

Adaptive Conditional Contrast-Agnostic Deformable Image Registration with Uncertainty Estimation

Yinsong Wang, Xinzhe Luo, Siyi Du and Chen Qin

Abstract— Deformable multi-contrast image registration is a challenging yet crucial task due to the complex, non-linear intensity relationships across different imaging contrasts. Conventional registration methods typically rely on iterative optimization of the deformation field, which is time-consuming. Although recent learning-based approaches enable fast and accurate registration during inference, their generalizability remains limited to the specific contrasts observed during training. In this work, we propose an adaptive conditional contrast-agnostic deformable image registration framework (AC-CAR) based on a random convolution-based contrast augmentation scheme. AC-CAR can generalize to arbitrary imaging contrasts without observing them during training. To encourage contrast-invariant feature learning, we propose an adaptive conditional feature modulator (ACFM) that adaptively modulates the features and the contrast-invariant latent regularization to enforce the consistency of the learned feature across different imaging contrasts. Additionally, we enable our framework to provide contrast-agnostic registration uncertainty by integrating a variance network that leverages the contrast-agnostic registration encoder to improve the trustworthiness and reliability of AC-CAR. Experimental results demonstrate that AC-CAR outperforms baseline methods in registration accuracy and exhibits superior generalization to unseen imaging contrasts. Code is available at <https://github.com/Yinsong0510/AC-CAR>.

Index Terms— Adaptive Conditional Feature Modulation, Latent Space Regularization, Contrast-Agnostic Image Registration (MRI), Uncertainty Estimation.

I. INTRODUCTION

MULTI-CONTRAST deformable image registration establishes dense spatial correspondences that align images with different imaging contrasts, which is crucial for downstream multi-contrast analysis and interpretation. Precise anatomical alignment between multi-contrast images provides complementary information for characterizing tissues of human bodies, which are widely used in both qualitative and quantitative imaging in clinical diagnosis [1]. However, due to the complex and non-linear relationship between their intensity distributions, registering images across imaging contrasts is extremely challenging.

Manuscript submitted for review on 2 April 2025. This work was partially supported by the Engineering and Physical Sciences Research Council under Grant EP/Y002016/1 and EP/X039277/1.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (e-mail: y.wang23@imperial.ac.uk; x.luo@imperial.ac.uk; s.du23@imperial.ac.uk; c.qin15@imperial.ac.uk).

Although there are many conventional approaches [2] on deformable image registration based on iterative optimization, most of them focus on mono-contrast scenarios and adopt intensity-based similarity measures like Mean Squared Error (MSE), which are not feasible in multi-contrast registration scenarios. To tackle this, previous multi-contrast registration approaches have leveraged information-theoretic similarity measures like mutual information [3], [4], or structural feature descriptors like MIND [5], [6] on misaligned image pairs to guide the registration.

However, the iterative-based optimization framework is usually time-consuming and therefore could be limited in real-time applications. Recently, deep learning-based registration approaches have shown their great potential for fast and accurate registration in both mono-contrast scenario [7], [8] and multi-contrast scenario [9], [10]. For learning-based multi-contrast registration, MIDIR [11] and ContraReg [10] managed to embed information-theoretic similarity measures or their lower bound into deep neural networks. Another type of approach, like UMDIR [9] and Arar et al. [12] proposed to reduce the multi-contrast registration problem into a mono-contrast one based on image-to-image translation. Nevertheless, these approaches can only deal with fixed contrasts that have seen during training, resulting in poor generalizability.

Some recent work attempted to address the contrast variations between training and inference. One kind of such work [13] tried to achieve contrast-agnostic registration by synthesizing images of arbitrary contrasts based on randomly generated labels or segmentation maps. However, it relies on the availability of segmentation maps or the quality of generated labels. Alternatively, several recent works proposed to train a neural network to learn a certain distance metric for multi-contrast registration [14], [15]. Nevertheless, this kind of approach focused on similarity approximation without explicitly enforcing contrast-agnostic learning.

In this work, we propose an adaptive conditional contrast-agnostic deformable image registration framework (AC-CAR) that requires only single contrast images for training but can be generalized to unseen contrasts during inference. Specifically, we propose a random convolution-based contrast augmentation scheme to generate arbitrary contrast images as input during training from single-contrast images. To learn contrast-invariant feature representations from the arbitrary contrast images generated, we propose an adaptive conditional feature modulator (ACFM). ACFM utilizes low-frequency

image information to adaptively modulate feature representation to inversely remove contrast-related information. To further enhance such modulation to capture contrast-invariant features, we propose contrast-invariant latent regularization via a contrast-invariance loss. This regularizes the latent space to enforce consistent feature representations of the images with similar structures but varying imaging contrasts. In addition, AC-CAR can rely on simple mono-contrast similarity losses, which enable network training to be more efficient and reliable. To equip AC-CAR with the capability to estimate contrast-agnostic uncertainty, we further integrate a separate contrast-agnostic variance network to estimate the variance map as the registration uncertainty for multi-contrast images. To the best of our knowledge, this has not been explored before, whereas it can be crucial for informing reliability and for further downstream analysis. We validate AC-CAR on T1 and T2-weighted brain MRI and cardiac T1 mapping data with different T1 weightings, which show that AC-CAR outperforms existing SOTA methods across various contrasts, even when the model has not seen most of those contrasts during training. Specifically, our contribution can be summarized as follows:

- 1) We propose a new adaptive conditional contrast-agnostic registration framework based on contrast augmentation that can generalize beyond the contrasts seen during training.
- 2) We propose a novel adaptive conditional feature modulator to adaptively modulate the features of varying contrasts coupled with a contrast invariance loss to learn contrast-invariant features.
- 3) We equip the proposed registration framework with a contrast-agnostic uncertainty estimation mechanism to provide registration uncertainty.
- 4) Experiments show that AC-CAR outperforms the state-of-the-art baseline methods in terms of registration accuracy while providing meaningful uncertainty estimation across various contrasts.

