [93]. By focusing on the vessel intensity recovery problem only in the small parts of vessel regions, the proposed vessel recovery method called VRBC-t-TNN can extract vessel layers with accurate recovery of vessel structures and intensities.

### 4.5.1 Global intensity mapping

A global logarithm mapping was carried out on the whole XCA image data to perfectly fit the X-ray attenuation sum model of angiograms. In X-ray imaging, photons coming through human body are attenuated by contrast agents and various human tissues. The intensity of rays is reduced exponentially by the sum of attenuation coefficients, as the following equation:

$$X_{\text{out}} = X_{\text{in}} \mathrm{e}^{-\int_d \mu \mathrm{d}x} \tag{4.28}$$

where $X_{\text{in}}$ and $X_{\text{out}}$ represent the intensities of X-rays that come into and out of human body, respectively, $\mu$ denotes the attenuation coefficient, and $d$ denotes the path of rays.

By applying the log operator on both sides, we got

$$-\ln(X_{\text{out}}/X_{\text{in}}) = \int_d \mu \mathrm{d}x. \tag{4.29}$$

The XCA image intensity normalized to the range [0,1] can be regarded as the normalization of its intensity, that is, the ratio of $X_{\text{out}}$ to $X_{\text{in}}$. Then we got the following equation:

$$-\ln(I_{\text{XCA}}) = -\ln(X_{\text{out}}/X_{\text{in}}) = \int_{d_1} \mu \mathrm{d}x + \int_{d_2} \mu \mathrm{d}x = A_F + A_B, \tag{4.30}$$

where $A_F$ and $A_B$ represent the attenuation sums caused by foreground vessels and complex backgrounds, respectively. This equation demonstrates that the XCA image is a sum of vessel/background layers in the logarithm domain, accordingly the multiplication of the two layers in the original image domain.

After this logarithm mapping, the linear sum model was ready for vessel/background separation via low-rank plus sparse matrix decomposition in RPCA (Illustrated in Section 4.3), as well as low-rank background plus foreground vessel extraction in tensor completion (Illustrated in Section 4.6). Therefore we used the logarithm operation as a preprocessing of image data and performed exponentiation operation afterward for the whole experiments in this work.

### 4.5.2 Background completion using t-TNN

Considering the vessel mask regions as missing entries of the background layer, we recovered the intensities of these entries for the construction of background layers by adopting the t-TNN tensor completion algorithm [93], which can effectively exploit the temporal redundancy and low-rank prior between the neighboring frames more efficiently. Specifically, the original XCA sequence was represented as a tensor $\mathcal{D}$ with each slice being a matrix representation of each frame. All areas except the vessel mask regions, denoted as $\Omega$, were presumed to be the known background layer pixels. Each vessel mask region was first morphologically dilated by a $5 \times 5$ mask to ensure that $\Omega$ does not contain edge pixels of the vessels. Therefore the background regions of all the frames constituted $\Omega$ for the subsequent t-TNN-based tensor completion.

The t-TNN model is based upon a tensor decomposition scheme called t-SVD [92,94]. Having a similar structure to the matrix SVD, t-SVD models a tensor in the matrix space through a defined t-product operation [92]. For a 3-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the notation $\mathcal{X}(k,:,:)$, $\mathcal{X}(:,k,:)$, and $\mathcal{X}(:,:,k)$ denote the $k$-th horizontal, lateral, and frontal slices, respectively. Particularly, $\mathcal{X}^{(k)}$ denotes $\mathcal{X}(:,:,k)$. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the $\mathcal{X}^{(k)}$ values can be used to form the block circulant matrix

$$\mathbf{bcirc}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \cdots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \cdots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \cdots & \mathcal{X}^{(1)} \end{bmatrix}, \tag{4.31}$$

The block vectorizing and its opposite operation are defined as

$$\mathbf{bvec}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix}, \mathbf{bvfold}(\mathbf{bvec}(\mathcal{X})) = \mathcal{X} \tag{4.32}$$

The **bdiag** operator which maps the tensor to the block diagonal matrix and its opposite operation **bdfold** are defined as

$$\mathbf{bdiag}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & & \\ & \ddots & \\ & & \mathcal{X}^{(n_3)} \end{bmatrix}, \mathbf{bdfold}(\mathbf{bdiag}(\mathcal{X})) = \mathcal{X}. \tag{4.33}$$

Based on these notations, the t-product is defined as follows [92]:

$$\mathcal{M} = \mathcal{A} * \mathcal{B} = \mathbf{bvfold}(\mathbf{bcirc}(\mathcal{A})\mathbf{bvec}(\mathcal{B})).$$

The t-product is analogous to the matrix product except that the circular convolution replaces the product operation between the elements. The following equation shows that original t-product corresponds to the matrix multiplication of the frontal slices in the Fourier domain:

$$\mathcal{M}_f^{(k)} = \mathcal{A}_f^{(k)} * \mathcal{B}_f^{(k)}, \quad k = 1, \cdots, n_3. \tag{4.34}$$

where $\mathcal{X}_f = \mathbf{fft}(\mathcal{X}, [\,], 3)$ denotes the Fourier transform of $\mathcal{X}$ along the third dimension. Accordingly, $\mathcal{X} = \mathbf{ifft}(\mathcal{X}_f, [], 3)$. Based on the definition of t-product, the tensor-SVD (t-SVD) [94] of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \tag{4.35}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, respectively. $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rectangular f-diagonal tensor in which all of its frontal slices are diagonal matrices, and the entries in $\mathcal{S}$ are called the singular values of $\mathcal{X}$.

