Multi-Modality Cardiac Image Computing: A Survey

Lei Li, Wangbin Ding, Liqun Huang, Xiaohai Zhuang*, Vicente Grau*

a Department of Engineering Science, University of Oxford, Oxford, UK
b School of Data Science, Fudan University, Shanghai, China
c College of Physics and Information Engineering, Fuzhou University, Fuzhou, China

Abstract

Multi-modality cardiac imaging plays a key role in the management of patients with cardiovascular diseases. It allows a combination of complementary anatomical, morphological and functional information, increases diagnosis accuracy, and improves the efficacy of cardiovascular interventions and clinical outcomes. Fully-automated processing and quantitative analysis of multi-modality cardiac images could have a direct impact on clinical research and evidence-based patient management. However, these require overcoming significant challenges including inter-modality misalignment and finding optimal methods to integrate information from different modalities.

This paper aims to provide a comprehensive review of multi-modality imaging in cardiology, the computing methods, the validation strategies, the related clinical workflows and future perspectives. For the computing methodologies, we have a favored focus on the three tasks, i.e., registration, fusion and segmentation, which generally involve multi-modality imaging data, either combining information from different modalities or transferring information across modalities. The review highlights that multi-modality cardiac imaging data has the potential of wide applicability in the clinic, such as trans-aortic valve implantation guidance, myocardial viability assessment, and catheter ablation therapy and its patient selection. Nevertheless, many challenges remain unsolved, such as missing modality, combination of imaging and non-imaging data, and uniform analysis and representation of different modalities. There is also work to do in defining how the well-developed techniques fit in clinical workflows and how much additional and relevant information they introduce. These problems are likely to continue to be an active field of research and the questions to be answered in the future.

Keywords: Multi-modality imaging, Cardiac, Registration, Fusion, Segmentation, Review

1. Introduction

Cardiovascular diseases (CVDs) remain one of the leading causes of mortality worldwide. Non-invasive cardiac imaging techniques can provide anatomical and functional information on health and pathology, and have been developed and employed extensively for the diagnosis and treatment of CVDs. Several imaging modalities are used for cardiac assessment, including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), echocardiography (Echo), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) [Laczay et al. 2021]. As Fig. 1 shows, each imaging modality manifests particular information, and a single modality may provide insufficient and/or ambiguous information regarding the condition of the heart that is a nonrigid and dynamic structure. Combining the complementary information from multi-modality cardiac images can be beneficial in establishing a more accurate diagnosis or assisting clinicians in the treatment management of CVDs [Valsangiacomo Buechel and Mertens 2012; Olsen et al. 2016; Puyol-Antón et al. 2021]. However, automated cardiac image computing is particularly challenging on account of normal or pathologic deformation of the heart during the respiratory and cardiac cycle. Furthermore, there exist misalignment and intensity variations among different modalities, and the imaging phases, field-of-views (FOVs), and resolution of different acquisitions can vary significantly. Several research methods have been developed to process multi-modality cardiac images, including registration, fusion, and segmentation of multi-modality images. With multi-modality images, image registration aims to align the available measurements (shape, function, pathology, etc.) into the same anatomical reference [Zhuang 2019]. Image fusion intends to integrate several imaging modalities in an optimal way for visualisation or decision-making. Image segmentation attempts to extract target regions by combining the complementary information from several imaging modalities. The three tasks exhibit overlap as well as distinct challenges. Specifically, their common issues are related to various sources of motion (patient, respiratory, and cardiac cycle). Unique challenges for registration and fusion may arise from the lack of a gold standard, which disables most current supervised learning-based models. The challenges associated with segmentation could originate from...
specific target regions, which can have large shape variations (e.g. the left atrium (LA)), can be small and/or diffuse (such as fibrotic regions and scars), or thin compared to image resolution (such as LA wall and coronary arteries). The proper selection of fusion techniques is complicated, as they should be easily interpretable in order to obtain traceable results. Also, the algorithms must satisfy several requirements, such as robustness to different acquisition conditions or tolerable computation times in a real-time system.

1.1. Study inclusion and literature search

In this work, we aim to provide readers with a survey of the state-of-the-art in multi-modality cardiac image computing techniques and the related literature. Fig. 2 illustrates the review scope of this survey. Note that in the computer vision area the modality might refer to text, audio and images, but in this work we restrict modality to imaging techniques. We consider multi-modality cardiac computing to include studies that involve diverse modalities, either combining or transferring information between cardiac image modalities. In the deep learning-based framework, different modalities may be present during training or more critically at the test phase at the same time. Therefore, we also include cardiac studies aiming at image translation or cross-modality domain adaptation, although different modality images may only appear separately in the training and test stages. Note that works that only test their methods in different cardiac imaging modalities for evaluation will not be included in this review. As for the cardiac computing tasks, we only consider the three major tasks at the current state of the art, i.e., registration, fusion, and segmentation. Classification is also a primary aim, but current classification works in the literature mainly integrate features from different cardiac signals, such as the electrocardiogram (ECG) and phonocardiogram (PCG), rather than imaging techniques (Kalidas and Tamil, 2016; Li et al., 2021d). There exist several cardiac image reconstruction works with the assistance of other modalities, mainly for cardiac motion correction (Polycarpou et al., 2021; Sang et al., 2021), which...
Table 2: Summary of the related representative review works. ML: machine learning; CAD: coronary artery disease; WH: whole heart; LA: left atrium; RV: right ventricle; DL: deep learning; TMI: IEEE Transactions on Medical Imaging; MedIA: Medical Image Analysis; CRP: Cardiology Research and Practice; CVIU: Computer Vision and Image Understanding; JBR: Journal of Biomedical Research; JHE: Journal of Healthcare Engineering; ELJ: Egyptian Informatics Journal; MAGMA: Magnetic Resonance Materials in Physics, Biology and Medicine; FCM: Frontiers in Cardiovascular Medicine; IPADCD: Image Processing for Automated Diagnosis of Cardiac Diseases; EP: EP Eurosense; CET: Cardiovascular Engineering and Technology; IET: IET Image Processing; FIP: Frontiers in Physiology; CCRR: Current Cardiovascular Risk Reports; EJNMMI: European journal of nuclear medicine and molecular imaging; JT: Jurnal Teknologi; CMIG: Computerized Medical Imaging and Graphics; JNM: Journal of Nuclear Medicine; CCIR: Current cardiovascular imaging reports. Note that here “single modality” refers to the methods designed for a single modality but may be evaluated on several different modalities, separately.

| Source | Venue | Scope | Inclusion |
|--------|-------|-------|-----------|
| Makela et al. (2002) | TMI | cardiac image registration | partially includes multi-modality |
| Stolzka and Baum (2009) | JNM | multi-modality image registration | software based; partially includes heart |
| Tavakoli and Amini (2013) | FIP | cardiac image segmentation and registration | single modality; shape-based methods |
| Ming et al. (2015) | JT | cardiac US and CT registration | US and CT registration |
| El-Gamai et al. (2016) | CRP | medical image registration and fusion | partially includes multi-modality or heart |
| Khalil et al. (2018) | JNM | cardiac image registration | partially includes multi-modality |
| Rischpler et al. (2013) | IPADCD | multi-modality cardiac image fusion | clinical review: PET/ MRI fusion |
| Piccinelli and Garcia (2013) | JBR | cardiac image fusion | partially includes multi-modality |
| Chauhan et al. (2021) | MedIA | cardiac short-axis MRI segmentation | single modality |
| Petitjean and Dacher (2011) | JHE | WH segmentation | single modality; conventional methods |
| Zhang (2013) | MAGMA | cardiac MRI segmentation | single modality |
| Peng et al. (2016) | EP | fibrosis segmentation from LGE MRI | single modality; thresholding methods |
| Pontecorbo et al. (2017) | FCM | LA cavity segmentation from LGE MRI | single modality; DL-based methods |
| Jamart et al. (2020) | CET | WH and chamber segmentation | single-modality |
| Habjaj et al. (2020) | JNM | cardiac PET and MRI fusion | clinical review: PET/ MRI fusion |
| Chen et al. (2020a) | IPADCD | cardiac image fusion | partially includes multi-modality |
| Ammari et al. (2021) | MedIA | LA-related segmentation from LGE MRI | single modality; LA-related segmentation |
| Wu et al. (2021) | JBR | multi-modality cardiac image fusion | clinical review: focus on CAD |
| Jia and Zhang (2021) | MedIA | cardiac short-axis MRI segmentation | single modality |
| Kwan et al. (2021) | JHE | WH segmentation | single modality; conventional methods |
| Li et al. (2022) | CCRR | DL-based cardiac image segmentation | single modality; learning-based methods |

Figure 3: The distributions of reviewed papers of multi-modality cardiac image computing per year and task.

However is under-researched yet.

To ensure comprehensive coverage, we have screened publications mainly from the last 12 years (2010-2021) related to this topic. Our main sources of reference were Internet searches using engines, including Google Scholar, PubMed, IEEE-Xplore, Connected Papers, and Citeseer. To cover as many related works as possible, flexible search terms have been employed when using these search engines, as summarized in Table 1. Both peer-reviewed journal papers and conference papers were included here. In the way described above, we have collected a comprehensive library of 147 papers. Fig. 3 presents the distributions of papers in multi-modality cardiac image computing per year/task. Note that we picked the most detailed and representative ones for this review when we encountered several papers from the same authors about the same subject. Moreover, as multi-modality cardiac image fusion works are more likely to be driven by a clinical objective rather than technical development, we only select several representative studies here to cover as many modality combinations as possible. Also, the work of fusion or segmentation may involve a registration step, but we usually only regard this work as fusion or segmentation when counting the number of papers, since these are the final aims.

1.2. Existing survey from literature

Table 2 lists current review papers related to the topic, i.e., multi-modality cardiac image computing. One can see that the scopes of current review works are different from ours though with partial overlaps. For example, current review works on cardiac image registration only summarized conventional registration methods (Makela et al., 2002; Tavakoli and Amini, 2013), only revolved around specific modalities (Ming et al., 2015), or partially included multimodality image registration (El-Gamai et al., 2016; Khalil et al., 2018). Cardiac image fusion survey works mainly focused on a specific CVD (Piccinelli and Garcia, 2013) or intra-modality fusion (Chauhan et al., 2021). Though
there are several cardiac segmentation review works (Pettitjean and Dacher, 2011; Zhuang, 2013; Peng et al., 2016; Chen et al., 2020a), they only concentrated on single-modality or single-structure cardiac imaging. In contrast, we provide a comprehensive review of multi-modality cardiac image analysis, excluding single-modality works.

1.3. Structure of this review
The rest of the survey is organized as follows. Section 2 summarizes widely used imaging modalities for cardiology. Section 3 covers the well-developed computing algorithms. In Section 4, publicly available datasets and performance evaluation measurements are listed. Discussion of current clinical applications, challenges and future perspectives of multi-modality cardiac image computing is given in Section 5 followed by a conclusion in Section 6.

2. Cardiac imaging
Many imaging techniques are available for cardiac analysis, as presented in Fig. 1. In general, these techniques either focus on the anatomical (wall thickness, volume, coronary artery morphology, etc.) or the functional (wall motion patterns, cardiac perfusion, etc.) aspects. This section briefly reviews the most commonly used imaging techniques for clinical investigation of the heart.

2.1. X-ray (fluoroscopy)
X-ray imaging captures a projection of a 3D structure on a 2D plane, with poor soft-tissue contrast but providing high contrast images of devices (Housden et al., 2012). It has an important role in the preliminary assessment of CVDs. For example, increased heart size and the presence of increased pulmonary vascular markings or pleural effusions in X-ray may suggest pulmonary congestion secondary to heart failure. X-ray angiography can provide information about coronary arteries (CA) and their stenos is, and therefore is a standard technique for imaging the aorta and large vessels being imaged, where the measured attenuation can be employed to reconstruct the 3D heart structure. Moreover, contrast agents can be utilized in CT to enhance the 3D visualization of CA (Mylonas et al., 2014), the localization of the myocardial infarction (MyoI) (Bauer et al., 2010), the detection of LA appendage (LAA) thrombus in atrial fibrillation (AF) patients (Spagnolo et al., 2021), and the estimation of myocardial (Myo) blood flow (Alessio et al., 2019). CT angiography (CTA) can give complementary information (clear anatomical information) about the CA tree but without any functional information in cardiac imaging (Zhou et al., 2012).

2.2. Computed tomography
Cardiac CT generates cross-sectional images of the human body, with excellent temporal and spatial resolution. It represents the X-ray attenuation properties of the tissue being imaged, where the measured attenuation can be employed to reconstruct the 3D heart structure. Moreover, contrast agents can be utilized in CT to enhance the 3D visualization of CA, the localization of the myocardial infarction (MyoI), the detection of LA appendage (LAA) thrombus in atrial fibrillation (AF) patients, and the estimation of myocardial (Myo) blood flow. CT angiography (CTA) can give complementary information (clear anatomical information) about the CA tree but without any functional information in cardiac imaging.

2.3. Magnetic resonance imaging
Cardiac MRI is an imaging technique with high resolution as well as high tissue contrast for the assessment of heart chambers, cardiac valves, cardiac vessels, and surrounding structures such as the pericardium (Earls et al., 2002). It can also be used to study myocardial perfusion and viability, and permits the calculation of ventricular function parameters such as left ventricle (LV) ejection fraction, end diastolic/systolic volumes or LV mass. Compared to nuclear imaging (PET or SPECT), the dependence of MRI signal on regional hypoperfusion is minimal (Kang et al., 2012). Note that within MRI, various pulse sequences can attenuate different tissue features to identify anatomical and functional information, resulting in images of varying contrasts. For example, cine MRI can provide clear structure and global ventricular performance via quantification of chamber volumes and function; late gadolinium enhanced (LGE) MRI provides fibrosis information by visually enhancing fibrosis regions; and T2-weighted images can be used to identify myocardial edema caused by inflammation or acute ischemia (Zhuang, 2019). In clinical practice, myocardial analysis of LGE/T2-weighted MRI is typically combined with cine MRI considering the poor image quality of LGE/T2-weighted MRI (Kim et al., 2009). Similar to CTA, MR angiography is a standard technique for imaging the aorta and large vessels of CA (Mieres et al., 2007).