II. RELATED WORKS

A. Conventional Multi-contrast Image Registration

Conventional multi-contrast deformable image registration approaches optimize the deformation field iteratively based on certain energy functions for each image pair. The problem can be formulated as follows:

$$\phi^* = \underset{\phi}{\operatorname{argmin}} \{ \mathcal{L}_{\text{sim}}(I_m \circ \phi, I_f) + \lambda \mathcal{L}_{\text{reg}}(\phi) \}, \quad (1)$$

where I_m and I_f denote the moving and fixed images defined over an n-D spatial domain $\Omega \in \mathbb{R}^n$. ϕ^* is the desired optimal deformation field. The objective function consists of a similarity term \mathcal{L}_{sim} and a regularization term \mathcal{L}_{reg} , where λ controls the trade-off between the two terms. However, intensity-based similarity measures like Mean Squared Error (MSE) often fail in multi-contrast registration due to the intensity matching criteria being violated by the distribution shifts across imaging contrast. Local Normalized Cross Correlation (LNCC) performs better due to it relaxes the matching criteria

to linear correlation. To better deal with multi-contrast images, Maes et al. [3], [4] proposed to use mutual information (MI) as the similarity to measure the statistical co-occurrence of image intensity values that is robust across different contrasts. However, MI is not sensitive to local variations as it focuses on histogram matching and lacks spatial information. Previous works tried to mitigate this by incorporating high-order MI [16] or combining spatial information [17], [18] within MI. Alternatively, Heinrich et al. [5] utilized the concept of self-similarity by proposing a Modality Invariant Neighborhood Descriptor (MIND) to extract distinctive structural features in a local neighborhood that are preserved across modalities. However, such structural descriptors are still an approximation of image similarities and may not be discriminative enough for subtle anatomical structures [19].

B. Learning-based Multi-Contrast Image Registration

Recently, deep learning has revolutionized medical image registration for its efficiency and precision. Deep learning-based image registration approaches estimate the deformation field ϕ through a neural network parameterized by θ , i.e., $\phi(\theta) = f_\theta(I_m, I_f)$. Instead of optimizing the deformation field directly, learning-based methods optimize the network parameters. After training, the network can estimate the deformation field with a single forward pass at the inference stage. Many deep learning approaches have adapted conventional multi-contrast metrics to learning-based frameworks. For example, Qiu et al. [11] extended conventional MI using a differentiable Parzen window [3] for network backpropagation.

In contrast, other works addressed multi-contrast registration from perspectives beyond conventional loss. For instance, Dey et al. [10] employed contrastive learning with a pre-trained autoencoder to learn an image distance between warped moving and fixed images using a multi-scale PatchNCE loss, which can be regarded as a lower bound of MI and thus inherits its limitations [19]. Alternatively, some researchers tried to use image translation-based methods to reduce the multi-contrast registration problem to a mono-contrast one. Qin et al. [9] proposed to use a GAN-based image-to-image translation framework to disentangle multi-contrast images into shape and appearance. The translated image can be reconstructed by the shape code of the source image and the appearance code of the target image. The translated image thus has the same contrast as the target image, reducing the multi-contrast registration to a mono-contrast one. Similarly, Arar et al. [12] proposed a geometry-preserving image translation network. They directly learned the translated image with L1 loss without disentanglement and iteratively exchanged the order of registration and translation during training to encourage geometry preservation. Deng et al. [20] disentangled multi-contrast images into features responsible for alignment (RA features) and not responsible for alignment (nRA features) by convolutional sparse encoding and used RA features solely for registration. However, all of the approaches mentioned above can only deal with fixed imaging contrasts, which means they can not generalize to contrasts not seen at the training stage.

C. Contrast-Agnostic Image Registration

To enable the registration network to be more generalizable to unseen imaging contrasts. Hoffmann et al. [13] proposed synthesizing arbitrary contrast of images based on segmentation or randomly generated labels and optimized registration network by Dice loss [21]. However, the employed Dice loss mainly focuses on the structure overlaps rather than pixel-level or volume-level differences. Several recent works proposed training a neural network to learn a certain distance metric for multi-modal registration. Sideri et al. [14] relied on modality and affine geometric augmentation for learning a contrast-agnostic distance measure based on image patch centers to tackle multi-contrast rigid registration. Ronchetti et al. [15] proposed to use a small CNN to learn a feature-level distance that approximates the conventional Linear Correlation of Linear Combination (LC²) similarity. Mok et al. [19] proposed to learn a modality-agnostic structural representation by adapting the structural descriptor [5] into a CNN. The learned structural representations were then used as the input of the registration network. However, this approach requires training two sub-networks, which demands high computational costs and may impede the registration network as its performance highly depends on how well the structural representations are learned.

Of particular relevance to this work, our earlier work [22] proposed a contrast-agnostic registration framework by simulating arbitrary contrast for training and regularizing the feature in latent space. In this work, to better handle the variability among different contrasts, we extend previous approaches by introducing an adaptive conditional feature modulator to adaptively modulate the features of varying contrast to capture contrast-invariant features. Additionally, a contrast-agnostic uncertainty estimation framework is introduced to equip the model with the capability to estimate uncertainty during registration. More comprehensive quantitative and qualitative evaluations of AC-CAR, including comparisons, generalization, and ablation studies, were conducted on a 3D brain MRI dataset and a 2D cardiac MRI dataset.