Based on the Fourier domain property of t-product as Eq. (4.34), t-SVD can be efficiently computed in the Fourier domain [92,94]. Each frontal slice of $\mathcal{U}_f$, $\mathcal{S}_f$, and $\mathcal{V}_f$ can be obtained via the matrix SVD, that is, $[\mathcal{U}_f^{(k)}, \mathcal{S}_f^{(k)}, \mathcal{V}_f^{(k)}] = \mathbf{SVD}(\mathcal{X}_f^{(k)})$. Then the t-SVD of $\mathcal{X}$ can be obtained by $\mathcal{U} = \mathbf{ifft}(\mathcal{U}_f, [], 3), \mathcal{S} = \mathbf{ifft}(\mathcal{S}_f, [], 3), \mathcal{V} = \mathbf{ifft}(\mathcal{V}_f, [], 3)$. The TNN of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the average of the nuclear norms of all the frontal slices of $\mathcal{X}_f$ [94], that is,

$$||\mathcal{X}||_{\circledast} = \frac{1}{n_3} \sum_{i=1}^{n_3} ||\mathcal{X}_f^{(i)}||_* = \frac{1}{n_3} ||\mathbf{bcirc}(\mathcal{X})||_* \tag{4.36}$$

Using the definition of TNN, the TNN-based tensor completion [93] can be represented by

$$\min_{\mathcal{X}} ||\mathcal{X}||_{\circledast}, \quad \text{s.t. } P_{\Omega}(\mathcal{X}) = P_{\Omega}(\mathcal{M}), \tag{4.37}$$

where $\mathcal{M}$ is the corrupted tensor, $\mathcal{P}_{\Omega}(\mathcal{X})$ refers to the projection of $\mathcal{X}$ on the observed entries $\Omega$. Accordingly, $\mathcal{P}_{\Omega^{\perp}}(\mathcal{X})$ is the complementary projection, that is, $\mathcal{P}_{\Omega}(\mathcal{X}) + \mathcal{P}_{\Omega^{\perp}}(\mathcal{X}) = \mathcal{X}$. The tensor completion problem can be solved using the t-SVD mentioned above.

TNN can simultaneously characterize the low-rankness of a tensor along various modes and therefore is a general model for 3D data completion problems. Based on the TNN, Hu et al. [93] proposed twist TNN (called t-TNN) via twist operator on the tensor. That is, for $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the twist tensor $\vec{\mathcal{X}}$ is an $n_1 \times n_2 \times n_3$ tensor whose lateral slice $\vec{\mathcal{X}}(:, k, :) = \mathcal{X}(:, :, k)$. Though this twist operation is simply a dimension shift of tensors, it emphasizes the temporal connections between frames. The t-TNN norm $\|\mathcal{X}\|_{\vec{\circledast}}$ of tensor $\mathcal{X}$ is then defined as follows:

$$\|\mathcal{X}\|_{\vec{\circledast}} = \frac{1}{n_3} \|\mathbf{bcirc}(\vec{\mathcal{X}})\|_{*}, \tag{4.38}$$

where the twist operation $\vec{\mathcal{X}}$ is a dimension shift of $\mathcal{X}$, and $\overleftarrow{\mathcal{Y}}$ shifts it back. By equalizing the nuclear norm of the block circulant matriculation of the twist tensor, t-TNN can not only exploit the correlations between all the modes simultaneously but also take advantage of the low-rank prior along a certain mode, for example, X-ray image sequence over the time dimension. Specifically, by exploiting the low-rank prior along the horizontal translation relationship between frames in XCA image sequences due to patient's breath and movement, we have found that the t-TNN model is more suitable than TNN for our XCA background completion work. Therefore by minimizing the t-TNN norm-based rank of the input tensor subject to certain constraints, the tensor completion work for XCA background recovery can be addressed by solving the following convex model [93]:

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\vec{\circledast}}, \quad \text{s.t. } \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{D}) \tag{4.39}$$

where $\mathcal{D}$ and $\mathcal{X}$ refer to the original corrupted data tensor (original XCA sequence) and the reconstructed tensor (background layer), respectively.

The alternating direction method of multipliers algorithm [95] can be used to solve the Eq. (4.39) by the following minimization model [93]:

$$\operatorname{argmin}_{\mathcal{X},\mathcal{Y},\mathcal{W}} \|\mathcal{Y}\|_{\vec{\circledast}} + \mathbf{1}_{\mathcal{X}_{\Omega}=\mathcal{D}_{\Omega}} + \langle \mathcal{W}, \mathcal{X} - \mathcal{Y} \rangle + \frac{\rho}{2} \|\mathcal{X} - \mathcal{Y}\|_{F}^{2}, \tag{4.40}$$

where a new variable $\mathcal{Y} = \mathcal{X}$ is introduced, $\mathbf{1}_{\mathcal{X}_{\Omega}=\mathcal{D}_{\Omega}}$ is an indicator function that indicates whether the elements of $\mathcal{X}$ and $\mathcal{D}$ on the support of $\Omega$ are equal, $\mathcal{W}$ is the Lagrangian

multiplier, and $\mu$ is a positive penalty scalar. Variables $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{W}$ can be optimized alternately with the other variable fixed. The detailed deduction of this algorithm is too long and out of the scope of this chapter. We refer the interested reader to the work in Ref. [93].

After the construction of background layer $\mathcal{X}$ by t-TNN, the final vessel layer can be obtained by subtracting $\mathcal{X}$ from the original data $\mathcal{D}$. This subtraction was done in the logarithm domain such that the corresponding operation for original image data would be division. The whole procedure of the t-TNN background completion step is shown in Algorithm 2.