2.4. Echocardiography
Echo (or US of the heart) is a real-time cardiac imaging technique without irradiation, and its temporal resolution is excellent. Echo is the most widely used imaging technique in cardiology as it is relatively inexpensive, radiation-free and portable. Dynamic Echo is the gold standard modality to assess myocardial strain. The limitations of Echo include small FOV, poor signal-to-noise
ratio, artifacts, and high sensitivity to operators and patients. There are various echocardiographic modalities, such as transthoracic Echo (TTE), transoesophageal Echo (TEE), and intracardiac Echo (ICE) (Hascoët et al., 2016). TTE can be used to identify the great majority of cardiac causes of shock (de Bruijn et al., 2006), diagnose pulmonary embolism (Fields et al., 2017), and evaluate patients with suspected myocarditis (Goitein et al., 2009). TEE allows for high-quality color flow imaging, which can be utilized for visualizing the cardiac structures near the upper chambers and assisting device positioning and deployment during cardiac treatment procedures (Khalil et al., 2018). However, TEE is an invasive imaging technique, as it is acquired via a thin tube connected with an Echo probe through the mouth of patients into the esophagus. ICE is a catheter-based technique that presents imaging within the heart as well as aortic root images, with a larger view than TEE. Several research works have presented the superiority of 3D over 2D Echo alone in the catheterization laboratory (Martin et al., 2013; Hascoët et al., 2015). Unfortunately, 3D Echo is not widely available in clinical practice and is more expensive compared with 2D Echo.

2.5. Positron emission tomography

Cardiac PET is a nuclear medical imaging technique that can measure myocardial perfusion, metabolism and regional/absolute myocardial blood flow (Slomka et al., 2009; Angelidis et al., 2017). It is particularly suitable for obese patients and patients with multi-vessel diseases to confirm or exclude balanced myocardial ischemia (Dewey et al., 2020). Its spatial resolution (about 3-4 mm) is lower compared to cardiac MRI or CT (about 1 mm), so PET usually can not provide clear anatomical information about the heart. Therefore, CT is usually utilized to perform attenuation correction for better resolution of PET images.

2.6. Single photon emission computed tomography

SPECT is the most commonly employed imaging technique for clinical myocardial perfusion, whereas PET is the clinical gold standard for myocardial perfusion quantification. With new detectors, SPECT can quantify myocardial blood flow and is suitable for patients with a higher body mass index (BMI) (Dewey et al., 2020). SPECT can also be used to detect ischemia and viability in heart failure (Angelidis et al., 2017), assess the physiological relevance of coronary lesions and provide a high prognostic predictive value (Kaufmann and Di Carli, 2009). Moreover, compared with PET, SPECT can accurately assess myocardial viability and has a lower cost of equipment and less expensive radiotracers, but with lower resolution imaging prone to artifacts and attenuation (Partington et al., 2011).

3. Multi-modality cardiac image computing

With these cardiac imaging techniques, one could combine their advantages for cardiac analysis in the clinic, e.g. for cardiac pathology identification, cardiac surgery guidance, and cardiac disease diagnosis and prognosis, as presented in Fig. 4. To achieve this, multi-modality cardiac image computing is required, mainly including multimodality registration, fusion and segmentation. In this section, we summarize techniques for the three multi-modality cardiac image computing tasks.

3.1. Registration

Multi-modality image registration is a native way to propagate knowledge between modalities as well as a primary step in integrating information of multi-modality images. It can align images with respect to each other and typically involves several components, i.e., feature extraction (optional), geometrical transformation, similarity...
measure, and optimization. Accurate multi-modality image registration can offer additional crucial clinical relevant information, which may be unavailable in a single imaging modality. However, registration is an ill-posed problem with several challenges, such as anatomy variations, possible partial overlaps between images, and high variability of tissue appearance under different modalities that leads to the lack of robust similarity measures.

Cardiac image registration is particularly challenging due to respiration and mixed motions during the cardiac cycles. Furthermore, multi-modality cardiac images are normally inconsistent in the cardiac phases, slice thickness, dimension, and state (static or dynamic), which introduce additional challenges. For example, for multi-sequence MRI registration, T2 imaging generally only includes a few slices with larger thickness compared to bSSFP and LGE imaging. Echo/CT image registration frameworks usually consist of two major steps: temporal synchronization and spatial registration. Temporal synchronization allows the echocardiographic time-series data to be time-stamped to identify frames that are in a similar cardiac phase to the CT volume. Spatial registration aims at producing an interpolated cardiac CT image that matches the Echo image, which sometimes involves a challenging 2D-3D registration. Despite these challenges, there exist several promising solutions for multi-modality image registration. We have coarsely divided these methods into three categories, namely intensity based registration, structural/anatomical information based registration, and image synthesis/disentangle based registration. Table 3 provides a summary of the state-of-the-art cardiac multi-modality registration methods according to this classification in chronological order.

3.1.1. Intensity based registration

A straightforward approach is to employ modality invariant similarity metrics, such as mutual information (MI) and its variants such as normalized MI and its weighted version can further ensure an organ-specific intensity co-occurrence. However, such global intensity-based similarity measures often ignore the spatial information of the target, so they are unsuitable in some situations. For instance, the tissue presented in Echo is inherently characterized by a specific spatial distribution of speckles rather than a specific distribution of gray scales that exist in MRI and CT (Sandoval and Dillenseger, 2013). Entropy images satisfy only certain requirements of a relaxed version of the theoretical properties, but they are computationally fast and lead to good alignment, facilitating its use as a practical solution. Laplacian images capture the intrinsic structure of imaging modalities, with the preservation of optimal locality. The local phase image can be a characterization of features detected in a signal, such as edges and ridges in a signal, so it has been shown appropriately for registration involving Echo images. In contrast to local phase representations, PEMs only highlight the image structures that are relevant for image registration, as presented in Fig. 5. Figure 5: An example of multi-modality cardiac image registration. Here, a generic and modality-independent image representation, namely probabilistic edge map (PEM), was proposed for cardiac CT and US registration. The PEM, by only highlighting the registration of relevant image structures, i.e., LV and atrial boundaries, is a more robust approach than registering every voxel in the images. Image adapted from Oktay et al. (2015) with permission.

3.1.2. Structural/anatomical information based registration

Instead of directly employing image intensities, one could utilize structural information, which is supposed to be independent to modalities. In this way, the influence of appearance differences among various modalities can be mitigated, and a simple L1 or L2 distance can be used for the multi-modality image registration. One could extract image representations, namely structural representations, such as entropy, Laplacian images, local phase images, and probabilistic edge maps (PEMs). Entropy images satisfy only certain requirements of a relaxed version of the theoretical properties, but they are computationally fast and lead to good alignment, facilitating its use as a practical solution. Laplacian images capture the intrinsic structure of imaging modalities, with the preservation of optimal locality. The local phase image can be a characterization of features detected in a signal, such as edges and ridges in a signal, so it has been shown appropriately for registration involving Echo images. In contrast to local phase representations, PEMs only highlight the image structures that are relevant for image registration, as presented in Fig. 5. Note that several anatomical structures that are visible in US images may not be visible in CT/MR images and vice versa, such as endocardial trabeculae. Therefore, only registering the region of interest (ROI) can be more robust than registering every voxel in the images.
Table 3: Summary of previously published works on the multi-modality cardiac image registration. reg: registration; seg: segmentation; MS-CMRSeg: paired cardiac bSSFP, T2-weighed, and LGE MRIs from Multi-sequence Cardiac MR Segmentation Challenge (Zhuang et al., 2019); X-rayAI: X-ray angiography image; CTA: CT angiography; CSA: coronary sinus angiogram; EAM: electroanatomical mapping; MI: mutual information; NMI: normalized MI; NCC: normalized cross correlation; RMS: root mean squared; MSD: mean square distance; ECC: entropy correlation coefficient; JE: joint entropy; PSMI: point similarity measure based on MI; CoR: correlation ratio; WC: Woods criterion; SEMI: spatially encoded MI; VWMI: Viola-Wells MI; MMI: Mattes MI; SSD: sum of squared difference; WJS: weighted Jensen-Shannon; KL: Kullback-Leibler; LCC: local correlation coefficient; MD: mean distance; ASD: average surface distance; CA: coronary arteries; CoC: correlation coefficient; FFD: free-form deformation; Syn/ Dis: image synthesis/ disentangle based registration.

| Study                  | Modality Target Method Similarity metric |
|------------------------|------------------------------------------|
| Pauna et al. (2003)    | MRI, PET WH intensity-based rigid reg CoR, CoC, MI |
| Guetter et al. (2005)  | CT, PET Myo intensity-based non-rigid reg KL divergence + MI |
| Cremers et al. (2006) | CT, PET Myo intensity-based non-rigid reg KL divergence + MI |
| Guetter et al. (2006)  | CT, SPECT Myo intensity-based rigid reg WJS divergence |
| Huang et al. (2009)    | CT, Echo WH intensity-based temporal + 2D-3D rigid reg MI |
| Papp et al. (2009)     | CT, PET, Myo intensity-based reg triple NMI |
| Intensity based registration |
| Zhang et al. (2006)    | MRI, Echo LV, RV, Myo local phase based affine reg |
| Cordere-Granier et al. 2012 | cine, LGE MRI Myo local entropy based reg NCC |
| Oktay et al. (2015)    | CT, Echo LV, RV, Myo PEM-based rigid + non-rigid reg NCC, LCC |
| Cho et al. (2016)      | MRI, X-ray LV normalized gradient field based rigid reg NCC |
| Kmitra et al. (2017)   | MRI, PET Myo motion field based affine reg + FFD NCC |
| Coussev et al. (2017)  | CT, X-ray CA regression based rigid reg N/A |
| Alehertia et al. (2020) | MRI, Echo Myo salient image based Eulerian and nonrigid reg NCC |
| Savi et al. (1995)     | Echo, PET Myo landmark-based rigid reg MSD |
| Mäkelä et al. (2003)   | MRI, PET, Echo LV, RV, Myo landmark and seg based reg region energy |
| Xiamei et al. (2005)   | Echo, SPECT Myo landmark-based 2D-3D reg MI |
| Nakane et al. (2006)   | MRI, X-ray LV, RV, Myo landmark-based rigid reg MSD |
| Baka et al. (2015)     | CT, X-ray CA landmark-based rigid reg 2D-3D distance |
| Aksoy et al. (2013)    | CT, X-ray CA landmark-based rigid reg 2D-3D distance |
| Doring et al. (2013)   | Echo, CSA Myo landmark-based reg N/A |
| Smith et al. (2014)    | PET, SPECT Myo landmark-based reg N/A |
| Layena et al. (2015)   | CT, X-rayAI CA landmark-based affine reg RMS |
| Khail et al. (2017)    | CT, Echo LA landmark-based rigid reg NMI |
| Faber et al. (1991)    | MRI, SPECT LV point set based rigid reg MD |
| Sinha et al. (1995)    | MRI, PET Myo point set based rigid reg MSD |
| Santaneru et al. (2001) | MRI, SPECT Myo point set based non-rigid reg MSD |
| Strum et al. (2001)    | MRI, Echo LV point set based rigid reg SSD |
| Tavard et al. (2017)   | CT, Echo, EAM LV point set based temporal reg SSD |
| Peoples et al. (2019)  | CT, Echo WH point set based non-rigid reg posterior |
| Scott et al. (2021)    | CT, Echo WH point set based rigid reg N/A |
| Mäkelä et al. (2001)   | MRI, PET Myo seg based rigid reg N/A |
| Woz et al. (2009)      | MRI, SPECT Myo seg based rigid + nonrigid reg SSD |
| Courtail et al. (2019) | cine MRI, CT WH seg based rigid reg NMI |
| Zhuang (2019)          | MS-CMRSeg Myo seg-based rigid reg + FFD posterior SEMI |
| Li et al. (2020)       | bSSFP, LGE WH seg-based affine reg + FFD SEMI |
| Ding et al. (2020)     | MRI, CT Myo seg based non-rigid reg Dice |
| Shi et al. (2012)      | MRI, tagged MRI Myo pseudo-anatomical MRI based nonrigid reg weighed NCC |
| Dey et al. (2012)      | CT, SPECT WH synthetic CT based rigid and non-rigid reg SSD |
| Li et al. (2013)       | CT, Echo LV synthetic CT based non-rigid reg MI |
| Charlaite et al. (2020)| bSSFP, LGE LV, Myo disentangled representation based affine reg MI |
| Structural / anatomical information based registration |
| Aksoy et al. (2013)    | CT, X-ray CA landmark-based rigid reg 2D-3D distance |
| Doring et al. (2013)   | Echo, CSA Myo landmark-based reg N/A |
| Smith et al. (2014)    | PET, SPECT Myo landmark-based reg N/A |
| Layena et al. (2015)   | CT, X-rayAI CA landmark-based affine reg RMS |
| Khail et al. (2017)    | CT, Echo LA landmark-based rigid reg NMI |
| Faber et al. (1991)    | MRI, SPECT LV point set based rigid reg MD |
| Sinha et al. (1995)    | MRI, PET Myo point set based rigid reg MSD |
| Santaneru et al. (2001) | MRI, SPECT Myo point set based non-rigid reg MSD |
| Strum et al. (2001)    | MRI, Echo LV point set based rigid reg SSD |
| Tavard et al. (2017)   | CT, Echo, EAM LV point set based temporal reg SSD |
| Peoples et al. (2019)  | CT, Echo WH point set based non-rigid reg posterior |
| Scott et al. (2021)    | CT, Echo WH point set based rigid reg N/A |
| Mäkelä et al. (2001)   | MRI, PET Myo seg based rigid reg N/A |
| Woz et al. (2009)      | MRI, SPECT Myo seg based rigid + nonrigid reg SSD |
| Courtail et al. (2019) | cine MRI, CT WH seg based rigid reg NMI |
| Zhuang (2019)          | MS-CMRSeg Myo seg-based rigid reg + FFD posterior SEMI |
| Li et al. (2020)       | bSSFP, LGE WH seg-based affine reg + FFD SEMI |
| Ding et al. (2020)     | MRI, CT Myo seg based non-rigid reg Dice |
| Shi et al. (2012)      | MRI, tagged MRI Myo pseudo-anatomical MRI based nonrigid reg weighed NCC |
| Dey et al. (2012)      | CT, SPECT WH synthetic CT based rigid and non-rigid reg SSD |
| Li et al. (2013)       | CT, Echo LV synthetic CT based non-rigid reg MI |
| Charlaite et al. (2020)| bSSFP, LGE LV, Myo disentangled representation based affine reg MI |
| Syn/ Dis |
| Shi et al. (2012)      | cine, tagged MRI Myo pseudo-anatomical MRI based nonrigid reg weighed NCC |
| Dey et al. (2012)      | CT, SPECT WH synthetic CT based rigid and non-rigid reg SSD |
| Li et al. (2013)       | CT, Echo LV synthetic CT based non-rigid reg MI |
| Charterie et al. (2020)| bSSFP, LGE LV, Myo disentangled representation based affine reg MI |
Modality-invariant anatomical landmarks, surfaces (contours or point clouds) or labels can also be used to define the transformation from one image to the other due to their inherent correspondences with original images. For example, Savi et al. (1995), Döring et al. (2013), Smith et al. (2014) employed the landmarks being identified in multi-modality images for registration. Ding et al. (2020) predicted a dense displacement field (DDF) for cardiac CT and MRI registration by aligning the anatomical labels of MRI and CT instead. Nevertheless, these methods are highly sensitive to the accuracy of segmentation or landmark, so manual segmentation or landmark detection is usually required, which is impractical in image-guided intervention.