D. Uncertainty Estimation in Image Registration

Precise quantification of registration uncertainty is essential to mitigate the risk of misdiagnosis resulting from erroneous image alignments. Conventional methods [23], [24] employed probabilistic image registration (PIR) to estimate distributions of transformation parameters. The registration uncertainty can be derived from the variance of distributions. However, this approach only reflects variability in the mean model prediction, which may not correlate with registration error.

Recently, uncertainty estimation has gained growing interest in learning-based image registration. Dalca et al. [25] integrated PIR-based uncertainty estimation into a learning-based framework. Although it benefits from fast inference of learning-based methods, it still inherits the limitations of conventional methods. To address this, Chen et al. [26] proposed to calibrate the registration uncertainty using the fixed image to replace the mean model prediction. Furthermore, Zhang et al. [27] proposed heteroscedastic uncertainty estimation by using a separate variance network to predict

the variance map between the warped moving image and the fixed image. Nevertheless, existing learning-based approaches are still limited to mono-contrast uncertainty estimation, and there is still a gap in uncertainty estimation for learning-based multi-contrast image registration.

III. METHODOLOGY

In this work, we propose an adaptive conditional contrast-agnostic registration framework that can deal with even unobserved contrasts. Specifically, we proposed a random-convolution-based contrast-augmentation scheme that can simulate arbitrary contrast of images as inputs (Section III-A). To enable the network to learn contrast-invariant features for contrast-agnostic registration, we propose an adaptive conditional feature modulator to modulate learned features (Section III-B) and a contrast invariance loss to enforce consistency among features of the similar structure (Section III-C). Additionally, an uncertainty estimation mechanism is incorporated into the AC-CAR for providing uncertainty estimation for contrast-agnostic registration (Section III-D).

A. Contrast Augmentation via Random Convolution

To enable a contrast-agnostic registration framework, the network needs to learn contrast-invariant features for downstream deformation prediction. However, most existing learning-based multi-contrast registration approaches [10], [11] fail to generalize to unseen imaging contrasts during inference due to their dependence on a fixed set of imaging contrasts during training, leading the network to learn contrast-dependent features. This leads to a strong dependency between the input imaging contrast and its learned feature, therefore resulting in reduced generalizability.

To mitigate this, we propose a contrast augmentation scheme that simulates diverse imaging contrasts using Random Convolution (RC), similar to prior work such as [14]. The contrast augmentation scheme stacks multiple randomly initialized convolutional kernels with kernel size equal to one and LeakyReLU activations to simulate non-linear mapping among different MRI contrasts. We follow the same network setting for the contrast augmentation scheme as our earlier proposed work [22]. The augmented images capture diverse contrast patterns, simulating potential distributions of unseen contrasts. They are then used as inputs to train our contrast-agnostic registration network. By varying the input imaging contrasts at every training iteration, this approach alleviates the network's reliance on contrast-specific information, enabling the network to predict the deformation field in a contrast-agnostic manner.

B. Adaptive Conditional Feature Modulator

Handling inputs of arbitrary imaging contrasts with a single encoder is challenging, as it may struggle to manage the variability in different imaging contrasts and result in a trivial solution instead of contrast-invariant feature representations. To overcome these limitations, we propose an Adaptive Conditional Feature Modulator (ACFM) that leverages conditional

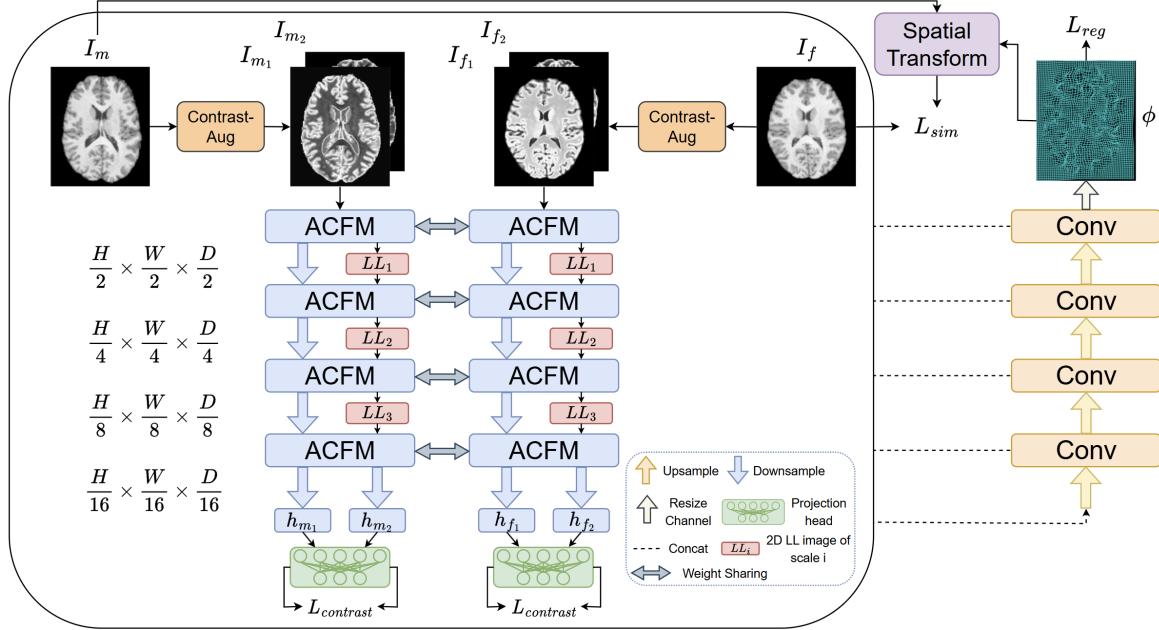


Fig. 1. Overview of the proposed registration framework. We first simulate two augmented images for each moving and fixed image with the contrast augmentation module. The augmented images are then used as the network inputs. At the encoder part, we use our proposed adaptive conditional feature modulator (ACFM) and contrast-invariant latent regularization (CLR) to extract contrast-invariant features. The learned features are then used to estimate the deformation field. The deformation field is then used to warp the pre-augmented moving image. The similarity loss is calculated on the warped image and the pre-augmented fixed image.