**Algorithm 2** t-TNN based background completion

**Input:** Original XCA data $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, non-vessel mask region $\Omega$.

1: **Initialize:** $\rho^0 > 0, \eta > 1, k = 0, \mathcal{X} = \mathcal{P}_\Omega(\mathcal{D}), \mathcal{Y} = \mathcal{W} = 0$.
2: **while** $\|\mathcal{X} - \mathcal{Y}\|_F / \|\mathcal{X}\|_F > tol$ **and** $k < K$ **do**
3: $\mathcal{X}$ **sub-problem:**
$\mathcal{X}^{k+1} = \arg\min_{\mathcal{X}: \mathcal{X}_\Omega = \mathcal{D}_\Omega^k} \left\| \mathcal{X} - \mathcal{Y}^k + \frac{1}{\rho^k}\mathcal{W}^k \right\|_F^2$
solved by:
$\mathcal{X}^{k+1} = \mathcal{P}_\Omega(\mathcal{D}^k) + \mathcal{P}_{\Omega^\perp}(\mathcal{Y}^k - \frac{1}{\rho^k}\mathcal{W}^k)$;
4: $\mathcal{Y}$ **sub-problem:**
$\mathcal{Y}^{k+1} = \arg\min_{\mathcal{Y}} \|\mathcal{Y}\|_{\overrightarrow{\circledast}} + \frac{\rho^k}{2} \left\| \mathcal{X}^{k+1} - \mathcal{Y} + \frac{1}{\rho^k}\mathcal{W}^k \right\|_F^2$
solved by:
$\tau = \frac{1}{\rho^k}, \tau' = \tau \left\lceil \frac{n_2+1}{2} \right\rceil, \mathcal{Z} = \mathcal{X}^{k+1} + \tau\mathcal{W}^k$;
$\overrightarrow{\mathcal{Z}}_f = \mathbf{fft}(\overrightarrow{\mathcal{Z}}, [], 3)$;
**for** $j = 1, \cdots, \left\lceil \frac{n_2+1}{2} \right\rceil$ **do**
$[\mathcal{U}_f^{(j)}, \mathcal{S}_f^{(j)}, \mathcal{V}_f^{(j)}] = \mathbf{SVD}(\overrightarrow{\mathcal{Z}}_f^{(j)})$;
$\mathcal{J}_f^{(j)} = \mathbf{diag}\{(1 - \frac{\tau'}{\mathcal{S}_f^{(j)}(i,i)})_+\}$;
$\mathcal{S}_{f,\tau'}^{(j)} = \mathcal{S}_f^{(j)} \mathcal{J}_f^{(j)}$;
$\mathcal{H}_f^{(j)} = \mathcal{U}_f^{(j)} \mathcal{S}_{f,\tau'}^{(j)} \mathcal{V}_f^{(j)^T}$
**end for**
**for** $j = \left\lceil \frac{n_2+1}{2} \right\rceil + 1, \cdots, n_2$ **do**
$\mathcal{H}_f^{(j)} = \mathrm{conj}\left(\mathcal{H}_f^{(n_2-j+2)}\right)$;
**end for**
$\mathcal{H} = \mathbf{ifft}(\mathcal{H}_f, [], 3), \mathcal{Y}^{k+1} = \overleftarrow{\mathcal{H}}$;
5: $\mathcal{W}^{k+1} = \mathcal{W}^k + \rho^k(\mathcal{X}^{k+1} - \mathcal{Y}^{k+1})$;
6: $\rho^{k+1} = \eta\rho^k, k = k + 1$
7: **end while**
8: Vessel layer $\mathcal{V} = \mathcal{D} - \mathcal{X}$;

**Output:** Background layer tensor $\mathcal{X}$, vessel layer tensor $\mathcal{V}$.

## 4.5.3 Experimental results

### 4.5.3.1 Synthetic X-ray coronary angiography data

To accurately evaluate the performance of vessel intensity recovery, we constructed 10 synthetic XCA images which include ground truth background images (GTBL) and foreground vessel layer images (GTVL). All the original 12 XCA sequences were obtained from Ren Ji Hospital of Shanghai Jiao Tong University. Each sequence has 80 frames whose image resolution is $512 \times 512$ pixels with 8 bits per pixel. All the experiments were approved by our institutional review board.

To get GTVLs, we performed a vessel extraction algorithm similar to Section 4.2 on the real XCA data. Then some artifacts were manually removed from the extracted rough vessel images to obtain the GTVLs. The GTBLs were the consecutive frames selected from the real XCA data. Because an XCA image is the product of the vessel layer and the background layer according to the X-ray imaging mechanism (see Section 4.5.1), we multiplied a sequence of GTVLs to the clean regions of GTBLs from a different sequence to obtain the synthetic XCA data. An example synthetic image with GTBL and GTVL is shown in Fig. 4.7A.

### 4.5.3.2 Experiment demonstration

We performed experiments similar to the algorithm shown in Fig. 4.1 to illustrate the effect of tensor completion on vessel intensity recovery. The real vessel segmentation algorithm (see details in Ref. [11]) is different from that of Section 4.4 in Fig. 4.1, it makes little influence on vessel intensity performances. We validated the performances from the VRBC-t-TNN and other layer separation methods for comparison purpose. The median subtraction method (MedSubtract) used by Baka et al. constructs a static background layer image as the median of the first 10 frames of a sequence and subtract it from all the frames [96]. Several open-source RPCA algorithms, including PRMF [97], MoG-RPCA [51], IALM-RPCA, [5], and MCR-RPCA in Section 4.3 [5] were tested. The proposed framework VRBC can use other matrix completion and tensor completion methods to replace t-TNN. We tested some open source data completion methods including PG-RMC [98], MC-NMF [99], ScGrassMC [100], LRTC [101], and tSVD [102] as comparison, whose codes were obtained from Sobral's library lrslibrary [103] and mctc4bmi [104].