3.1.3. Image synthesis/ disentangle based registration

An alternative way is to reduce the multi-modality registration into a mono-modality problem, where most existing mono-modality registration methods then can be applied (Li et al., 2013). One could either synthesize one modality from another one or transfer multiple modalities into a common domain (Chen et al., 2017, Arar et al., 2020). The main idea behind this is to mitigate the large appearance gap among different modalities. To achieve this, we can take advantage of prior knowledge of the physical properties of imaging devices (Roche et al., 2001) or capture their intensity relationships (Cao et al., 2017). As for mapping multiple modalities to a common space, these modalities need to share the same anatomical structure or features in order to build a meaningful correspondence. A major issue of current image synthesis based registration methods is that the image synthesis is normally performed in a single direction. Specifically, current image synthesis is often adapted from the imaging modality with rich anatomical details (e.g., MRI) to the imaging modality with limited anatomical details (e.g., CT/ US). Note that it is quite challenging to achieve accurate local deformation prediction based on an imaging modality with limited anatomical details, especially in the presence of large local deformations in cardiac images. One could also disentangle the anatomy and modality information from images, and then a mono-modality registration can be adopted on the latent embedding space (Chartsias et al., 2020). A major issue with current disentanglement based methods is that they generally can not explicitly impose the disentanglement, so the learned representations may be exposed to information leakage (Ouyang et al., 2021).

3.1.4. Discussion

Multi-modality cardiac image registration is a crucial step for further cardiac analysis, providing complementary information for cardiac image fusion (Sec 3.2) and facilitating the cardiac image segmentation (Sec 3.3). It allows for a more comprehensive analysis of the cardiac functions and pathologies, and has many applications in the clinic, such as monitoring the progression of diseases, quantifying the effectiveness of treatment mechanisms, surgery planning, and intra-operative navigation (Döring et al., 2013, Giammogliou et al. (2006), Khalil et al. (2017a). During the registration procedure, different combinations of dimensions, such as 2D-to-2D, 3D-to-3D, 2D-to-3D, and 3D-to-4D, may be involved, and they usually have different applications and optimization targets. For example, 2D-to-3D image registration is normally employed for establishing correspondence between X-ray and 3D volumes. Its main application is to facilitate image-guided surgery (Khalil et al., 2017b, Dori et al., 2011), so computational complexity must be minimized without compromising accuracy.

We have summarized various cardiac image registration methods according to the key elements of methodologies. One can see that structural/ anatomical information based registration is the most popular for multi-modality cardiac image registration. This can be attributed to the fact that the structural/ anatomical information among different modalities is the same, ensuring the accuracy of registration. Also, structural/ anatomical information based registration can use simple similarity metrics instead of computationally expensive MI-based metrics, and thus accelerate the registration procedure. Pure rigid transformations have been applied in several works, but rigid cardiac image registration is generally not enough to describe the spatial relationship due to the existence of cardiac motion. Therefore, there is considerable research extending the transformation to the combination of rigid and non-rigid ones.

Conventional cardiac image registration basically is an iterative-based optimization procedure, which could be quite slow, especially for non-rigid image registration. DL-based registration procedures can be computationally efficient (Boveiri et al., 2020), and have been applied in multi-modality cardiac images (Gouveia et al., 2017, Ding et al., 2020, Luo and Zhuang, 2020). These methods generally require extensive anatomical labels for supervised network training. To solve this, Ding et al. (2021) investigated un-
supervised DL-based cardiac multi-modality image registration by introducing a modality-invariant structural representation, i.e., spatially encoded gradient information. One can also exploit the image similarity analogous to conventional intensity-based image registration for unsupervised DL-based cardiac image registration (de Vos et al., 2019), or employ image disentangle learning to embed an image onto a domain-invariant latent space for the registration (Qin et al., 2019). Also, groupwise registration has recently emerged and has been applied for multi-modality registration and segmentation of cardiac images (Kouw and Loog, 2019; Zhuang, 2019; Luo and Zhuang, 2020). Compared to pairwise registration, groupwise registration is able to handle several imaging modalities simultaneously in an unbiased way (see Fig. 6). Therefore, DL-based multi-modality image registration can be further facilitated by introducing groupwise learning. Nevertheless, there are no general automatic methods due to the wide variety of modalities and clinical scenarios in cardiology.

3.2. Fusion

Multi-modality cardiac image fusion aims at integrating relevant information from several images into a single one, to obtain a more informative and representative integrated image, as presented in Fig. 7. The fusion process can improve image quality and reduce randomness and redundancy, in order to augment clinical accuracy and robustness of the analysis for an accurate diagnosis. For decades, images from different modalities have been used to develop image fusion frameworks either invasive or non-invasive, anatomical and functional, leading to the emergence of hybrid devices such as PET/ SPECT/ CT and MRI (Piccinelli, 2020), as shown in Fig. 8. Specifically in the field of cardiology, image fusion is mainly used to integrate functional imaging, such as SPECT or PET, and anatomical imaging, such as CT or MR. Also, it permits a combination of the advantages of each modality, such as excellent soft-tissue contrast and/or higher spatial resolution (CT or MRI) and high temporal resolution (X-ray fluoroscopy; Echo) for detailed real-time feedback. Among these hybrid cardiac imaging techniques, PET (or SPECT) / CT (CTA) is currently the most widespread (Lawonn et al., 2018). As cardiac MRI and CT are both high-resolution images, their fusion is relatively rare and could be more challenging (Lawonn et al., 2018). Table 4 summarizes the applied modalities in cardiac image fusion with their applications in cardiac imaging studies. Note that as some of these reviewed papers provide limited details about fusion methodology, we classified image fusion works based on their applied modalities instead of the methodologies. Also, we only include parts of representative image fusion studies without pretending to be exhaustive.

3.2.1. PET/ SPECT and CT/ MRI fusion

Healthy cardiac function depends on a close interplay between anatomy and physiology, which is disrupted in many CVDs; thus, their joint evaluation can significantly boost a physician’s ability to diagnose and plan treatment (Piccinelli, 2020). Initial attempts to fuse anatomical and functional information date back to the early 1990s and have gradually become established techniques, demonstrating the greater diagnostic power of fused images compared to single-mode or side-by-side interpretation (Peifer et al., 1990; Piccinelli et al., 2018). SPECT and PET can evaluate myocardial perfusion, metabolism, and regional absolute myocardial blood flow, while CT/ MRI provides coronary anatomy and allows for multi-parametric evaluation of cardiac morphology, ventricular function, myocardial perfusion, and viability. Integrated SPECT/ PET and CT/ MRI allow true simultaneous analysis of the structure and function of the heart and has been applied in clinical practice, as presented in the first row of Ta-
Table 4: Summary of previously published works on the multi-modality cardiac image fusion. PV: pulmonary vessel; PC: pulmonary conduit; PA: pulmonary artery; LSVC: left superior vena cava; MV: mitral valve; MyoI: myocardial infarction; CCAA: complex CA anomalies; PCI: percutaneous coronary intervention; AF: atrial fibrillation; TAVI: trans-aortic valve implantation; CHD: congenital heart disease; CRT: cardiac resynchronization therapy; MBF: myocardial blood flow; MFR: myocardial flow reserve; RFR: relative flow reserve; iAS: inherited arrhythmia syndromes; SHD: structural heart disease; EM: electromagnetic; STDME: simultaneous two-screen display of multidetector-CT and real-time echogram; TE: transcendicular; VNE: virtual native enhancement; MSF: multi-sequence MRI fusion.

| Study | Modality | Target | Method | Motivation/Application |
|-------|----------|--------|--------|------------------------|
| Grani et al. (2017) | PET, CT | CA | CardIQ fusion software | assess the impact of CCAA on Myo perfusion |
| Aguadé-Bruix et al. (2017) | PET, CT | Aorta | N/A | diagnose and evaluate the extent of endocarditis |
| Piccinelli et al. (2020) | PET, CTA | Myo | rigid reg | extract MBF, MFR and RFR along CAs |
| Quail and Smisshaven et al. (2017) | PET, MRI | Myo | N/A | visualize each CA and strain in its territory |
| Lizardo-Garcia et al. (2020) | PET, MRI | Myo | OsiriX DICOM viewer | explore the effect of histone deacetylas on the heart |
| Zanuto et al. (2018) | PET, MRI | Myo | landmark based rigid reg | diagnose patients with cardiac sarcoidosis |
| Degraeuwe et al. (2017) | PET, MRI, CT | Myo, PV | N/A | assess cardiac parangangliomas |
| Gaemperli et al. (2007) | SPECT, CT | Myo | CardIQ fusion software | combine information of CA and lesion |
| Sazonova et al. (2017) | SPECT, CT | PC | hybrid devices | diagnose infectious endocarditis |
| Siebermair et al. (2020) | SPECT, CT | LV, RV | N/A | evaluate chamber-specific patterns of automatic innervation in IAS |
| Koukouraki et al. (2013) | SPECT, CTA | Myo | CardIQ Fusion software | clinical management of patients with suspected CAD |
| Nakahara et al. (2016) | SPECT, CTA | Myo | fusion based bull’s eye | determine hemodynamically relevant coronary vessels |
| Kirisli et al. (2014) | SPECT, CTA | Myo | point-cloud based reg | allow Myo perfusion defect correlations with corresponding CA |
| Piccinelli et al. (2018) | SPECT, CTA | Myo | rigid reg | detect and localize CAD |
| Yoneyama et al. (2019) | SPECT, CTA | Myo | manual registration | defect correlations with corresponding CA |
| Mallessant et al. (2017) | Echo, CT | CA | N/A | visualize each CA and strain in its territory |
| Lazzari et al. (2014) | Echo, CT, EAM | LV | N/A | extraction of local electro-mechanical delays |
| Bruge et al. (2015) | Echo, CT | CA | N/A | select the LV pacing sites |
| Watanabe et al. (2021) | Echo, CT | CA | STDME technique | assess conduit stenosis in complex adult CHD |
| Kiss et al. (2011) | Echo, MRI | LV, RV, Myo | landmark based reg | assess Myo viability and diagnose ischemia |
| Kiss et al. (2013) | Echo, MRI | LV | landmark based rigid reg | guide the echocardiographic acquisition |
| Gomez et al. (2020) | Echo, MRI | WH | landmark based reg | 3D printed heart models |
| Hatt et al. (2013) | Echo, MRI, X-ray | LV | N/A | precise targeting for TE therapeutic delivery |
| Zhou et al. (2014) | X-ray, SPECT | LV | vessel-surface rigid reg | guide LV lead implantation for CRT |
| Ma et al. (2010a) | X-ray, (LGE) MRI | LV, scar | manual reg | guide LV lead implantation for CRT |
| Dori et al. (2013) | X-ray, MRI | PA | landmark based rigid reg | facilitate cardiac catheterization of CHD |
| Faranesh et al. (2013) | X-ray, MRI | AO, CA | affine motion model | facilitate cardiac interventions |
| Abou Hazeem et al. (2014) | X-ray, MRI | PA, LVSC | landmark based reg | facilitate cardiac catheterization of CHD |
| McGuirt et al. (2016) | X-ray, MRI | PA, LVSC | landmark based reg | facilitate cardiac catheterization of CHD |
| Choi et al. (2016) | X-ray, MRI | LV | edge based rigid reg | guide LV lead implantation for CRT |
| Grant et al. (2019) | X-ray, MRI | AO | landmark based rigid reg | select transcatheter CHD interventions |
| Ma et al. (2010a) | X-ray, MRI | LV | 2D-3D manual reg | guide LV lead implantation for CRT |
| Tomko et al. (2011) | X-ray, (LGE) MRI | LV, scar | custom software | guide catheter-based TE delivery |
| Ghoshhajra et al. (2017) | X-ray, CT | CA | reg with manual correction | guide chronic total occlusion of PCI |
| Vennikovska et al. (2018) | X-ray, CTA | AO | reg with manual correction | facilitate TAVI procedures |
| Ma et al. (2010a) | X-ray, Echo | LV, RV | intensity-based rigid reg | guide cardiac catheterization procedures |
| Clegg et al. (2015) | X-ray, Echo | MV | probe tracking | guide percutaneous SHD interventions |
| Housien et al. (2012) | X-ray, Echo | LA, RA | probe based 2D-3D reg | facilitate catheter ablation and TAVI |
| Zorinas et al. (2017) | X-ray, Echo | AO | EchoNavigator system | assist catheter-based cardiac operations |
| Hadeed et al. (2018) | X-ray, Echo | N/A | EchoNavigator system | guide interventional procedures for CHD |
| Ebel et al. (2020) | X-ray, Echo | LAA | TrueFusion technology | improve the procedure of LAA closure |
| Cordero-Grande et al. (2012) | cine, LGE MRI | Myo | local entropy based reg | assess Myo |
| Zhang et al. (2021a) | cine, T1 mapping | Myo | DL VNE generator | generate gadolinium-free LGE-like MRI |
ble [1]. Compared to PET-CT, hybrid PET-MRI offers some advantages in that it does not require exposure to ionizing radiation from CT and iodinated contrast agents (Bergquist et al., 2017). It can be applied for the assessment of Myo [Lee et al., 2012], cardiac sarcoidosis (Zandieh et al., 2018), atherosclerosis (Cuadrado et al., 2016), non-ischemic cardiomyopathies (Zhang et al., 2022), myocarditis (Nensa et al., 2018), vasculitis (Laurent et al., 2019), and cardiac tumors (Rinuncini et al., 2016). The fusion of SPECT with CT (especially CTA) images is regarded as the most successful example of fusion imaging [Kramer and Narula, 2010], and has been widely employed for the assessment of CAD (Koukouraki et al., 2013; Piccinelli et al., 2018).