instance normalization (CIN) [28] to enable the network to adaptively modulate the feature representation according to different imaging contrasts. This is achieved by conditioning the modulation on the low-frequency component extracted via the discrete wavelet transform (DWT). Unlike previous wavelet-based registration methods such as WiNet [7], which use DWT to directly represent the deformation field, ACFM instead employs the DWT-extracted low-frequency component to inversely remove contrast-related variations from the learned features through CIN. Specifically, ACFM is conditioned on the low-frequency component of the input image, based on the assumption that contrast-related information is predominantly concentrated in low-frequency bands. The details of the ACFM are shown in Fig. 2. Within the ACFM, we first extract the low-frequency components (LL_i) of the image by using DWT to modulate the feature representation in a multi-scale manner, where i denotes the current scale. The first ACFM is conditioned on the original image LL_0 at the original scale. Each following ACFM is conditioned on the low-frequency components of the conditioned image from the previous ACFM by performing another DWT. Each conditioned image of ACFM shares the same size as the feature of that layer, allowing it to modulate the feature for each scale. To use the LL_i for CIN, we first expand LL_i into a vector v_i . The v_i are then fed into a projection head of a linear layer to generate the scale and shift parameters, α and β . The operation of CIN is summarized as

$$h'_i = \alpha_{\theta,i}(v_i) \left(\frac{h_i - \mu(h_i)}{\sigma(h_i)} \right) + \beta_{\theta,i}(v_i), \quad (2)$$

where h_i denotes the input of the current ACFM layer in channel i , $\mu(h_i)$ and $\sigma(h_i)$ are the mean and standard devi-

ation of h_i , and $\alpha_{\theta,i}(v_i)$ and $\beta_{\theta,i}(v_i)$ are the scale and shift parameters for each channel in the hidden feature map h_i . By conditioning the feature representation on the low-frequency component of the input image, we enable the network to modulate the learned feature representations adaptively to remove their contrast-related information inversely.

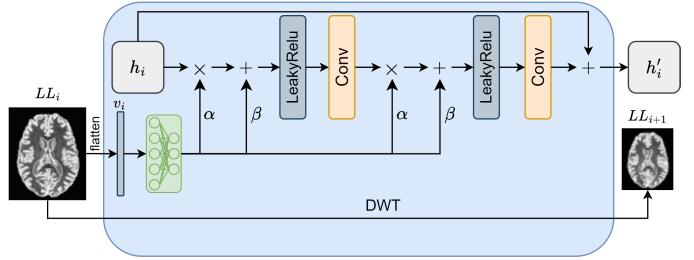


Fig. 2. Overview of the proposed Adaptive Conditional Feature Modulation Module (ACFM).

C. Contrast-Invariant Latent Regularization (CLR)

The proposed ACFM (Section III-B) can adaptively modulate the feature according to varying input contrasts to encourage contrast-invariant feature learning. However, this modulation is not explicitly constrained to achieve this. To further enhance contrast-invariant feature learning, we propose regularizing the feature representation in latent space using our proposed contrast invariance loss. Specifically, we aim to pull the feature representations of the images with the same shape information close in latent space under the assumption that contrast-invariant feature representations should be highly correlated with the shape information of the images. We generate two sets of contrast-augmented image pairs $\{I_{m1}, I_{f1}\}$

and $\{I_{m_2}, I_{f_2}\}$ based on a mono-contrast image pair $\{I_m, I_f\}$ with two separate contrast augmentation. Then, the contrast invariant loss can be formulated as follows,

$$\mathcal{L}_{\text{contrast}} = \mathbb{E}_{\Omega} \left[\|\text{Proj}(h_{m_1}) - \text{Proj}(h_{m_2})\|_2^2 + \|\text{Proj}(h_{f_1}) - \text{Proj}(h_{f_2})\|_2^2 \right], \quad (3)$$

where Proj represents the projection head, $h_{m_1}, h_{m_2}, h_{f_1}$, and h_{f_2} represent the corresponding latent features after the encoder. The projection head consists of a linear layer. The contrast invariance loss aims to pull the feature representation with the same shape as close as possible in the latent space.

It serves two critical purposes during training: 1) It compels the encoder to learn intrinsic structural features that are robust to variations in imaging contrasts and appearances [29]. 2) It provides guidance for the modulations of ACFM to encourage contrast-invariant feature representations to remove contrast-related information inversely.

D. Contrast-Agnostic Uncertainty Estimation Network

Beyond contrast-agnostic registration, we enhance the framework with a contrast-agnostic uncertainty estimation module to capture errors from heteroscedastic noise and highlight regions at higher risk of misregistration. Existing uncertainty estimation approaches for learning-based image registration are limited to mono-contrast scenarios [26], [27]. However, in multi-contrast settings, the intensity correspondence between contrasts is highly nonlinear, making it more difficult to disentangle uncertainty arising from true misalignment from that induced by contrast differences. This challenge complicates the estimation of reliable uncertainty. To address this issue, we adapt heteroscedastic uncertainty modelling to the contrast-agnostic setting using our learned contrast-invariant feature representations. Inspired by Zhang et al. [27], we use a separate variance network that predicts the input-specific heteroscedastic variance map for each multi-contrast image pair. To allow the variance network to be contrast-agnostic, the variance network is constructed on the contrast-agnostic encoder of AC-CAR, which yields contrast-invariant feature representations. These features are then used to predict the variance map of the warped moving image. The network details are shown in Fig. 3. The variance network thus has a weight-sharing encoder with the AC-CAR network and a separate decoder, which is used to output the heteroscedastic variance map. Specifically, the variance network takes the warped moving augmented image $I_{m_1} \circ \phi$ and the fixed augmented image I_{f_1} as the input. We optimize the variance network based on the β -NLL objective [30], similar to Zhang et al. [27]. Same as the registration network, we use the pre-augmented mono-contrast image pairs for loss supervision to facilitate the β -NLL objective in a multi-contrast registration scenario. The loss function for optimizing the variance network can be formulated as follows.