### 4.5.3.3 Performance comparison

The layer separation results of the synthetic data are shown in Fig. 4.7. As can be seen, all these layer separation methods can remove noises and increase the vessel visibility to some extent. Among these algorithms, because MedSubtracted constructs a static background layer image that does not change over time, the extracted vessel layer image with lots of remaining noise was the worst among these extracted results. The four RPCA-based methods achieved much better vessel extraction results with more

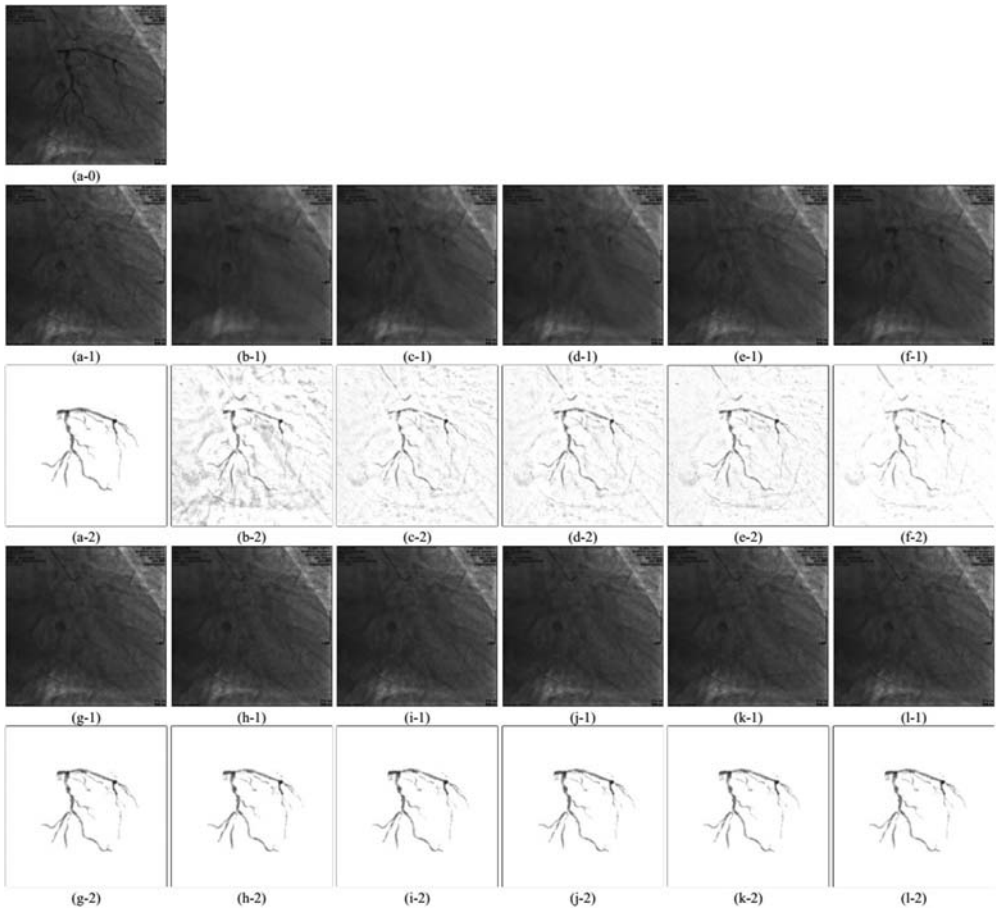


**Figure 4.7** Examples of vessel layer extraction results from synthetic data [21]. Each group of results contains a background layer image labeled 1 and a vessel layer image labeled 2. (a-0) Synthetic XCA image. (a-1,2) Ground truth background layer and vessel layer image. (b)–(l) Layer separation results: (b ) Med Subtract [83]. (c) PRMF [84]. (d) MoG-RPCA [49]. (e) IALM-RPCA [51]. (f) MCR-RPCA [5]. (g) VRBC-PG-RMC. (h) VRBC-MC-NMF. (i) VRBC-ScGrassMC. (j) VRBC-LRTC [21]. (k) VRBC-tSVD. (l) VRBC-t-TNN [21].

noises being removed. Among these four RPCA-based methods, MCR-RPCA [5] achieved the best vessel extraction results with the least residual artifacts. However, though RPCA-based methods can nicely capture the vessel structures in the vessel layer images, the vessel intensities were not fully extracted since obvious vessel residuals can be observed in their resulting background layer images. This observation is consistent with the fact that some contrast agents being pushed against the vessel wall have reduced their flow rate and behaved low-rankness property, such that these contrast agents remained in the backgrounds after RPCA-based decomposition.

In contrast, the VRBC framework embedded within all the tensor-completion methods greatly improved the performances of vessel extraction and intensity recovery. The recovered vessel intensities were further compared to those from RPCA algorithms. Among all these algorithms, the result images of VRBC-t-TNN achieved the best visual performances. Both the background layer images and the vessel layer images were visually appealing and seemed to be well recovered in terms of structure and intensity recovery. To quantitatively measure the accuracy of vessel intensity recovery, we directly calculated the differences between the extracted vessel layers and the ground truths. The reconstruction error of vessels is defined as follow:

$$E_{\text{recon}} = \frac{\sum_{(x,y)\in V} |I_{\text{result}}(x,y) - I_{\text{groundtruth}}(x,y))|}{\sum_{(x,y)\in V} I_{\text{groundtruth}}(x,y)} \tag{4.41}$$

where $V$ denotes the vessel regions, $I_{\text{result}}$ and $I_{\text{groundtruth}}$ denote the intensities of the resulting vessel layer images and the ground truth vessel layer images, respectively. For each synthetic XCA sequence, the $E_{\text{recon}}$ of the whole sequence is calculated. Fig. 4.8 shows the general performances of different algorithms on the 10 synthetic sequences.