3.2.2. Echo and CT/ MRI fusion

Echo has great potential and offers clear benefits when complementing other modalities due to its ability to provide real-time information. The limited FOV and noise in Echo data could be alleviated by integrating it with additional imaging modalities. For example, TEE can be used for the assessment of morphology, function and hemodynamics in most adult congenital heart disease (CHD) patients, while CT enables extensive anatomical analysis due to its high spatial resolution. Several methods have been developed for synchronized display of real-time Echo images and multi-planar reconstruction images of CT or MRI (Watanabe et al., 2021). The fused images can provide additional findings, allowing accurate assessment of the diagnosis and severity of diseases, compared to Echo alone. For example, it can be employed in cardiac surgery, longitudinal studies, and to derive a more complete picture of the heart (motion from Echo and tissue perfusion from MRI) (Takaya and Ito, 2020).

3.2.3. X-ray and CT/ MRI/ Echo fusion

X-ray fluoroscopy is commonly used for guiding minimally invasive cardiac interventions due to its excellent catheter and device visualization. However, it is limited by the 2D projection nature of the images, exposure to radiation, and poor soft-tissue contrast, which however can be solved with the use of diagnostic quality 3D images, such as MRI, CT and real-time 3D Echo. For example, the fusion of X-ray and MRI can be employed to facilitate cardiac catheterization of CHD patients with significant reductions in contrast and radiation exposure and decide the optimal pacing of the LV lead during CRT procedures. The introduction of additional LGE MRI during this fusion can further assist cardiologists to avoid scarring regions, as positioning an LV lead within the scarring areas may reduce response to CRT (Ma et al., 2010b). The fusion of X-ray and CT (or CTA) has been used to guide percutaneous procedures (Dori et al., 2011), guide chronic total occlusion of percutaneous coronary intervention (Ghoshhajra et al., 2017), and facilitate trans-aortic valve implantation (TAVI) procedures (Vernikousskaya et al., 2018). Real-time 3D stereo echocardiography (such as 3D TEE) can be fused with X-ray fluoroscopy to guide percutaneous structural heart disease interventions (Clegg et al., 2015) and facilitate AF catheter ablation as well as TAVI procedures (Housden et al. 2012). As opposed to fluoroscopy, 3D TEE provides excellent detail of 3D anatomy and soft tissue structures and offers “real-time” intraoperative guidance (Clegg et al., 2015). However, its advantages are limited by Echo shadowing, which reduces visualization of the catheters and metallic structures [Zorinas et al., 2017]. Nevertheless, compared to CT and MRI, 3D Echo is widely available in hospitals at a much lower cost and provides real-time cardiac anatomical and function information for the fusion. With the recent availability of commercially available fusion packages on commercial X-ray systems, a wider range of clinical applications has been developed based on image fusion related to X-rays. Fig. 9 presents a clinical application example of X-ray and Echo image fusion.

3.2.4. Multi-sequence MRI fusion

As MRI has different imaging sequences, one could fuse multi-sequence MRI to combine complementary information from different sequences. For example, Cordero-Grande et al. (2012) fused cine and LGE MRIs for precise representation of both the Myo boundaries and scars. Zhang et al. (2021a) fused cine and T1 mapping MRIs to generate gadolinium-free LGE-like MRI via a DL-based generator. Reducing the need for gadolinium can dramatically shorten scan times, lower the cost of associated consumables, minimize patient preparation time, and eliminate the need for physician presence. T1 mapping appears the most promising gadolinium-free technique but its clinical utility is still limited by the lack of standardized interpretation and post-processing. Therefore, one could combine T1 mapping and cine MRIs to exploit and enhance existing contrast and signals within them and display them in a standardized presentation.

3.2.5. Discussion

Image fusion typically includes two stages: (a) image registration (please refer to Section 3); (b) fusion of relevant features, which could be gray-level based or component-
based analysis, from the aligned images. To avoid the registration step, several hybrid scanners have been developed to integrate multiple modalities (Sazonova et al., 2017; Lavernom et al., 2018). With the scanners, there still exist registration errors mainly due to cardiac or respiratory motion for cardiac imaging, so a post-alignment process remains desired (García et al., 2009). The feature fusion involves the identification and selection of the features for specific clinical assessment purposes. In this study, we have found that the most common hybrid version is combining low-resolution molecular images (PET/ SPECT) with anatomical context from high-resolution MR/ CT; The most common target regions of fusion are LV Myo and LV, which are mainly involved with the management of CAD and the lead implantation for CRT, respectively.

Compared to other medical image fusion (brain, breast, liver, etc.) where many effective fusion schemes have been developed, cardiac image fusion is still at an early stage (James and Dasarathy, 2014). The works we surveyed did not provide many details on the fusion schemes, and most of them achieved static fusion. There were only a few software products commercially available for real-time fusion between dynamic images. For example, EchoNavigator is a software tool that enables real-time image synchronization and fusion of 2D and 3D TEE with fluoroscopy images. The coordinate system of the two imaging modalities can be aligned based on the localization and tracking of the TEE probe. This software might facilitate complex percutaneous procedures and decrease procedure length as well as radiation dose. Note that the software was installed only in a few hospitals at the time of publication (Biaggi et al., 2015), which limits the clinical evaluation of fusion images. Recently, many DL-based image fusion methods have been developed for brain image fusion (Zhang et al., 2021b), but they have not been applied to cardiac image fusion yet. Nevertheless, there are several DL-based works that achieved implicit cardiac image fusion in the latent space for multi-modality image segmentation (please c.f. Section 3.3.2 for details). Given a large amount of data, deep neural networks are powerful for learning hierarchical feature representations. Note that there are no general criteria for network architecture design, and the questions of what, when and how, applied to cardiac image fusion, remain unresolved. Moreover, current research mainly focuses on the fusion of two modalities, and the fusion of three or more modalities is rarely studied in literature (Tirupal et al., 2021). Despite extensive research and test applications, multi-modality image fusion techniques have not yet been extensively translated into clinical routine. In conclusion, fused imaging further increases the variety of cardiac imaging modalities (modality augmentation), but effective techniques are still under development.

3.3. Segmentation

Multi-modality cardiac image segmentation has received a substantial research attention, including both anatomical segmentation (Zhuang et al., 2020) and pathology segmentation (Zhuang and Li, 2020). It needs to solve the difference in data distribution among different modalities normally via registration, fusion, or domain adaptation (modality transfer or disentanglement). Fig. 10 provides the sketch maps of these modality processing schemes. The image registration employs spatial transformation parameters to align images of different modalities and then performs segmentation on the anatomy-aligned images (Luo and Zhuang, 2020; Zhuang, 2019). Based on the registration results, one could further fuse several modalities into a single one, and then perform segmentation on the integrated images/ features (Zhuang and Shen, 2016). Moreover, one could utilize transfer learning to transfer different modalities to the target domain and then perform segmentation on the modality-aligned images/ features (Chen et al., 2019c). Another similar idea is to disentangle the anatomy and modality information from cardiac images, and the target region can be directly segmented on the disentangled anatomy images/ features (Chartsias et al., 2019a). Table 5 summarizes the representative multi-modality cardiac image segmentation methods, the applied modalities and the target substructures of the heart.

3.3.1. Registration based segmentation

A direct way to propagate anatomy knowledge between modalities is multi-modality registration, and these methods can be categorized as registration based segmentation. Among registration based segmentation methods, atlas registration based approaches are commonly used in multi-modality scenarios. They register one or multiple atlases to the target image, followed by propagation of (usually manual) labels. We refer the reader to Sec. 3 for multi-modality registration. If several atlases are available, labels from individual atlases can be combined into a final segmentation via a label fusion strategy (Zhuang and Shen, 2016; Ding et al., 2020). The number of atlases determines the potential optimal performance of multi-atlas segmentation (MAS), and thus considering multi-modality atlases is beneficial when they are available. Conventional multi-modality MAS based methods are generally computationally expensive. This is because these methods perform the registration step in an iterative fashion and typically employ patch-based label fusion (Bai et al., 2013; Zhuang and Shen, 2016; Sanroma et al., 2018). To solve this, one could achieve both image registration and label fusion by deep neural networks for a computationally efficient MAS framework (Ding et al., 2020).

However, these DL-based MAS frameworks can not be optimized in an end-to-end fashion, as the image registration and label fusion are separated into two tasks. Recently, Luo and Zhuang (2020) proposed a probabilistic image registration framework based on a multivariate mixture model and neural network estimation. In their framework, MAS was unified by groupwise registration, and registration and segmentation of multi-modality cardiac images were achieved simultaneously. The joint distribution of multi-modality images was modeled as multivariate
Table 5: Summary of previously published representative works on multi-modality cardiac segmentation. MM-WHS: unpaired cardiac CT and MRI dataset from Multi-Modality Whole Heart Segmentation Challenge [Zhuang et al., 2020]; MyoPS: paired cardiac bSSFP, T2-weighed, and LGE MRIs from Myocardial Pathology Segmentation Combining Multi-sequence CMR Challenge [Li et al., 2022a]; X-rayAI: X-ray angiography image; Ti-CE: contrast-enhanced T1 imaging; MAS: multi-atlas segmentation; CNN: Convolutional neural network; MTL: multi-task learning; HM: histogram matching; UDA: unsupervised domain adaptation; FCNN: full convolutional neural network; MMD: maximum mean discrepancy; Reg-based seg: registration based segmentation; \(^1\) denotes that the evaluation dataset also includes non-cardiac or single-modality dataset which is out of the scope of this study.