$$\mathcal{L}_{\text{var}} = \mathbb{E}_{\Omega} \left[\lfloor \hat{\sigma}_I^{2\beta} \rfloor \left(\frac{1}{\hat{\sigma}_I^2} \|I_m \circ \lfloor \phi \rfloor - I_f\|_2^2 + \log \hat{\sigma}_I^2 \right) \right], \quad (4)$$

where $\hat{\sigma}_I$ denotes the estimated variance map, β denotes the hyperparameter in the β -NLL objective. $\lfloor \cdot \rfloor$ denotes gradient stopping to prevent duplicate back-propagation by the two subnetworks. Similar to Zhang et al. [27], we output $\log \hat{\sigma}_I^2$ for the variance network for numerical stability.

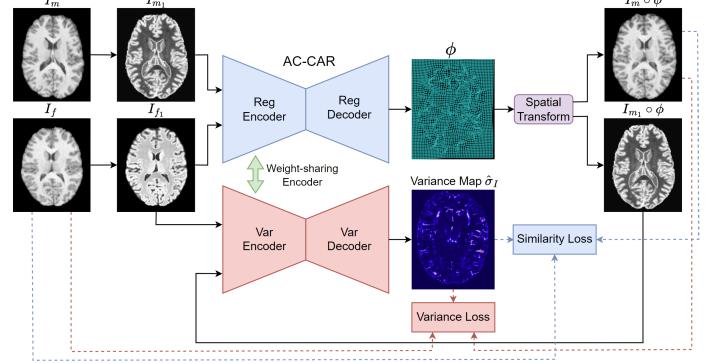


Fig. 3. Overview of the proposed contrast-agnostic uncertainty estimation framework.

E. Overall Network Architecture and Loss function

The overall registration framework is illustrated in Fig. 1. It follows a U-shaped architecture [31], with two Siamese encoders for the moving and fixed images. Skip connections [32] are employed between the two Siamese encoders and the decoder. The moving and fixed images are input into their respective encoders, producing latent features h_m and h_f , each reduced to 1/16 of the original image size. The two learned latent features are then concatenated and fed into the decoder, which is used to estimate the final deformation field.

The loss function of the registration network includes three terms: a similarity loss \mathcal{L}_{sim} , a regularization loss \mathcal{L}_{reg} , and the proposed contrast invariance loss $\mathcal{L}_{\text{contrast}}$. The loss function for the variance network is given in Eq. 4. As the input contrast-augmented image pairs are generated from the mono-contrast image pair $\{I_m, I_f\}$, we propose using the pre-augmented mono-contrast image pairs for loss supervision with a mono-contrast similarity measure. This avoids designing a multi-contrast similarity loss. The loss function of the registration network can be formulated as

$$\begin{aligned} \mathcal{L}_{\text{total}} = & \mathbb{E}_{\Omega} \left[\lfloor \hat{\sigma}_I^{2\beta-2} \rfloor \|I_m \circ \phi - I_f\|_2^2 + \lambda_1 \|\nabla \phi\|_2^2 \right] + \\ & \lambda_2 \mathcal{L}_{\text{contrast}} - \lambda_3 \text{LNCC}(I_m \circ \phi, I_f), \end{aligned} \quad (5)$$

where $\lfloor \cdot \rfloor$ denotes gradient stopping. λ_1 , λ_2 , and λ_3 are the respective hyperparameters of the diffusion regularizer, contrast invariance loss, and the additional LNCC loss. Similar to Zhang et al [27], we adopt an alternating training strategy for the registration and variance networks.

IV. EXPERIMENTS SETUP

A. Datasets and Preprocessing

1) Inter-subject 3D brain MRI registration of multi-contrast images: For the brain registration task, we use two multi-contrast datasets. (1) The public Cambridge Centre for Ageing