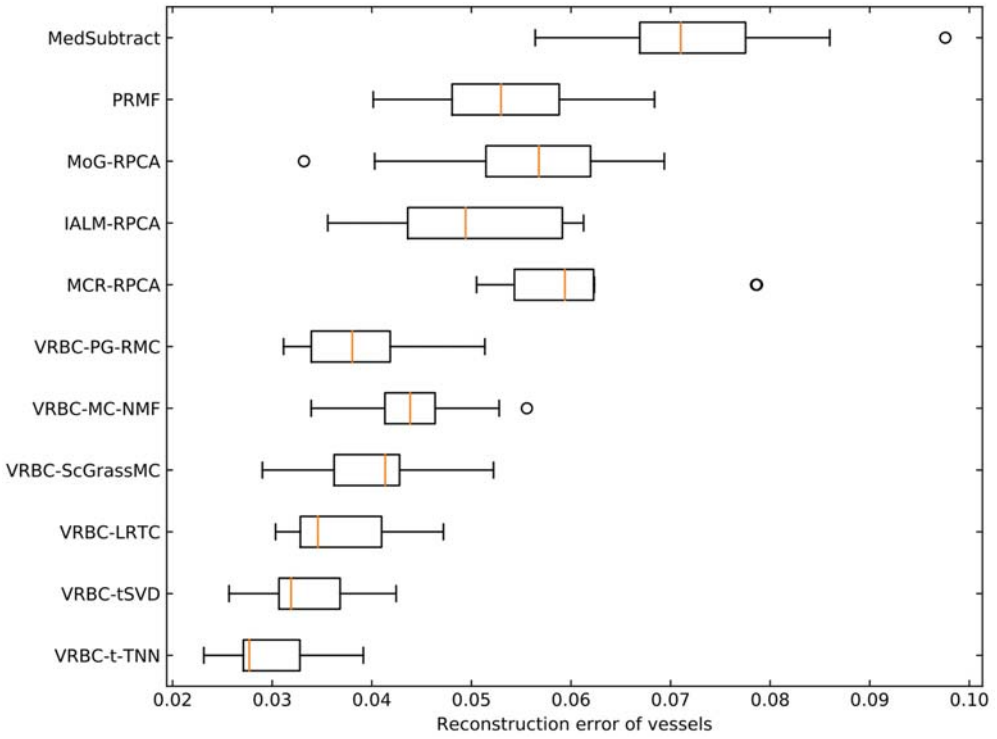


**Figure 4.8** $E_{\text{recon}}$ values in vessel regions on synthetic data [21].

$E_{\text{recon}}$ measured the vessel intensity difference between the separation result and the ground truth. A small $E_{\text{recon}}$ indicated an accurate vessel layer extraction. We can see that VRBC framework achieved smaller $E_{\text{recon}}$ values than other existing methods. Among them, VRBC-t-TNN achieved the best performance. This $E_{\text{recon}}$ evaluation indicated that VRBC-t-TNN can accurately recover the intensities of contrast-filled vessels from XCA images.

## 4.6 Conclusion

In this chapter, we demonstrated the entire framework to effectively extract the contrast-filled vessels and their intensities from the complex and noisy backgrounds in XCA sequences. To the best of our knowledge, this is the first work to perform such complete work for the benefit of microcirculation analysis in minimally invasive vascular interventions. Current XCA-guided coronary artery revascularization via stent implantation solely performs epicardial artery-based revascularization. However, whether this artery revascularization could improve the distal microcirculation and accordingly lead to a desired therapeutic effect on the acute myocardial ischemic heart diseases is not easily evaluated in preo perative and postoperative assessment in current vascular interventions. We hope our work could be helpful to appropriately assess the microcirculation and other cardiovascular diseases by extracting and analyzing the heterogeneous XCA vessels.

## Acknowledgments

This work was supported in parts by the Science and Technology Commission of Shanghai Municipality under Grant 19dz1200500 and Grant 19411951507, the National Natural Science Foundation of China under Grant 61271320 and 82070477, Shanghai ShenKang Hospital Development Center (SHDC12019X12), and the Shanghai Jiao Tong University Cross Research Fund for Translational Medicine under Grant ZH2018ZDA19, YG2021QN122 and YG2021QN99.

## References

[1] Theodoridis S. Machine learning: a Bayesian and optimization perspective. 2nd ed. Elsevier; 2020.
[2] Lin Z, Zhang H. Low-rank models in visual analysis: theories, algorithms, and applications. Academic Press; 2017.
[3] Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? J ACM 2011;58(3):1–37.
[4] Bouwmans T, Javed S, Zhang H, Lin Z, Otazo R. On the applications of robust PCA in image and video processing. Proc IEEE 2018;106(8):1427–57.
[5] Jin M, Li R, Jiang J, Qin B. Extracting contrast-filled vessels in X-ray angiography by graduated RPCA with motion coherency constraint. Pattern Recognit 2017;63:653–66.
[6] Ma H, Hoogendoorn A, Regar E, Niessen WJ, van Walsum T. Automatic online layer separation for vessel enhancement in X-ray angiograms for percutaneous coronary interventions. Med Image Anal 2017;39:145–61.