| Study | Modality | Target | Method | Type |
|-------|----------|--------|--------|------|
| Peters et al. (2010) | CT, MRI, X-rayAI | WH, LA | simulated search | supervised |
| Zhuang and Shen (2016) | CT, MRI | WH | multi-scale and multi-modality MAS | supervised |
| Zheng et al. (2019) | MS-CMRSeg | LV, RV, Myo | DL-based registration and segmentation | supervised |
| Ding et al. (2020) | MM-WHS | Myo | DL-based cross-modality MAS | supervised |
| Zhuang et al. (2019) | MS-CMRSeg | Myo | multivariate mixture model | supervised |
| Luo and Ziaian (2020) | MS-CMRSeg | LV, RV, Myo | DL-based mixture model | supervised |
| Paknejad et al. (2020) | tagged, cine MRI | Myo | deformation-based segmentation | supervised |
| Mortazi et al. (2017) | MM-WHS | WH | multi-planar network with an adaptive fusion strategy | supervised |
| Tong et al. (2017) | MM-WHS | WH | modality normalization and fusion | supervised |
| Zhao et al. (2020) | MyoPS | LV scar and edema | stacked and parallel U-Nets for fusion | supervised |
| Jiang et al. (2021) | MyoPS | LV scar and edema | max-fusion U-Net | supervised |
| Li et al. (2021) | MS-CMRSeg, MyoPS | LV, RV, Myo, LV | deep co-training | supervised |
| Ankenbrand et al. (2020) | MyoPS | LV scar and edema | ensemble U-Net | supervised |
| Zhai et al. (2020) | MyoPS | LV scar and edema | coarse-to-fine weighted ensemble model | supervised |
| Martin-Isla et al. (2020) | MyoPS | LV scar and edema | stacked BCDU-net with MRI synthesis | supervised |
| Wang et al. (2022) | MM-WHS, MS-CMRSeg | LA, LV, RV, Myo, AO | affinity-guide CNN | supervised |
| Liao et al. (2020) | MM-WHS | WH | cycle- and shape-consistency GAN | supervised |
| Yuan et al. (2019) | CT, MRI | WH | cycle- and shape-consistency GAN | supervised |
| Cai et al. (2016) | MS-CMRSeg | LV, RV, Myo | shape-transfer GAN | supervised |
| Ly et al. (2019) | MS-CMRSeg | LV, RV, Myo | image-to-image translation | supervised |
| Wang et al. (2019) | MS-CMRSeg | LV, RV, Myo | style data augmentation | supervised |
| Liao et al. (2020) | MM-WHS | LA, LV, Myo, AO | HM and domain adversarial learning | supervised |
| Chen et al. (2016) | MM-WHS | LA, LV, Myo, AO | plug-and-play adversarial UDA | supervised |
| Chen et al. (2019) | MM-WHS | LA, LV, Myo, AO | VAE-based feature prior matching | supervised |
| Chen et al. (2020) | MM-WHS, MS-CMRSeg | LA, LV, RV, Myo, AO | domain adaptation | supervised |
| Liao et al. (2020) | MM-WHS | WH | multi-modality transfer learning | semi-supervised |
| Yu et al. (2020) | CT, MRI | Myo | MTL + adversarial reverse mapping | supervised |
| Liu and Liu (2020) | MS-CMRSeg | LV, RV, Myo | explicit domain discrepancy | supervised |
| Li et al. (2020) | MM-WHS | WH | dual-teacher++ | semi-supervised |
| Xue et al. (2020) | MM-WHS | WH | dual-task and hierarchical learning | supervised |
| Bian et al. (2020) | MM-WHS\(^1\) | LA, LV, Myo, AO | uncertainty-aware domain alignment | unsupervised |
| Wu and Zhuang (2021) | MM-WHS, MS-CMRSeg | WH, LV, RV, Myo | variational approximation | unsupervised |
| Wang et al. (2021) | MM-WHS | LA, LV, Myo, AO | zero-shot learning | unsupervised |
| Dou et al. (2021) | MM-WHS\(^1\) | LA, LV, Myo, AO | information bottleneck GAN | unsupervised |
| Chen et al. (2021) | MM-WHS\(^1\) | LA, LV, Myo, AO | diverse data augmentation GAN | unsupervised |
| Cui et al. (2021) | MM-WHS | LA, LV, Myo, AO | bidirectional UDA | unsupervised |
| Wu et al. (2021) | MM-WHS | LA, LV, Myo, AO | hybrid domain-invariant information | unsupervised |
| Cui et al. (2021) | MM-WHS\(^1\) | LA, LV, Myo, AO | spatial adaptive normalization | unsupervised |
| Chen et al. (2021) | MM-WHS, MS-CMRSeg | LV, RV, Myo | point-cloud shape adaptation | unsupervised |
| Cui et al. (2021) | MM-WHS\(^1\) | LA, LV, Myo, AO | semantic consistent UDA | unsupervised |
| Cui et al. (2021) | MM-WHS, CT, Echo | LV, Myo | few-shot multi-level semantic adaptation | semi-supervised |
| Cui et al. (2018) | MM-WHS | LA, LV, Myo, AO | symmetric FCNN with attention | semi-supervised |
| Cui et al. (2020) | MM-WHS\(^1\) | LA, LV, Myo, AO | deep co-training | semi-supervised |
| Dou et al. (2020) | CT, MRI | LA, LV, Myo, WH | groupwise; spectral decomposition | unsupervised |
| Chartsias et al. (2019) | MM-WHS\(^1\) | LA, LV, Myo, AO | knowledge distillation | supervised |
| Chartsias et al. (2019) | bSSFP + LGE MRI | LV, Myo | disentangled representation learning | semi-supervised |
| Chartsias et al. (2020) | MM-WHS\(^1\) | LV, Myo | disentangled representation learning | semi-supervised |
| Chartsias et al. (2020) | bSSFP, LGE MRI\(^1\) | LV, Myo | disentangled align and fuse network | semi-supervised |
| Wang et al. (2021) | MM-WHS, MS-CMRSeg | LV, RV, Myo | cycle-consistent model | unsupervised |
mixtures, and the model was formulated with transformation. Segmentation can be performed on a virtual common space, where all the images were simultaneously registered. Instead of the commonly used Expectation Maximization (EM) algorithm (Zhuang 2019), they utilized neural networks to efficiently estimate the parameters of multivariate mixture model (Luo and Zhuang 2020).

### 3.3.2. Fusion based segmentation

To the best of our knowledge, there are only a few works that employ image fusion for multi-modality cardiac segmentation. Most fusion based segmentation methods do not explicitly generate a new integrated image by image fusion (also namely explicit fusion), but only perform an implicit fusion at different levels. The only available explicit fusion based segmentation, up to our knowledge, is from Tong et al. (2017), where cardiac MRI and CT were fused based on their normalized mean intensity values. With the fused images, the size of the training set was increased, and therefore the segmentation model yielded better results. With the popularization of DL methods, DL-based image fusion has emerged, as it can learn high-level features from data in an incremental manner compared with conventional methods (Zhou et al., 2019b). It mitigates the need for domain expertise, simplifies the feature extraction steps, and can be optimized in an end-to-end style.

Fig. 11 presents the three typical types of DL-based modality fusion for multi-modality cardiac image segmentation. Many works simply concatenate different modalities into different channels and focus on the subsequent segmentation network architecture, namely input-level fusion (Wang et al., 2019b; Yu et al., 2020b; Zhang et al., 2020). Similarly, one could perform fusion by fusing the model outputs from different modality inputs, namely decision-level fusion. For the decision-level fusion, many fusion strategies can be used, such as averaging and majority voting (Rokach, 2010). Both input- and decision-level fusion are simple, but ignore the relationship among different modalities. A potential solution is to utilize two-stage networks which consider the prior spatial relationship between different modalities. One typical application is MyoPS, where anatomical information can be extracted from one modality (bSSFP MRI) at the first stage; with obtained segmentation masks from the first stage, the functional information can be learned by combining other modalities (LGE and T2 MRIs) at the second stage (Zhuang and Li, 2020). Moreover, fusion can be performed on latent feature space (Mortazi et al., 2017; Jiang et al., 2020; Li et al., 2020; Zhao et al., 2020), namely feature-level fusion, and multi-scale information fusion can be achieved via densely connected layers (Zhao et al., 2020). For example, Jiang et al. (2020) proposed a max-fusion U-Net for multi-sequence cardiac pathology segmentation, where the features from different modalities were fused via a pixel-wise maximum operator. The operator was expected to guide the network to keep informative features extracted by each modality. Furthermore, the contribution of each modality can be weighted via cross-modal convolutions (Tseng et al., 2017).

Compared with input/ decision-level fusion, feature-level fusion can pay more attention to learning the complex relationship between different modalities (Zhou et al., 2019b). However, the features to be merged are not modality-invariant, and therefore may affect the multi-modality segmentation. To solve this, one could combine disentangled representation learning with fusion schemes for the segmentation. For example, Chartsias et al. (2020) disentangled modality and anatomy features from multi-modality images, and then fused the disentangled anatomy features for cardiac segmentation of bSSFP and LGE MRIs. Besides modality fusion, model ensemble scheme is also widely employed for multi-modality cardiac segmentation by integrating several predictions from different models (Ankenbrand et al., 2020; Zhai et al., 2020; Martín-Isla et al., 2020).

### 3.3.3. Domain adaptation based segmentation

For multi-modality cardiac analysis, the annotation of all modalities via supervised learning can be tedious, time-consuming, and significantly variable across imaging modalities. An intuitive solution is to transfer the knowledge from the annotation-rich modalities (denoted as source domain) to another annotation-poor modality (referred to as target domain). Note that here the source and target domains could be collected from the same or different subjects. Due to the existence of domain shift, the model trained on the source domain usually fails on the target domain (Wu and Zhuang, 2020; Li et al., 2022b). Domain adaptation is therefore proposed to generalize models from

---

**Figure 10:** Four typical multi-modality image processing schemes. Here, $I_{\text{target}}$ denotes target image, while $I_{\text{source}}$ refers to source image. $I_{\text{source}}$, $I_{\text{fused}}$, $I_{\text{source}}$ are deformed source image, fused image, and target-style source image, respectively. $I_{\text{anatomy}}$ and $I_{\text{modality}}$ are the disentangled anatomy image and modality information, respectively. Note that in this illustration the modality fusion is performed on the pre-aligned images, i.e., $I_{\text{source}}$ and $I_{\text{target}}$. 

---

| Image Registration | Image Fusion | Modality Transfer | Modality Disentangle |
|--------------------|--------------|------------------|---------------------|
| $I_{\text{source}}$ | $I_{\text{fused}}$ | $I_{\text{target}}$ | $I_{\text{anatomy}}$, $I_{\text{modality}}$ |

---

| $I_{\text{source}}$ | $I_{\text{target}}$ | $I_{\text{fused}}$ | $I_{\text{source}}$, $I_{\text{fused}}$, $I_{\text{source}}$, $I_{\text{anatomy}}$, $I_{\text{modality}}$ | $I_{\text{anatomy}}$, $I_{\text{modality}}$ |
|-------------------|-------------------|-------------------|--------------------------|--------------------------|
| Source image | Target image | Fused image | Source image, Fused image, Source image, Anatomy, Modality | Anatomy, Modality |

---

| $I_{\text{source}}$ | $I_{\text{target}}$ | $I_{\text{fused}}$ | $I_{\text{source}}$, $I_{\text{fused}}$, $I_{\text{source}}$, $I_{\text{anatomy}}$, $I_{\text{modality}}$ | $I_{\text{anatomy}}$, $I_{\text{modality}}$ |
|-------------------|-------------------|-------------------|--------------------------|--------------------------|
| Source image | Target image | Fused image | Source image, Fused image, Source image, Anatomy, Modality | Anatomy, Modality |

---

| $I_{\text{source}}$ | $I_{\text{target}}$ | $I_{\text{fused}}$ | $I_{\text{source}}$, $I_{\text{fused}}$, $I_{\text{source}}$, $I_{\text{anatomy}}$, $I_{\text{modality}}$ | $I_{\text{anatomy}}$, $I_{\text{modality}}$ |
|-------------------|-------------------|-------------------|--------------------------|--------------------------|
| Source image | Target image | Fused image | Source image, Fused image, Source image, Anatomy, Modality | Anatomy, Modality |

---

| $I_{\text{source}}$ | $I_{\text{target}}$ | $I_{\text{fused}}$ | $I_{\text{source}}$, $I_{\text{fused}}$, $I_{\text{source}}$, $I_{\text{anatomy}}$, $I_{\text{modality}}$ | $I_{\text{anatomy}}$, $I_{\text{modality}}$ |
|-------------------|-------------------|-------------------|--------------------------|--------------------------|
| Source image | Target image | Fused image | Source image, Fused image, Source image, Anatomy, Modality | Anatomy, Modality |
Figure 11: An example of multi-modality cardiac image segmentation. Here, we present three different deep learning based fusion strategies for myocardial pathology segmentation combining multi-sequence MRIs. Note that here bSSFP, T2-weighted, and LGE MRIs have been aligned before fusion.

Figure 12: An example of unsupervised domain adaptation based cross-modality cardiac image segmentation. Image adapted from Chen et al. (2019b) with permission.

3.3.4. Discussion

Multi-modality image segmentation benefits from both aforementioned image registration and fusion especially when involving a combination of multiple modalities. However, currently image registration or explicit image fusion is mostly used as a pre-processing step for multi-modality segmentation (Li et al., 2022a; Tong et al., 2017). Also, no work is able to simultaneously achieve multi-modality fusion from unregistered data for cardiac image segmentation, as far as we know. Only a few studies achieved simultaneous registration and segmentation for multi-modality cardiac images via multivariate mixture models (Zhuang, 2019; Luo and Zhuang, 2020). Unified registration and segmentation frameworks eliminate the need for a separate registration model, and segmentation can facilitate registration and vice versa. In the future, it would be desirable to achieve simultaneous registration and fusion of multi-modality data for cardiac segmentation.

Domain adaptation can be regarded as a new application of multi-modality imaging. It does not combine information from different modalities but transfers information across modalities, to assist the image segmentation of the target domain which usually has limited available annotated data. Modality information transfer can also be employed in cardiac image reconstruction, where cardiac motion priors can be obtained from other modalities for motion correction (Sang et al., 2021). An alternative idea to domain adaptation that appears in multi-modality image registration is the disentangled representation. It is reasonable as both multi-modality image registration and segmentation care about the structural information instead of the modality information. Moreover, compared to cardiac image registration and fusion, DL-based models have been widely employed for multi-modality cardiac image segmentation, thanks to the release of public datasets.

4. Data and evaluation measures

4.1. Public multi-modality cardiac datasets

The Cardiac Atlas Project provides a public imag-
ing database for computational modeling and statistical atlases of the heart (Fonseca et al., 2011). It includes images targeted to different cardiac structures as well as diseases and acquired with different modalities from different centers. Other challenge events have released public multi-modality cardiac datasets, as summarized in Table 6. One can see that different from the computer vision datasets, there are a limited number of public multi-modality cardiac image datasets. Also, only some specific modalities are covered, including MRI, CT and Echo. Nevertheless, these datasets have already substantially facilitated the development of multi-modality cardiac imaging analysis, as we summarized in Sec 3.3. In the future, more multi-modality cardiac images, either paired or unpaired, are expected to be released to boost the development of multi-modality cardiac image analysis.