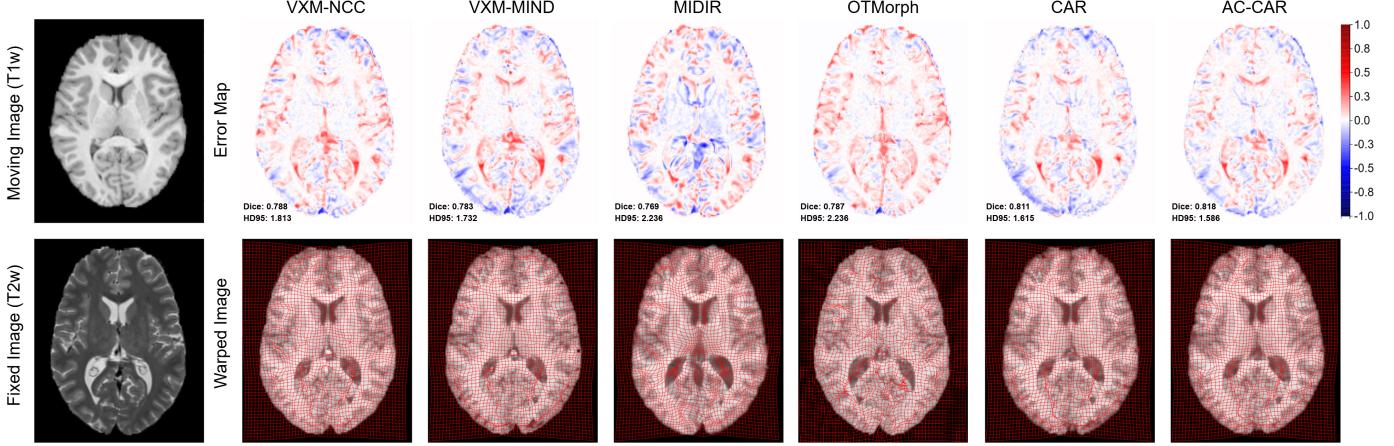


Fig. 4. Qualitative results on CamCAN dataset. We present the registration results of the middle slices along the z-axis of the same volume for illustration. The first row shows the error map of our proposed method against the baseline. The Dice score and HD95 of the whole volume are shown at the bottom left of the error map. The second row shows the warped images overlaid with the deformation fields.

and Neuroscience (CamCAN) project [33], [34]. The dataset has 3D MRI T1-weighted (T1w) and T2-weighted (T2w) images of 652 subjects. All images have a spatial resolution of $1 \times 1 \times 1$ mm and are cropped to a size of $160 \times 192 \times 160$. The dataset was randomly split into 600, 10, and 42 subjects for training, validation, and testing, where subjects were randomly selected to form T1w-T2w image pairs. (2) The Information eXtraction from Images (IXI) database. The dataset comprises 3D MRI T1w, T2w, and Proton Density-weighted (PD) images of 397 subjects. All images are processed using Freesurfer [35] for skull-stripping. SimpleITK [36] is adopted for bias-field correction and affine alignment to normalize all the images into a common space and have the same spatial resolution. All images are cropped to a size of $192 \times 160 \times 144$. MALPEM [37] is used for automated segmentation of 138 cortical and subcortical structures, categorized into 5 groups. The dataset was randomly split into 347, 10, and 40 subjects for training, validation, and testing, where subjects were randomly selected to form T1w-T2w, T1w-PD, and T2w-PD image pairs. For training AC-CAR and CAR, image pairs are formed using T1w images only for both datasets.

2) Intra-subject registration of multiple contrasts of T1 mapping acquisition in cardiac MRI: For the cardiac registration task, we use the public 2D CMRxRecon cardiac dataset [38], [39]. CMRxRecon dataset consists of two parts: (1) Cine MRI, and (2) T1 mapping with images of nine different T1 weightings (nine different imaging contrasts, $TI_{i=1:9}$). Both Cine MRI and T1 Mapping data in the dataset contain 167 subjects. Each subject includes 7 to 12 or 4 to 5 short-axis view slices for Cine MRI data and T1 Mapping data, respectively. The images in Cine and T1 Mapping data are cropped into a size of 128×128 . The dataset was randomly split into 137, 10, and 20 subjects for training, validation, and testing. The image sequence within each slice is considered free from motion artefacts. Therefore, we simulate random deformations to distort images using artificially generated Free-Form Deformations (FFDs) with various mesh spacings following Luo et al. [40]. To simulate realistic deformation, we restrict the deformation within the cardiac region. For

sequences within the same slice, we selected the original image TI_1 and a spatially randomly deformed image from $TI'_{i=2:9}$ as an image pair. For training AC-CAR and CAR, the image pairs are formed by the original image TI_1 and its own deformed pair TI'_1 .

3) Intra-subject registration of abdominal MR-CT: To evaluate the performance of AC-CAR beyond multi-contrast images, we also conduct experiments on abdominal CT-MR registration using the Learn2Reg MICCAI challenge dataset [41]. The Learn2Reg dataset contains 3D T1-weighted MR and CT abdominal images. The data were resampled to $3 \times 3 \times 3$ mm and cropped to a size of $112 \times 96 \times 112$. Training was conducted on 40 unpaired MR and 50 unpaired CT images, with CT-MR pairs formed through random selection. 8 CT-MR pairs were used for testing. The masks provided by the dataset were used to confine the information used for registration. Whereas VXM-MIND and MIDIR are trained directly on CT-MR pairs, AC-CAR is trained solely on T1w MR image pairs. Informed consent was obtained during the public data release for all datasets.

B. Evaluation Metrics

The registration accuracy is evaluated by the Dice score and the Hausdorff Distance. The Dice score measures the degree of overlap between the anatomical segmentation of the warped moving image and the fixed image. The Hausdorff Distance (HD) captures the maximum distance between corresponding anatomical regions in the warped moving and fixed images. To mitigate the impact of outliers, we report the 95th percentile of the HD (HD95). Additionally, we measure extreme deformations by calculating the folding ratio—the percentage of points with a negative Jacobian determinant ($J_{<0\%}$), and evaluate deformation smoothness through the magnitude of the spatial gradient of the Jacobian determinant $|\nabla_J|$ [11].