[7] Song S, Du C, Ai D, Huang Y, Song H, Wang Y, et al. Spatio-temporal constrained online layer separation for vascular enhancement in X-ray angiographic image sequence. IEEE Trans Circuits Syst Video Technol 2020;30(10):3558–70.
[8] Xia S, Zhu H, Liu X, Gong M, Huang X, Xu L, et al. Vessel segmentation of X-ray coronary angiographic image sequence. IEEE Trans Biomed Eng 2020;67(5):1338–48.
[9] Zhu L, Jiang X, Li J, Hao Y, Tian Y. Motion-aware structured matrix factorization for foreground detection in complex scenes. ACM Trans Multimedia Comput Commun Appl (TOMM) 2020;16(4):1–23.
[10] Hu Z, Nie F, Wang R, Li X. Low rank regularization: a review. Neural Netw 2021;136:218–32.
[11] Jin M, Hao D, Ding S, Qin B. Low-rank and sparse decomposition with spatially adaptive filtering for sequential segmentation of 2D + t vessels. Phys Med Biol 2018;63(17):17LT01.
[12] Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. Front Cardiovasc Med 2020;7:25.
[13] Nasr-Esfahani E, Karimi N, Jafari MH, Soroushmehr SMR, Samavi S, Nallamothu BK, et al. Segmentation of vessels in angiograms using convolutional neural networks. Biomed Signal Process Control 2018;40:240–51.
[14] Shin SY, Lee S, Yun ID, Lee KM. Deep vessel segmentation by learning graphical connectivity. Med Image Anal 2019;58:101556.
[15] Hao D, Ding S, Qiu L, Lv Y, Fei B, Zhu Y, et al. Sequential vessel segmentation via deep channel attention network. Neural Netw 2020;128:172–87.
[16] Samuel PM, Veeramalai T. VSSC Net: vessel specific skip chain convolutional network for blood vessel segmentation. Comput Methods Prog Biomed 2021;198:105769.
[17] Zhu X, Cheng Z, Wang S, Chen X, Lu G. Coronary angiography image segmentation based on PSPNet. Comput Methods Prog Biomed 2021;200:105897.
[18] Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, et al. Multichannel fully convolutional network for coronary artery segmentation in X-ray angiograms. IEEE Access 2018;6:44635–43.
[19] Wang L, Liang D, Yin X, Qiu J, Yang Z, Xing J, et al. Coronary artery segmentation in angiographic videos utilizing spatial-temporal information. BMC Med Imaging 2020;20(1):1–10.
[20] Zhang J, Wang G, Xie H, Zhang S, Huang N, Zhang S, et al. Weakly supervised vessel segmentation in X-ray angiograms by self-paced learning from noisy labels with suggestive annotation. Neurocomputing 2020;417:114–27.
[21] Qin B, Jin M, Hao D, Lv Y, Liu Q, Zhu Y, et al. Accurate vessel extraction via tensor completion of background layer in X-ray coronary angiograms. Pattern Recognit 2019;87:38–54.
[22] Kirbas C, Quek F. A review of vessel extraction techniques and algorithms. ACM Comput Surv (CSUR) 2004;36(2):81–121.
[23] Lesage D, Angelini ED, Bloch I, Funka-Lea G. A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med Image Anal 2009;13(6):819–45.
[24] Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms-review of methods, datasets and evaluation metrics. Comput Methods Prog Biomed 2018;158:71–91.
[25] Zhao F, Chen Y, Hou Y, He X. Segmentation of blood vessels using rule-based and machine-learning-based methods: a review. Multimed Syst 2019;25(2):109–18.
[26] Mookiah MRK, Hogg S, MacGillivray TJ, Prathiba V, Pradeepa R, Mohan V, et al. A Review of machine learning methods for retinal blood vessel segmentation and artery/vein classification. Med Image Anal 2020;68:101905.
[27] Jia D, Zhuang X. Learning-based algorithms for vessel tracking: a review. Computerized Med Imaging Graph 2021;89:101840.
[28] Çimen S, Gooya A, Grass M, Frangi AF. Reconstruction of coronary arteries from X-ray angiography: a review. Med Image Anal 2016;32:46–68.
[29] Kawabe M, Ohnishi T, Nakano K, Kato H, Ooka Y, Haneishi H. Vascular roadmap generation by registration and blending of multiple enhanced X-ray angiograms. IEEE Access 2021;9:36356–67.
[30] Qin B, Shen Z, Zhou Z, Zhou J, Lv Y. Structure matching driven by joint-saliency-structure adaptive kernel regression. Appl Soft Comput 2016;46:851–67.
[31] Qin B, Shen Z, Fu Z, Zhou Z, Lv Y, Bao J. Joint-saliency structure adaptive kernel regression with adaptive-scale kernels for deformable registration of challenging images. IEEE Access 2017;6:330–43.