4.2. Evaluation measures

Thorough validation is necessary for clinical acceptance. Verifying the accuracy of multi-modality image registration or fusion is a difficult task as a good quality gold standard is normally unavailable. Visual assessment, sometimes used to evaluate the accuracy, is subjective and thus not reproducible. For an objective evaluation, external labels, anatomical landmarks and/or external fiducial frames are required (Mäkelä et al., 2002). In contrast, the evaluation of multi-modality image segmentation is relatively simple as manual segmentation can be used as the gold standard. Nevertheless, the comparison of measures from literature is difficult as different metrics can be applied to evaluate one method. Note that sometimes different evaluation metrics could lead to different conclusions regarding the performance of an algorithm, indicating the potential limitation of current metrics. In this section, we summarize several common measures employed in each multi-modality image computing task.

4.2.1. Image registration measures

For assessing the performance of cardiac multi-modality image registration, a range of different measures have been explored (Van de Kraats et al., 2005). The registration may involve different combinations of dimensions, with the most common combinations being 3D-to-3D and 2D-to-3D. For 3D-to-3D registration, one could employ mean landmark distance (mLD) (Sinha et al., 1995), mean target registration error (mTRE), and surface distance (Sturm et al., 2003; Mäkelä et al., 2001) for evaluation. Here, mLD can be defined as,

$$mLD(\hat{T}) = \frac{1}{N_l} \sum_{i,j=1}^{N_l} \| \hat{T}(l_i) - l_j \|,$$  (1)

where $\hat{T}$ is the predicted transformation, $N_l$ is the number of landmarks, and $l_i$ and $l_j$ are paired landmarks from source and target images, respectively. mTRE is computed as the mean 3D distance between the point $p_j$ from the target image space and the corresponding point $p_i$ from the source image space transformed with the predicted transformation $\hat{T}$,

$$mTRE(\hat{T}) = \frac{1}{N_p} \sum_{i,j=1}^{N_p} \| \hat{T}(p_i) - p_j \|,$$  (2)

where $N_p$ is the number of points. mTRE is a good measure for patient-to-image registration for navigation, as it calculates the misalignment on specific points of interest within the target volume where the operation will be performed. Surface distance can be defined as the mean 3D Euclidian distance, standard deviation, and root mean square distance between the transformed source image and target image surfaces (Sturm et al., 2003). Note that here the surface is sampled after registration in the format of paired corresponding points. Therefore, one could employ the percent number of overlapped point pairs as a registration measure. Checkerboard visualization for visual assessment is also commonly used (Mäkelä et al., 2003; Camara et al., 2009; Courtial et al., 2019), as presented in Fig. 13. Due to the absence of the gold standard, one could instead evaluate the anatomical labels of transformed source images in terms of Dice and Hausdorff distance (HD) (Turco et al., 2016; Atelhörtiu et al., 2020), defined as,

$$\text{Dice}(V_1, V_2) = \frac{2|V_1 \cap V_2|}{|V_1| + |V_2|},$$  (3)

and

$$\text{HD}(X, Y) = \max \left[ \sup_{x \in X, y \in Y} d(x, y), \sup_{y \in X, x \in X} d(x, y) \right],$$  (4)

where $V_1$ and $V_2$ denote the set of pixels in the target and transformed source labels, respectively; $| \cdot |$ refers to the number of pixels in set $V$; $X$ and $Y$ represent two

| Reference          | Dataset             | Target                  | Disease | Type |
|--------------------|---------------------|-------------------------|---------|------|
| Utah (2012)        | 155 bSSFP, LGE MRI | LA cavity and scar      | AF      | paired |
| Tobon-Gomez et al. (2013) | 15 MRI, 3D Echo | Myo                     | healthy | paired |
| Tobon-Gomez et al. (2015) | 30 MRI, 30 CT | LA cavity               | AF      | unpaired |
| Karim et al. (2018) | 10 MRI, 10 CT      | LA wall                 | AF      | unpaired |
| Zhuang et al. (2019) | 60 MRI, 60 CT      | WH                      | AF, CHD, others | unpaired |
| Zhuang et al. (2020) | 45 bSSFP, T2 weighted, LGE MRI | LV, RV, Myo | Myo | paired |
| Li et al. (2022a)  | 45 bSSFP, T2 weighted, LGE MRI | LV scar and edema | Myo | paired |
sets of contour points; and \(d(x, y)\) indicates the Euclidean distance between the two points \(x\) and \(y\); Besides, one could randomly transform the source image with known transformations for the known gold standard to quantify registration accuracy (Pauna et al., 2003). For 2D-to-3D registration, visual inspection (Aksoy et al., 2013), mTRE\(_{\text{proj}}\) (Gouveia et al., 2017), and centreline 2D distance measure (Baka et al., 2013; Van de Kraats et al., 2005) have been utilized for evaluation. Here, mTRE\(_{\text{proj}}\) refers to the mTRE in the projection direction and is defined as,

\[
mTRE_{\text{proj}}(\hat{T}) = \frac{1}{N_p} \sum_{i,j=1}^{N_p} \| (\hat{T}(p_i) - p_j) \cdot \hat{n} \|, \tag{5}\]

where \(\hat{n}\) is the normal to the projection plane. However, 2D error measures may be more appropriate than mTRE, as 2D-to-3D registration normally projects a 3D image onto a 2D image plane. Another solution is employing mean reprojection distance (mRPD), which calculates the minimum distance between the line (eg. centreline of CA) from 2D projected points to the corresponding 3D position of that point. Here, mRPD can be defined as,

\[
mRPD(\hat{T}) = \frac{1}{N_p} \sum_{i,j=1}^{N_p} \text{Dist}(L_i(\hat{T}(\text{source}, p_i)), p_j), \tag{6}\]

where \(\text{Dist}(L_i, p_j)\) denotes the minimum distance between the 3D point \(p_j\) and a line \(L_i\) through \(p_i\) and the source image (2D image) space. Besides, the Jacobian determinant can measure the plausibility of estimated displacement fields for non-rigid registration (Ashburner, 2007). The Jacobian matrix of a displacement field \(\Phi\) at a spatial point \(p\) is defined as,

\[
J_{\Phi}(p) = \nabla\Phi(p) \tag{7}
\]

where \(\Phi(p)\) denotes the displacement value at \(p\). Jacobian determinant \(\det(J_{\Phi}(p)) \leq 0\) indicates that \(\Phi(p)\) is not topology-preserving, while the values near 1 represent smooth fields (Dalca et al., 2019).

4.2.2. Image fusion measures

Although there are several multi-modality cardiac image fusion works, most of them are performed in clinical studies generally with visual evaluation (Chauhan et al., 2021). There are many available objective evaluation measures for image fusion, including entropy, MI, standard deviation (SD), peak signal to noise ratio (PSNR), structural similarity index measure, average gradient, Q:\(AB/\)f metric, root mean square error (RMSE), edge intensity, visual information fidelity, spatial frequency, and spectral discrepancy (Tirupal et al., 2021; Bhavana and Krishnappa, 2015). This qualitative analysis mainly aims to check the quality of fused images from different perspectives. However, specific to the cardiology field, the measure of image fusion is mainly performed at the registration stage instead of the fusion stage, as to our knowledge. This may be due, at least in part, to the limited methodology development of image fusion in cardiology.

4.2.3. Image segmentation measures

The most commonly used measure for multi-modality cardiac image segmentation is the Dice score. Jaccard index, HD, and average surface distance (ASD) have also been used (Li et al., 2021b,a; Zhuang, 2019), defined as:

\[
\text{Jaccard}(V_{\text{auto}}, V_{\text{manual}}) = \frac{|V_{\text{auto}} \cap V_{\text{manual}}|}{|V_{\text{auto}} \cup V_{\text{manual}}|}, \tag{8}\]

and

\[
\text{ASD}(X,Y) = \frac{1}{2} \left(\frac{\sum_{x \in X} \min_{y \in Y} d(x,y)}{\sum_{x \in X} 1} + \frac{\sum_{y \in Y} \min_{x \in X} d(x,y)}{\sum_{y \in Y} 1}\right) \tag{9}\]

where \(V_{\text{auto}}\) and \(V_{\text{manual}}\) denote the set of pixels in the automatic and manual segmentation, respectively. Statistical measurements can also be employed (Li et al., 2021b; Montazeri et al., 2017), i.e., Accuracy (Acc), Specificity (Spe), Sensitivity (Sen), and Precision (Pre). Acc refers to the proportion of pixels that have received the correct label among the total number of subjects examined. Spe and Sen (also known as Recall) are utilized to reflect the success rates of algorithms for the background and the foreground segmentation, respectively. Precision (also called positive predictive value) indicates the fraction of relevant instances among the retrieved instances. These measures can appear in various combinations when evaluating multi-modality cardiac image segmentation, and Dice is usually included (Cai et al., 2016; Ding et al., 2020).

5. Discussion and future perspectives

The last decade has witnessed an enormous amount of efforts in the specific field of multi-modality image computing for cardiac analysis. Cardiac hybrid imaging can provide valuable diagnostic and prognostic information for patients with CVDs, compared with side-by-side evaluation from a single imaging modality (Gimelli and Liga, 2013). Table 7 summarizes the potential clinical applications of the developed computing algorithms. With the development of hybrid imaging devices, there has been a
Table 7: Summary of potential clinical applications of combining multi-modality cardiac imaging for cardiac analysis. PAD: peripheral artery disease; ID: infectious endocarditis; DCM: dilated cardiomyopathy; HCM: hypertrophic cardiomyopathy; CS: cardiac sarcoidosis; DMR: degenerative mitral valve regurgitation; AS: aortic stenosis; MVD: mitral valve disease; AVS: aortic valve stenosis; TAVR: trans-aortic valve replacement.

| Modality       | Clinical application                                                                 | Pathology | Representative study                                                                 |
|---------------|--------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------|
| CT, MRI       | WH segmentation; LA cavity segmentation; LA wall thickness measurement; cardiac index estimation | AF, CHD, etc | Zhuang et al. (2019); Hobon-Gomez et al. (2015); Karim et al. (2018); Yu et al. (2020a) |
| CTA, MRI      | CA quantification                                                                   | CAD       | Sturm et al. (2003)                                                                  |
| CTA, SPECT    | Myo/ CA quantification                                                              | CAD, ID   | Koukouraki et al. (2013); Piccinelli and Garcia (2013); Sazonova et al. (2017)       |
| CTA, X-ray    | Angioplasty; TAVI guidance                                                          | CAD, AVS  | Baka et al. (2013); Gouveia et al. (2017)                                            |
| MRI, X-ray    | invasive cardiovascular procedures (such as CRT); LV contour detection; target endomyocardial injections | PAD, CHD, MyoI | Oost et al. (2003); Gutiérrez et al. (2007); Ma et al. (2010a); Faranesh et al. (2013) |
| CT, Echo      | LA wall thickness measurement; real-time surgical navigation; ablation/ TAVI/ TAVR guidance | AF, AS, MVD, CHD | Karim et al. (2018); Huang et al. (2009); Sanna-oval and Diffenseger (2013); Khalil et al. (2017a); Watanabe et al. (2021) |
| MRI, Echo     | cardiac motion estimation; Echo acquisition guidance; Myo viability assessment        | DCM, HCM, MyoI | Puuy-Anton et al. (2017); Atehortua et al. (2020)                                    |
| MRI, PET      | practical MRI acquisition; Myo viability, perfusion and metabolism assessment; cardiac motion correction | MyoI, CAD, CS | Kokilisch et al. (2017); Sinha et al. (1995); Mäkelä et al. (2001); Zandieh et al. (2018); Polygarpoz et al. (2021) |
| X-ray, Echo   | LAA closure procedure guidance; ablation and TAVI guidance                           | CHD, AF, SHD | Houwden et al. (2012); Hadeed et al. (2018); Ebel et al. (2020)                      |
| bSSFP, LGE, T2 weighted MRIs | LV (scar and edema), RV, Myo segmentation generate gadolinium-free LGE-like MRI | MyoI | Zhang et al. (2020); Li et al. (2022a); Wang et al. (2022) |
| bSSFP, T2 mapping MRIs | LA fibrosis and scar segmentation Myo strain analysis; motion estimation | AF | Li et al. (2020a); Wu et al. (2018) |
| tagged, cine MRIs | weighted MRIs

reduction in complex image processing through software. This is because the superposition of images using a hybrid device, although it may not be obtained simultaneously, is obtained by successive measurements. Nevertheless, many techniques have not yet been translated into clinical routine. This is, at least in part, attributed to the fact that the pre-processing step, i.e., registration, usually requires intensive manual interaction. In this discussion, we aim to identify the most important challenges that may drive future research in multi-modality cardiac image analysis.

5.1. Multi-modality cardiac computing with missing modality

In multi-modality cardiac studies, one or more sub-modalities may be missing due to poor image quality (e.g. imaging artifacts), failed acquisitions or counterindications. However, current multi-modality learning algorithms generally assume that all modalities are available (Azad et al., 2022). Nevertheless, there exist a number of methods to deal with missing modalities, which can be roughly categorized into three types. Firstly, different models can be designed for every potential missing modality combination, which is complicated and time-consuming. Secondly, many works attempt to impute or synthesize missing modalities to obtain the complete imaging modalities (Havaei et al., 2016). However, an additional network is normally required for synthesis, and the model might be sensitive to the synthesized image quality. Thirdly, one could learn a shared feature representation through all sub-modalities and then project it to a single model (Zhou et al., 2019a). The third solution is more efficient than the previous two approaches though it is difficult to learn a shared latent representation. The most straightforward way is employing a mean function to map all available sub-modalities into a unified representation, though this may lose important information (Lau et al., 2019). Feature disentanglement and domain adaptation can also be utilized to generate a shared feature representation for missing modalities (Chen et al., 2019a; Shen and Gao, 2019). However, current studies mainly focus on segmentation (especially brain tumor segmentation), while the issue of missing modalities occurring in multi-modality cardiac analysis has not been well investigated.