C. Implementation Details

The registration network comprises two Siamese encoders followed by a decoder. For 3D brain registration, encoder

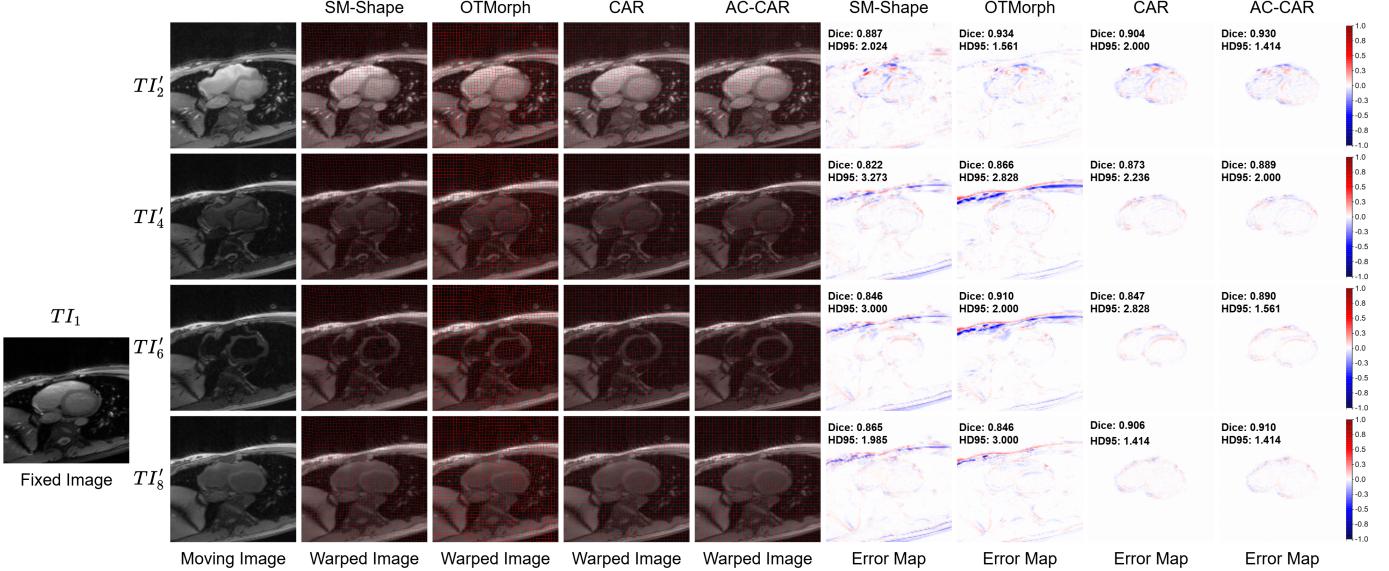


Fig. 5. Qualitative results on CMRxRecon dataset using T1 Mapping data for training. We show the results of registering misaligned images of $TI_{i=2,4,6,8}$ to TI_1 . Columns 2-5 present the warped images overlaid with the deformation fields, and columns 6-10 present the error map of our proposed method against the baseline. The Dice score and HD95 are shown at the top left of the error map.

TABLE I

QUANTITATIVE RESULTS OF ON BOTH CAMCAN AND CMRxRECON DATASETS. * REPRESENTS FOR VALUES THAT AC-CAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Methods	CamCAN (T1w→T2w)				CMRxRecon			
	Dice \uparrow	$ \nabla J \downarrow$	$J_{<0\%} \downarrow$	HD95 \downarrow	Dice \uparrow	$ \nabla J \downarrow$	$J_{<0\%} \downarrow$	HD95 \downarrow
Unregistered	0.609* \pm 0.078	-	-	5.418* \pm 1.829	0.783* \pm 0.079	-	-	2.749* \pm 0.546
SyN	0.756* \pm 0.023	0.043 \pm 0.005	0.430 \pm 0.146%	2.033* \pm 0.199	0.601* \pm 0.271	0.019* \pm 0.006	0.034* \pm 0.073%	6.142* \pm 4.469
VXM-LNCC	0.753* \pm 0.029	0.065* \pm 0.005	0.549* \pm 0.122%	2.098* \pm 0.187	0.838* \pm 0.085	0.051* \pm 0.015	0.004 \pm 0.008%	2.301* \pm 0.760
VXM-MIND	0.746* \pm 0.032	0.065* \pm 0.006	0.619* \pm 0.228%	2.157* \pm 0.198	0.837* \pm 0.077	0.051* \pm 0.015	0.007* \pm 0.020%	2.158* \pm 0.532
MIDIR	0.761* \pm 0.023	0.052* \pm 0.005	0.600* \pm 0.146%	2.152* \pm 0.175	0.795* \pm 0.092	0.023* \pm 0.005	0.018* \pm 0.045%	2.670* \pm 0.728
SM-Shape	0.698* \pm 0.063	0.074* \pm 0.009	0.513* \pm 0.112%	3.305* \pm 1.764	0.836* \pm 0.078	0.049* \pm 0.005	0.006* \pm 0.016%	2.302* \pm 0.576
CAR	0.784* \pm 0.029	0.063* \pm 0.007	0.558* \pm 0.145%	1.824* \pm 0.239	0.860* \pm 0.073	0.014* \pm 0.005	0.004 \pm 0.010%	2.016* \pm 0.468
OTMorph	0.775* \pm 0.026	0.072* \pm 0.006	0.607* \pm 0.211%	1.953* \pm 0.188	0.876 \pm 0.056	0.032* \pm 0.009	0.004 \pm 0.006%	2.273* \pm 0.681
UTSRMorph	0.771* \pm 0.024	0.062* \pm 0.006	0.604* \pm 0.145%	1.955* \pm 0.188	0.859* \pm 0.061	0.024* \pm 0.027	0.005 \pm 0.010%	2.557* \pm 0.781
AC-CAR (w/o NCC)	0.787* \pm 0.023	0.059* \pm 0.007	0.703* \pm 0.165%	1.766* \pm 0.191	0.863* \pm 0.065	0.012 \pm 0.004	0.003 \pm 0.009%	1.943* \pm 0.456
AC-CAR (Ours)	0.808 \pm 0.026	0.056* \pm 0.006	0.494* \pm 0.101%	1.652 \pm 0.236	0.871* \pm 0.068	0.015* \pm 0.005	0.003 \pm 0.009%	1.862 \pm 0.464