[32] Smets BMN, Portegies J, St-Onge E, Duits R. Total variation and mean curvature PDEs on the homogeneous space of positions and orientations. J Math Imaging Vis 2021;63:237–62.
[33] Wang C, Oda M, Hayashi Y, Yoshino Y, Yamamoto T, Frangi AF, et al. Tensor-cut: a tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation. Med Image Anal 2020;60:101623.
[34] Zhang W, Wang X, Chen J, You W. A new hybrid level set approach. IEEE Trans Image Process 2020;29:7032–44.
[35] Liu X, Hou F, Qin H, Hao A. Robust optimization-based coronary artery labeling from X-ray angiograms. IEEE J Biomed Health Inform 2016;20(6):1608–20.
[36] Chen D, Zhang J, Cohen LD. Minimal paths for tubular structure segmentation with coherence penalty and adaptive anisotropy. IEEE Trans Image Process 2019;28(3):1271–84.
[37] Fang H, Ai D, Cong W, Yang S, Zhu J, Huang Y, et al. Topology optimization using multiple-possibility fusion for vasculature extraction. IEEE Trans Circuits Syst Video Technol 2020; 30(2):442–56.
[38] Yang G, Lv T, Shen Y, Li S, Yang J, Chen Y, et al. Vessel structure extraction using constrained minimal path propagation. Artif Intell Med 2020;105:101846.
[39] Liu M, Chen W, Wang C, Peng H. A multiscale ray-shooting model for termination detection of tree-like structures in biomedical images. IEEE Trans Med Imaging 2020;38(8):1923–34.
[40] Huang Z, Zhang Y, Li Q, Zhang T, Sang N. Spatially adaptive denoising for X-ray cardiovascular angiogram images. Biomed Signal Process Control 2018;40:131–9.
[41] Hariharan SG, Strobel N, Kaethner C, et al. Data-driven estimation of noise variance stabilization parameters for low-dose x-ray images. Phys Med Biol 2020;65(22):225027.
[42] Fazlali HR, Karimi N, Soroushmehr SMR, Shirani S, Nallamothu BK, Ward KR, et al. Vessel segmentation and catheter detection in X-ray angiograms using superpixels. Med Biol Eng Comput 2018;56(9):1515–30.
[43] Lv H, Fu S, Zhang C, Liu X. Non-local weighted fuzzy energy-based active contour model with level set evolution starting with a constant function. IET Image Process 2019;13(7):1115–23.
[44] Mei K, Hu B, Fei B, Qin B. Phase asymmetry ultrasound despeckling with fractional anisotropic diffusion and total variation. IEEE Trans Image Process 2020;29:2845–59.
[45] Shukla AK, Pandey RK, Pachori RB. A fractional filter based efficient algorithm for retinal blood vessel segmentation. Biomed Signal Process Control 2020;59:101883.
[46] Kovesi P. Image features from phase congruency. Videre: J Comput Vis Research 1999;1(3):1–26.
[47] Zhao Y, Zheng Y, Liu Y, Zhao Y, Luo L, Yang S, et al. Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter. IEEE Trans Med Imaging 2018; 37(2):438–50.
[48] Reisenhofer R, King EJ. Edge, ridge, and blob detection with symmetric molecules. SIAM J Imaging Sci 2019;12(4):1585–626.
[49] Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med Image Anal 2020;63:101693.
[50] Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. IEEE Trans Pattern Anal Mach Intell 2021.
[51] Zhao Q, Meng D, Xu Z, Zuo W, Zhang L. Robust principal component analysis with complex noise[C]//International conference on machine learning. PMLR 2014;55–63.
[52] Bouwmans T, Sobral A, Javed S, Jung SK, Zahzah EH. Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset. Comput Sci Rev 2017;23:1–71.
[53] Lin Z, Chen M, Ma Y The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv:1009.5055*; 2010.
[54] Ren Z, Chia LT, Rajan D, Gao S Background subtraction via coherent trajectory decomposition. In: Proceedings of the twenty-first ACM international conference on multimedia; 2013, pp. 545–548.
[55] Goldstein T, Osher S. The split Bregman method for L1-regularized problems. SIAM J Imaging Sci 2009;2(2):323–43.

[56] Gao Z, Cheong LF, Wang YX. Block-sparse RPCA for salient motion detection. IEEE Trans Pattern Anal Mach Intel 2014;36(10):1975–87.
[57] Zhou X, Yang C, Yu W. Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans Pattern Anal Mach Intel 2012;35(3):597–610.
[58] Lin Z., Wei S. A block Lanczos with warm start technique for accelerating nuclear norm minimization algorithms. arXiv preprint arXiv:1012.0365, 2010.
[59] Rodriguez P., Wohlberg B. Fast principal component pursuit via alternating minimization. 2013 IEEE International Conference on Image Processing. IEEE, 2013;69-73.
[60] Zhou T., Tao D. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. Proceedings of the 28th International Conference on Machine Learning, ICML 2011.
[61] Becker SR, Candès EJ, Grant MC. Templates for convex cone problems with applications to sparse signal recovery. Math Program Comput 2011;3(3):165.
[62] Wang N., Yao T., Wang J., et al. A probabilistic approach to robust matrix factorization. European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012:126-139.
[63] Xian Z, Wang X, Yan S, Yang D, Chen J, Peng C. Main coronary vessel segmentation using deep learning in smart medical. Math Probl Eng 2020;2020:8858344.
[64] Yang S, Kweon J, Roh JH, Lee JH, Kang H, Park LJ, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. Sci Rep 2019;9(1):1–11.
[65] Jun TJ, Kweon J, Kim YH, Kim D. T-net: nested encoder–decoder architecture for the main vessel segmentation in coronary angiography. Neural Netw 2020;128:216–33.
[66] Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J. A survey on U-shaped networks in medical image segmentations. Neurocomputing 2020;409:244–58.
[67] Ronneberger O, Fischer P, Brox T U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the international conference on medical image computing and computer-assisted intervention. Cham: Springer; 2015, pp. 234–241.
[68] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016, pp. 770–778.
[69] Zagoruyko S, Komodakis N Wide residual networks. *arXiv:1605.07146*; 2016.
[70] Wang W, Lai Q, Fu H, et al. Salient object detection in the deep learning era: an in-depth survey. IEEE Trans Pattern Anal Mach Intell 2021. Available from: https://doi.org/10.1109/TPAMI.2021.3051099.
[71] Liu GH, Yang JY, Li ZY. Content-based image retrieval using computational visual attention model. Pattern Recognit 2015;48(8):2554–66.
[72] Qin B, Gu Z, Sun X, Lv Y. Registration of images with outliers using joint saliency map. IEEE Signal Process Lett 2010;17(1):91–4.
[73] Gu Z, Qin B. Nonrigid registration of brain tumor resection MR images based on joint saliency map and keypoint clustering. Sensors 2009;9(12):10270–90.
[74] Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. Med Image Anal 2011;15(4):622–39.
[75] Li H, Ngan KN. A co-saliency model of image pairs. IEEE Trans Image Process 2011;20(12):3365–75.
[76] Gong X, Liu X, Li Y, Li H. A novel co-attention computation block for deep learning based image co-segmentation. Image Vis Comput 2020;101:103973.
[77] Wang W, Shen J, Yang R, Porikli F. Saliency-aware video object segmentation. IEEE Trans Pattern Anal Mach Intell 2018;40(1):20–33.
[78] Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. J Big Data 2019;6(1):1–54.
[79] Ma J, Chen J, Ng M, Huang R, Li Y, Li C, et al. Loss Odyssey in medical image segmentation. Med Image Anal 2021;71:102035.
[80] Kervadec H, Bouchtiba J, Desrosiers C, et al. Boundary loss for highly unbalanced segmentation. Med Image Anal 2021;67:101851.
[81] Coye T A novel retinal blood vessel segmentation algorithm for fundus images. MATLAB Cent File Exch (Jan 2017); 2017.