There exist several studies that attempt to find al-
ternative cardiac imaging modalities for missing modalities, namely modality replacement. Modality replacement aims to find “cheap” modalities (such as CTA and cine/T1 mapping MRIs) to replace the “expensive” ones (such as LGE MRI) by employing the inter-relationship among multi-modality images. For example, the LV scarring areas presented in LGE MRI are correlated to LV Myo thicknesses measured by CTA (Takigawa et al., 2019). Zhang et al. (2019) found that LV myocardial wall motion present in non-enhanced cine MRI could be used to predict chronic Myol areas. Therefore, one could employ CTA or cine MRI to detect scar regions instead of LGE MRI which is time-consuming to acquire and requires contrast agents (O’Brien et al., 2021). Also, native T1 mapping MRI can be utilized to replace LGE MRI as it exhibits sensitivity to a variety of cardiac diseases (Zhang et al., 2021a). So far, only a few modality replacement works (Zhang et al., 2021a; Takigawa et al. 2019) have been proposed and are mainly devoted to LV Myo pathology analysis. It would be desirable to continue the study of modality relationships and develop modality replacement technologies for other cardiac substructure analyses.

5.2. Integration of cardiac imaging with non-imaging information

In addition to multi-modality images, there are many available non-imaging information sources associated with CVDs (Bai et al., 2020; Meyer et al., 2020). Take UK Biobank data as an example, which provides gender, age, BMI, biological information, heart and lung function measurements, etc (Bycroft et al., 2018). Among these non-imaging modalities, ECGs can provide a substantial amount of information through its indirect recording of the electrical activity of the heart. This includes scar regions after Myol, different arrhythmias, the effects of hypertension, as well as information about recovery and cardiac motion. However, its interpretation requires considerable human expertise, and ECGs are limited in their ability to spatially locate and characterize CVDs. The structural information from cardiac imaging data may be complementary to the information provided by ECGs. Body surface ECG mapping data, which provides a larger amount of measurements and includes more spatial information, may be preferable for this purpose (Cochet et al., 2014; Pereira et al., 2020). As presented in Fig. 14 (a), one could combine cardiac geometry obtained from CT/MRI with body surface potentials for a more comprehensive noninvasive assessment of cardiac arrhythmias. Specifically, the combined multi-modality information might be able to identify the mechanism of the arrhythmia and locate the circuit or source based on cardiac anatomy and myocardial substrate (Cochet et al., 2014). It could then be embedded into a 3D mapping system to assist in catheter ablation therapy for atrial or ventricular arrhythmias. Electroanatomical mapping (EAM) can also be integrated with cardiac imaging, such as CT (Itoh et al., 2010), MRI (Reddy et al., 2004), X-ray (Scaglione et al., 2011), to guide ablation procedures (see Fig. 14 (b)). Wang et al. (2018) combined cine MRI with intra-ventricular pressure recording data to estimate diastolic myocardial stiffness and stress for personalized biomechanical analysis. Non-imaging information can also be integrated to diagnose cardiac abnormalities, for example, combining ECG and phonocardiogram signals (Chakir et al., 2020; Li et al., 2020d).

The combination of imaging and non-imaging data will provide a transition from image space to patient-level information. However, it could be more challenging than the integration of different imaging modalities. This is because it brings a data format variation, resulting in the requirement of intermediate conversion steps for each data. For example, cardiac imaging data usually needs to be converted into a surface, while ECG normally needs to be recorded on a body surface via a multi-electrode vest. It is necessary to develop algorithmic techniques to learn high-level features from two types of data sources for a more elegant combination (Bacoyannis et al., 2021).

5.3. Unified framework for multi-center multi-modality cardiac image analysis

Current multi-modality image processing steps do not include much user interaction, resulting in the limited ability to adjust upstream processing steps based on downstream requirements. Furthermore, these processing stages are not optimized for a clinical endpoint. We therefore expect a unified framework for joint optimization of the whole pipeline with respect to a clinical endpoint, as shown in Fig. 15. For example, Pan et al. (2021) recently developed a joint image synthesis and disease diagnosis framework using incomplete multi-modality neuroimages. This study emphasized the importance of performing a disease-image-oriented joint optimization for disease identification and prediction.

Due to different clinical acquisition protocols across hospitals, acquiring all modalities from all centers is costly and often impossible in clinical settings. Federated learning (also referred to as collaborative learning) is a promising technique that can train models on multiple distributed centers that hold local data samples (Li et al., 2020c). It ensures data sharing while protecting patient data privacy, which would increase the number of available cardiac imaging with different modalities. However, different centers may own misaligned imaging modalities, which raises a realistic challenge for federated learning. Recently, Chang et al. (2022) designed a privacy secure decentralized multimodality adaptive learning architecture, namely Modality-Bank. Modality-Bank consists of a domain-specific modulation parameters bank, a central generator, and multiple distributed discriminators located in a variety of medical centers. It can synthesize multi-modality images by switching different sets of configurations to complete missing modalities across data centers. Nevertheless, the domain generalization ability of current DL-based models needs to be significantly improved to handle multi-center multi-modality cardiac datasets (Li et al., 2021c). In the
Figure 14: Two examples of combining cardiac imaging with non-imaging datasets. (a) The procedure of electrocardiographic imaging (ECGi), which combines CT/MRI with ECG signal using mathematical algorithms. Here, body surface potentials are recorded from several electrodes, while the patient-specific heart-torso geometry is obtained from thoracic CT or MRI. Image adapted from Pereira et al. (2020) with permission; (b) The procedure of combining LGE MRI and EAM data for LA scar localization and visualization. Image adapted from Nuñez García et al. (2018) with permission.

Figure 15: Traditional vs. future expected multi-center multi-modality cardiac image analysis (MM-CIA) procedures. Conventional MM-CIA is usually only optimized in a single direction and usually not for a clinical endpoint, while unified MM-CIA can transfer information in both directions. Imaging-free MM-CIA even can directly achieve CVD diagnosis from imaging acquisition in the absence of actual imaging.

Future, with deep neural networks we even could achieve direct CVD diagnosis from imaging acquisition without the requirement of imaging, which however could be too ambitious for now.

5.4. Polar map of multi-modality cardiac imaging

The visualization of multi-modality images requires a common coordinate system, where the spatial information about cardiac anatomy and physiology from different modalities can be combined. One can utilize a unified 3D cardiac visualization, but it requires viewing rotated images or a rotating display. In contrast, a 2D polar map can display the entire cardiac surface and the coronary anatomy in a single image, even though it comes at the expense of undesired geometric distortions. The 2D polar map offers cardiac regional features, permits a comparison of spatial data across patients, and allows a standardized analysis across different imaging modalities. Therefore, it is promising for multi-modality cardiac image analysis and could be useful in clinical practice.

The most popular application of polar maps to the heart is the shaped LV, which is often mapped into a planar bull’s-eye plot (Cerqueira et al., 2002). In the bull’s-eye polar map, the LV can be divided into different segments (AHA-16, 17 or 18) for simplification and standardization (Lang et al., 2015). Fig. 16 (a) presents the most common AHA-17 map, which can be used to visualize regional function information about the LV, including wall motion, perfusion and distribution of scars and edema. Bull’s eye maps have also been applied to visualize the CA (Nakahara et al., 2016; Fukushima et al., 2020) and coronary sinus (Ma et al., 2012). Specific to multi-modality image analysis, Nakahara et al. (2016) proposed a fusion based bull’s eye map to describe the SPECT–coronary CTA data in a single image. Tavard et al. (2014) employed a bull’s eye map to visualize the fused anatomical, functional and electrical data acquired from CT, speckle tracking Echo, and electro-anatomical mappings (EAM). Pann et al. (2017); Bayer et al. (2018); Schuler et al. (2021) provided more general ventricle representations that included both LV and RV.

Compared to the ventricles, it is more challenging to generate a standard planar mapping for the atria which usually have more complex shapes (Roney et al., 2019; Li et al., 2022). There are only a few studies targeting atrial polar mapping (Williams et al., 2017; Roney et al., 2019; Nuñez-Garcia et al., 2019, 2020). Their templates were all designed for the most common LA topology with four pulmonary veins and are sensitive to LA topological
variants. Fig. 16 (b) presents a universal atrial coordinate mapping system for 2D visualization of both the LA and right atrium. For unified multi-modality image visualization and analysis, we expect more comprehensive and robust cardiac polar maps, which can cover more substructures of the heart and adapt for cardiac variations across patients.

6. Conclusion

The development of cardiac multi-modality imaging has already refined clinical decision-making and improved the treatment of patients with CVD. Many studies have been proposed to integrate multiple imaging modalities, to facilitate our understanding of the complex anatomy of the heart and its behavior and a move to more patient-specific interpretation. This review summarizes recent advances in multi-modality cardiac imaging techniques, computing algorithms, public datasets, evaluation measures, clinical applications and future perspectives. We conclude that although the development of machine learning, especially deep learning, has promoted multi-modality cardiac image analysis, there are still many unresolved issues. Note that although the review focuses on the multi-modality cardiac image computing, the key issues and obstacles of the computing techniques can be generalized to other medical image analysis applications and utilized in mainstream clinical practice.

Acknowledgment

This work was funded by the CompBioMed 2 Centre of Excellence in Computational Biomedicine (European Commission Horizon 2020 research and innovation programme, grant agreement No. 823712). X. Zhuang was supported by the National Natural Science Foundation of China (grant no. 61971142, 62111530195 and 62011540404) and the development fund for Shanghai talents (no. 20200015). L. Li was partially supported by the SJTU 2021 Outstanding Doctoral Graduate Development Scholarship.

References

Abu Hazem, A.A., Dori, Y., Whitehead, K.K., Harris, M.A., Fogel, M.A., Gillespie, M.J., Rome, J.J., Glatz, A.C., 2014. X-ray magnetic resonance fusion modality may reduce radiation exposure and contrast dose in diagnostic cardiac catheterization of congenital heart disease. Catheterization and Cardiovascular Interventions 84, 795–800.

Aguadé-Bruix, S., Pizzi, M., Roque, A., Cababria, H.C., Fernández-Hidalgo, N., Alujas, M.G., 2017. Diagnostic value of 18F-FDG PET/cardiac CT in late prosthetic aortic endocarditis with periprosthetic abscess. Rev Esp Med Nucl Imagen Mol 36, 59–60.

Aksoy, T., Unal, G., Demirci, S., Navab, N., Degertekin, M., 2013. Template-based CTA to x-ray angiographic registration of coronary arteries in frequency domain with automatic x-ray registration. Medical physics 40, 101903.

Alessio, A.M., Bindschadler, M., Busey, J.M., Shuman, W.P., Caldwell, J.J., Branch, K.R., 2019. Accuracy of myocardial blood flow estimation from dynamic contrast-enhanced cardiac CT compared with PET. Circulation: Cardiovascular Imaging 12, e008323.

Ammari, A., Mahmoudi, R., Himida, B., Saouli, R., Bedoui, M.H., 2021. A review of approaches investigated for right ventricular segmentation using short-axis cardiac MRI. IET Image Processing 8, 557–569.

Angelidis, G., Giamouzis, G., Karagiannis, G., Butler, J., Tsougos, I., Valotassios, V., Giannakoulas, G., Dimakopoulos, N., Xanthopoulos, A., Skoularigis, J., et al., 2017. SPECT and PET in ischemic heart failure. Heart failure reviews 22, 243–261.

Ankenbrand, M.J., Lühr, D., Schreiber, L.M., 2020. Exploring ensemble applications for multi-sequence myocardial pathology segmentation, in: Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge, Springer. pp. 60–67.

Arar, M., Ginger, Y., Danon, D., Bermano, A.H., Cohen-Or, D., 2020. Unsupervised multi-modal image registration via geometry preserving image-to-image translation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 13410–13419.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Atelhoťa, A., Garreau, M., Simon, A., Donal, E., Lederlin, M., Romero, E., 2020. Fusion of 3D real-time echocardiography and cine MRI using a saliency analysis. International journal of computer assisted radiology and surgery 15, 277–285.

Azad, R., Khosravi, N., Dehghananshadi, M., Cohen-Adad, J., Merhof, D., 2022. Medical image segmentation on MRI images with missing modalities: A review. arXiv preprint arXiv:2203.06217.

Bacoyannis, T., Ly, B., Cedilnik, N., Cochet, H., Sermeusant, M., 2021. Deep learning formulation of electrocardiographic imaging integrating image and signal information with data-driven regularization. EP Europace 23, 155–162.

Bai, W., Shi, W., O’Regan, D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D., 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration.
Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1999. Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging 16, 187–198.

Maffessanti, F., Patel, A.R., Patel, M.B., Walter, J.J., Medratta, A., Medvedovsky, D., Kachenoura, N., Lang, R.M., Mor-Avi, V., 2017. Non-invasive assessment of the haemodynamic significance of coronary stenosis using fusion of cardiac computed tomography and 3D echocardiography. European Heart Journal-Cardiovascular Imaging 18, 670–680.

Mäkelä, T., Clarysse, P., Löjtjönen, J., Sipiä, O., Laurerka, K., Hänninen, H., Pyökkönen, E.P., Nenonen, J., Knutti, J., Katila, T., et al., 2001. A new method for the registration of cardiac PET and MR images using deformable model based segmentation of the main thorax structures, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 557–564.

Mäkelä, T., Clarysse, P., Sipiä, O., Pauna, N., Pham, Q.C., Katila, T., Magnin, I.F., 2002. A review of cardiac image registration methods. IEEE Transactions on medical imaging 21, 1011–1021.