[82] Kerkeni A, Benabdallah A, Manzanera A, Bedoui MH. A coronary artery segmentation method based on multiscale analysis and region growing. Computerized Med Imaging Graph 2016;48:49–61.
[83] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. IEEE Trans Med Imaging 2016;35(11):2369–80.
[84] Hu J, Wang H, Gao S, Bao M, Liu T, Wang Y, et al. S-unet: a bridge-style u-net framework with a saliency mechanism for retinal vessel segmentation. IEEE Access 2019;7:174167–77.
[85] Ambrosini P, Ruijters D, Niessen WJ, Moelker A, van Walsum T Fully automatic and real-time catheter segmentation in X-ray fluoroscopy. In: Proceedings of the international conference on medical image computing and computer-assisted intervention. Cham: Springer; 2017, pp. 577–585.
[86] Guo S, Wang K, Kang H, Zhang Y, Gao Y, Li T. BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. Int J Med Inform 2019;126:105–13.
[87] Clough J.R., Byrne N., Oksuz I., et al. A topological loss function for deep-learning based image segmentation using persistent homology. IEEE Trans Pattern Anal Mach Intel 2021. Preprint.
[88] Song Q, Ge H, Caverlee J, Hu X. Tensor completion algorithms in big data analytics. ACM Trans Knowl Discovery Data (TKDD) 2019;13(1):1–48.
[89] Kolda TG, Bader BW. Tensor decompositions and applications. SIAM review 2009;51 (3):455–500.
[90] Panagakis Y, Kossaifi J, Chrysos GG, et al. Tensor Methods in Computer Vision and Deep Learning. Proc IEEE 2021;109(5):863–90.
[91] Ely G, Aeron S, Hao N, Kilmer ME. 5D and 4D pre-stack seismic data completion using tensor nuclear norm (TNN). In: SEG technical program expanded abstracts 2013. Society of Exploration Geophysicists; 2013, pp. 3639–3644.
[92] Kilmer ME, Braman K, Hao N, Hoover RC. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM J Matrix Anal Appl 2013;34(1):148–72.
[93] Hu W, Tao D, Zhang W, Xie Y, Yang Y. The twist tensor nuclear norm for video completion. IEEE Trans Neural Netw Learn Syst 2017;28(12):2961–73.
[94] Lu C, Feng J, Chen Y, Liu W, Lin Z, Yan S. Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans Pattern Anal Mach Intell 2020;42(4):925–38.
[95] Boyd S, Parikh N, Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc.; 2011.
[96] Baka N, Metz CT, Schultz CJ, van Geuns RJ, Niessen WJ, van Walsum T. Oriented Gaussian mixture models for nonrigid 2D/3D coronary artery registration. IEEE Trans Med imaging 2014; 33(5):1023–34.
[97] Wang N, Yao T, Wang J, Yeung DY A probabilistic approach to robust matrix factorization. In: European conference on computer vision. Berlin, Heidelberg: Springer; 2012, pp. 126–139.
[98] Cherapanamjeri Y., Gupta K., Jain P. Nearly optimal robust matrix completion. In: International conference on machine learning. PMLR; 2017, pp. 797–805.
[99] Xu Y, Yin W, Wen Z, Zhang Y. An alternating direction algorithm for matrix completion with nonnegative factors. Front Math China 2012;7(2):365–84.
[100] Ngo TT, Saad Y Scaled gradients on grassmann manifolds for matrix completion. In: NIPS 2012; 2012, pp. 1412–1420.
[101] Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. IEEE Trans Pattern Anal Mach Intell 2013;35(1):208–20.
[102] Zhang Z, Ely G, Aeron S, Hao N, Kilmer M Novel methods for multilinear data completion and de-noising based on tensor-SVD. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014, pp. 3842–3849.
[103] Sobral A, Bouwmans T, Zahzah E. Lrslibrary: low-rank and sparse tools for background modeling and subtraction in videos. Robust low-rank and sparse matrix decomposition: applications in image and video processing. CRC Press, Taylor and Francis Group; 2016.
[104] Sobral A, hadi Zahzah E. Matrix and tensor completion algorithms for background model initialization: a comparative evaluation. Pattern Recognit Lett 2017;96:22–33.