Martin, R., Hascœ̈t, S., Dulac, Y., Peyre, M., Mejean, S., Hadeed, F., Caravet, A., Lebon, B., Acar, P., 2013. Comparison of two- and three-dimensional transthoracic echocardiographic measurement of aortic annulus diameter in children. Archives of cardiovascular diseases 106, 492–500.

Martín-Isla, C., Asadi-Aghbolaghí, M., Glontra, P., Campello, V.M., Escalera, S., Lekadir, K., 2020. Stacked BCNU-Net with semantic CMR synthesis: Application to myocardial pathology segmentation challenge, in: Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge, Springer. pp. 1–16.

Martinez-Möller, A., Sovatzoglou, M., Navah, N., Schwagier, M., Nekolla, S.G., 2007. Artifacts from misaligned CT in cardiac perfusion PET/CT studies: frequency, effects, and potential solutions. Journal of Nuclear Medicine 48, 188–193.

McGuirt, D., Mazal, J., Rogers, T., Faranesh, A.Z., Schenke, W., Grant, L., Lederman, R.J., 2016. X-ray fused with magnetic resonance imaging to guide endomyocardial biopsy of a right ventricular mass. Radiologic technology 87, 622–626.

Meyer, H.V., Dawes, T.J., Serrani, M., Bai, W., Tokarczuk, P., Cai, J., 2012. Noninvasive assessment of right ventricular mass. Radiologic technology 87, 622–626.

Mikes, J., Makaryus, A.N., Redberg, R., Shaw, L., 2007. Noninvasive cardiac imaging. American family physician 75, 1219–1228.

Ming, C.T., Omar, Z., Mahmood, N.H., Kadiman, S., 2015. A literature survey of ultrasound and computed tomography-based cardiac image registration. Jurnal Teknologi 74.

Mortazi, A., Burt, J., Bagci, U., 2013. Multi-planar deep segmentation networks for cardiac substructures from MRI and CT, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer. pp. 199–206.

Mylonas, I., Alam, M., Amuly, N., Small, G., Chen, L., Yan, V., Hibbert, B., Chow, B.J., 2014. Quantifying coronary artery calcification from a contrast-enhanced cardiac computed tomography angiography study. European Heart Journal–Cardiovascular Imaging 15, 210–215.

Nakahara, T., Iwabuchi, Y., Murakami, K., 2016. Diagnostic performance of 3D virtual eye display of SPECT and coronary CTA fusion. JACC: Cardiovascular Imaging 9, 705–711.

Nens, F., Kloth, J., Tegzah, E., Puepel, T.D., Heusch, P., Goebel, J., Nassenstein, K., Schlosser, T., 2018. Feasibility of FDG-PET in myocarditis: comparison to CMR using integrated PET/MRI. Journal of Nuclear Cardiology 25, 785–794.
18F-FDG-PET/CT imaging. International journal of cardiology 212, 94–96.

Rischpler, C., Nekolla, S.G., Dregely, I., Schwaiger, M., 2013. Hybrid PET/MR imaging of the heart: potential, initial experiences, and future prospects. Journal of Nuclear Medicine 54, 402–415.

Roche, A., Pennec, X., Malandain, G., Ayache, N., 2001. Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information. IEEE transactions on medical imaging 20, 1038–1049.

Rokach, L., 2010. Ensemble-based classifiers. Artificial intelligence review 33, 1–39.

Roney, C.H., Pashaei, A., Meo, M., Dubois, R., Boyle, P.M., Trayanova, N.A., Cohet, H., Niederer, S.A., Vigmund, E.J., 2019. Universal atrial coordinates applied to visualisation, registration and construction of patient specific meshes. Medical image analysis 55, 65–75.

Sandoval, Z.L., Dillenseger, J.L., 2013. Evaluation of computed tomography to ultrasound 2D image registration for atrial fibrillation treatment, in: Computing in Cardiology 2013. IEEE. pp. 245–248.

Sang, Y., Cao, M., Mcnett-Gray, M., Gao, Y., Hu, P., Yan, R., Yang, Y., Ruan, D., 2021. Inter-phase 4D cardiac MRI registration with a motion prior derived from CTA. IEEE Transactions on Biomedical Engineering.

Sanroma, G., Benkarim, O.M., Piella, G., Camara, O., Wu, G., Shen, D., Gispert, J.D., Molinuevo, J.L., Ballester, M.A.G., Initiative, A.D.N., et al., 2018. Learning non-linear patch embeddings with neural networks for label fusion. Medical image analysis 44, 143–155.

Sanzarelli, P., Benassai, P., Marcheschi, L., Landini, L., Marzullo, P., Benassi, A., 2001. Multimodal cardiac image fusion by geometrical features registration and warping, in: Computers in Cardiology 2001. Vol. 28 (Cat. No. 01CH37287), IEEE. pp. 277–280.

Savi, A., Gilardi, M.C., Rizzo, G., Pepi, M., Landoni, C., Rossetti, C., Lucignani, G., Bartorelli, A., Fazio, F., 1995. Spatial registration of echocardiographic and positron emission tomographic heart studies. European journal of nuclear medicine 22, 243–247.

Sazonova, S.I., Nikolaeva Iluyushenkov, J., Valevich Zavadovskiy, K., Borisovich Lishmanov, Y., 2017. 99mTc-hmpao-labeled autologous leukocyte SPECT/CT for diagnosis of bacterial endocarditis of the prosthetic pulmonary conduit: A clinical case. Iranian Journal of Radiology 14.

Scaglione, M., Biasco, L., Caponi, D., Anselmino, M., Negro, A., Di Donna, P., Corleto, A., Montefusco, A., Gaita, F., 2011. Visualization of multiple catheters with electroanatomical mapping reduces X-ray exposure during atrial fibrillation ablation. Europace 13, 955–962.

Schuler, S., Pavlo, N., Potyagaylo, D., Loewe, A., 2021. Cobiveco: Consistent biventricular coordinates for precise and intuitive description of position in the heart—with matlab implementation. Medical Image Analysis 74, 102247.

Scott, K., Stuart, D., Peoples, J.J., Bisleri, G., Ellis, R.E., 2021. Efficient automatic 2D/3D registration of cardiac ultrasound and CT images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 9, 438–446.

Shade, J.K., Prakosa, A., Popescu, D.M., Yu, R., Okada, D.R., Chrispin, J., Trayanova, N.A., 2021. Predicting risk of sudden cardiac death in patients with cardiac sarcoidosis using multimodality imaging and personalized heart modeling in a multivariable classifier. Science Advances 7, eaba0202.

Shen, Y., Gao, M., 2019. Brain tumor segmentation on MRI with missing modalities, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 417–428.

Shi, W., Zhuang, X., Wang, H., Duckett, S., Luong, D.V., Tobon-Gomez, C., Tung, K., Edwards, P.J., Rhode, K.S., Razavi, R.S., et al., 2012. A comprehensive cardiac motion estimation framework using both untagged and 3-D tagged MR images based on nonrigid registration. IEEE transactions on medical imaging 31, 1263–1275.

Siebermair, J., Kholmovski, E.G., Marrouche, N., 2017. Assessment of left atrial fibrosis by late gadolinium enhancement magnetic resonance imaging: methodology and clinical implications. JACC: Clinical Electrophysiology 3, 791–802.

Siebermair, J., Lehner, S., Sattler, S.M., Rizas, K.D., Beckmann, B.M., Becker, A., Schiller, J., Metz, C., Zacherl, M., Wonderlin, N., et al., 2020. Left-ventricular innervation assessed by 123I-SPECT/CT is associated with cardiac events in inherited arrhythmia syndromes. International journal of cardiology 312, 129–135.

Sinha, S., Sinha, U., Czornin, J., Porenta, G., Schelbert, H., 1995. Noninvasive assessment of myocardial perfusion and metabolism: feasibility of registering gated MR and PET images. AJR. American journal of roentgenology 164, 301–307.

Slomka, P.J., Baam, R.P., 2009. Multimodality image registration with software: state-of-the-art. European journal of nuclear medicine and molecular imaging 36, 44–55.

Smith, M.F., Gilbert-Serra, X., Arrate, F., Chellappa, R., Chodnicki, K., Tian, J., Dickfeld, T., Dilisizian, V., 2014. CardioViewer: a novel modular software tool for integrating cardiac electrophysiology voltage measurements and PET/SPECT data, in: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE. pp. 1–3.

Spagnolo, P., Giglio, M., Di Marco, D., Cannao, P.M., Della Bella, P.E., Monti, C.B., Sardanelli, F., et al., 2021. Diagnosis of left atrial appendage thrombus in patients with atrial fibrillation: delayed contrast-enhanced cardiac CT. European Radiology 31, 1236–1244.

Sturm, B., Powell, K.A., Stillman, A.E., White, R.D., 2003. Registration of 3D CT angiography and cardiac MR images in coronary artery disease patients. The International journal of cardiovascular imaging 19, 281–293.

Takaya, Y., Ito, H., 2020. New horizon of fusion imaging using echocardiography: its progress in the diagnosis and treatment of cardiovascular disease. Journal of echocardiography 18, 9–15.

Takigawa, M., Martin, R., Chentli, G., Kitamura, T., Vlachos, K., Frontera, A., Martin, C.A., Courier, F., Lam, A., Pillois, X., et al., 2019. Detailed comparison between the wall thickness and voltages in chronic myocardial infarction. Journal of Cardiovascular Electrophysiology 30, 195–204.

Tao, X., Wei, H., Xue, W., Ni, D., 2019. Segmentation of multimodal myocardial images using shape-transfer GAN, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer. pp. 271–279.

Tavakoli, V., Amini, A.A., 2013. A survey of shaped-based registration and segmentation techniques for cardiac images. Computer Vision and Image Understanding 117, 966–989.

Tavard, F., Simon, A., Di Marco, D., Donal, E., Hernández, A.I., Garreau, M., 2014. Multimodal registration and data fusion for cardiac resynchronization therapy optimization. IEEE transactions on medical imaging 33, 1363–1372.

Tayebi, R.M., Wirza, R., Sulaiman, P.S., Dimon, M.Z., Khalid, P., Al-Surmi, A., Mazaheri, S., 2015. 3D multimodal cardiac data reconstruction using angiography and computerized tomographic angiography registration. Journal of cardiothoracic surgery 10, 1–25.

Tirupal, T., Mohan, B.C., Kumar, S.S., 2021. Multimodal medical image fusion techniques—a review. Current Signal Transduction Therapy 16, 142–161.

Tobon-Gomez, C., De Craene, M., Meleod, K., Tautz, L., Shi, W., Hennemuth, A., Prakosa, A., Wang, H., Carr-White, G., Kapetanakis, S., et al., 2013. Benchmarking framework for myocardial tracking and deformation algorithms: An open access database. Medical image analysis 17, 632–648.

Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., Pinto, K., Patel, R., Ammar, M., Daoudi, A., Margeta, J., Sandoval, Z., 2015. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. IEEE transactions on medical imaging 34, 1460–1473.

Tomar, D., Lortkipanidze, M., Vray, G., Bozortgarba, B., Thiran, J.P., 2021. Self-attentive spatial adaptive normalization for cross-
Burrage, M.K., Lukaschuk, E., Shanmuganathan, M., Popescu, I.A., Nikolaidou, C., Mills, R., Werys, K., Hann, E., Barutcu, A., et al., 2021a. Toward replacing late gadolinium enhancement with artificial intelligence virtual native enhancement for gadolinium-free cardiovascular magnetic resonance tissue characterization in hypertrophic cardiomyopathy. Circulation 144, 589–599.

Zhang, W., Noble, J.A., Brady, J.M., 2006. Real time 3-D ultrasound to MR cardiovascular image registration using a phase-based approach, in: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006., IEEE. pp. 666–669.

Zhang, X., Noga, M., Panithakumar, K., 2020. Fully automated deep learning based segmentation of normal, infarcted and edema regions from multiple cardiac MRI sequences, in: Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge, Springer. pp. 82–91.

Zhang, Y., Sidibé, D., Morel, O., Moriaudeau, F., 2021b. Deep multimodal fusion for semantic image segmentation: A survey. Image and Vision Computing 105, 104042.

Zhao, Z., Boutry, N., Puybareau, É., 2020. Stacked and parallel U-nets with multi-output for myocardial pathology segmentation, in: Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge, Springer. pp. 138–145.

Zheng, R., Zhao, X., Shen, J., 2019. Deep learning based multi-modal cardiac MR image segmentation, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer. pp. 263–270.

Zhou, C., Chan, H.P., Chughtai, A., Patel, S., Hadjiiski, L.M., Wei, J., Kazerooni, E.A., 2012. Automated coronary artery tree extraction in coronary CT angiography using a multiscale enhancement and dynamic balloon tracking (MSCAR-DBT) method. Computerized Medical Imaging and Graphics 36, 1–10.

Zhuang, X., Shen, J., 2016. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical image analysis 31, 77–87.

Zhang, X., Xu, J., Luo, X., Chen, C., Ouyang, C., Rueckert, D., Campello, V.M., Lekadir, K., Vesal, S., Ravikumar, N., et al., 2020. Cardiac segmentation on late gadolinium enhancement MRI: a benchmark study from multi-sequence cardiac MR segmentation challenge. arXiv preprint arXiv:2006.12434.

Zorinas, A., Janusauskas, V., Davidavicius, G., Puodziukaite, L., Zakarkaite, D., Kramena, R., Cypienė, R., Bilikis, V., Rucinskas, K., Ašliditis, A., et al., 2017. Fusion of real-time 3D transesophageal echocardiography and cardiac fluoroscopy imaging in transapical catheter-based mitral paravalvular leak closure. Advances in Interventional Cardiology 13, 263.