layers have 32 channels and decoder layers 64 channels; for 2D cardiac registration, encoder layers have 128 channels and decoder layers 256 channels. The projection head for CLR employs a 1×1 convolution with 16 output channels for 3D tasks and 32 for 2D tasks. The random convolution-based contrast augmentation uses 4 RC layers, with kernel weights sampled from a uniform distribution $U(0, 10)$, re-normalized to be zero-centered, and followed by a LeakyReLU activation with a negative slope of 0.2. Configurations of contrast augmentation are based on preliminary ablation studies from our previous work [22]. Haar wavelet transform is adopted for the DWT in our proposed ACFM. To reduce computational costs while maintaining performance, a single 2D slice, the middle slice along the z-axis, is used as the condition of ACFM in 3D experiments, as it is expected to encapsulate the complete contrast information of the entire 3D image. λ_1 , λ_2 , and λ_3 are set to 0.15, 0.1, and 0.8 for 3D experiments and 0.3, 0.2, and 0.8 for 2D experiments, chosen through empirical tuning on a validation subset to balance deformation smoothness, registration accuracy, and CLR. The parameter β in Eq. 4

and Eq. 5 is fixed at 0.5 for all experiments, following [27], [30]. Experiments were conducted using Adam Optimizer [42] on an NVIDIA A100 GPU with a batch size of 1 for 3D experiments and 8 for 2D experiments. The learning rates are set as default values from official implementations for all baselines, and are set to be 1×10^{-4} for AC-CAR and CAR.

D. Baseline Methods

AC-CAR is first compared with a conventional iterative-based registration method, SyN [43], with mutual information as the similarity metric and a default Gaussian smoothing of 3 and three scales with 180, 80, 40 iterations, respectively, following the settings of TransMorph [26]. AC-CAR is also compared with SOTA learning-based methods, including VoxelMorph [31] using LNCC (VXM-LNCC) and MIND [5] (VXM-MIND), respectively, a multi-contrast mutual information-based registration approach MIDIR [11], a contrast-agnostic registration framework, SynthMorph [13], and our previous work on contrast-agnostic registration, CAR [22], as well as two recently proposed methods, OTMorph

TABLE II

QUANTITATIVE RESULTS ON THE IXI DATASETS. * REPRESENTS FOR VALUES THAT AC-CAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST. THE RUNNING TIME FOR MODEL TRAINING (HR) AND THE INFERENCE TIME PER VOLUME PAIR (SEC) ARE REPORTED.

Methods	T1w→T2w		T1w→PD		T2w→PD	
	Dice ↑	$J_{<0\%} \downarrow$	Dice ↑	$J_{<0\%} \downarrow$	Dice ↑	$J_{<0\%} \downarrow$
Unregistered	0.685* \pm 0.062	-	0.685* \pm 0.062	-	0.685* \pm 0.062	-
SyN	0.782* \pm 0.035	0.571* \pm 0.118%	0.768* \pm 0.038	0.481* \pm 0.122%	0.744* \pm 0.043	0.489* \pm 0.125%
VXM-LNCC	0.780* \pm 0.036	0.460 \pm 0.082%	0.768* \pm 0.039	0.560* \pm 0.089%	0.756* \pm 0.045	0.621* \pm 0.116%
VXM-MIND	0.774* \pm 0.035	0.627* \pm 0.110%	0.763* \pm 0.036	0.513* \pm 0.097%	0.751* \pm 0.043	0.618* \pm 0.112%
MIDIR	0.777* \pm 0.032	0.473 \pm 0.103%	0.766* \pm 0.034	0.464 \pm 0.096%	0.753* \pm 0.036	0.346 \pm 0.075%
SM-Shape	0.701* \pm 0.044	0.459 \pm 0.118%	0.698* \pm 0.045	0.483* \pm 0.093%	0.700* \pm 0.059	0.350 \pm 0.132%
CAR	0.794* \pm 0.035	0.489* \pm 0.087%	0.787* \pm 0.035	0.433 \pm 0.078%	0.782* \pm 0.042	0.439 \pm 0.082%
OTMorph	0.775* \pm 0.034	0.833* \pm 0.115%	0.761* \pm 0.038	0.841* \pm 0.110%	0.748* \pm 0.041	0.939* \pm 0.145%
UTSRMorph	0.781* \pm 0.034	0.569* \pm 0.094%	0.763* \pm 0.038	0.574* \pm 0.091%	0.764* \pm 0.042	0.620* \pm 0.109%
AC-CAR (Ours)	0.805\pm0.036	0.438\pm0.074%	0.796\pm0.035	0.427\pm0.079%	0.791\pm0.041	0.409\pm0.097%

[44] and UTSRMorph [45]. Synthetic images for training SynthMorph are generated from random label maps (SM-Shape). Note that the channel dimensions of all convolutional layers in VXM-LNCC, VXM-MIND, SM-Shape, and OTMorph are set to 64 for 3D experiments and 256 for 2D experiments to ensure a fair comparison with AC-CAR, and the channel dimension of CAR is the same as that of AC-CAR. In addition, we implemented the large version of UTSRMorph (UTSRMorph-large) for evaluation.

V. RESULTS

A. Comparison Studies on Registration Accuracy

1) *3D Inter-patient Brain MRI Registration on CamCAN dataset:* The left side of Table I summarizes the quantitative results for the 3D inter-patient registration task on the CamCAN dataset. It can be observed that AC-CAR achieved the best registration accuracy while maintaining a low folding ratio compared to other learning-based approaches. Compared to MIDIR, OTMorph, and CAR, AC-CAR achieves registration accuracy improvements of 4.7%, 3.3%, and 2.4% in terms of Dice score, respectively. It should be noted that both AC-CAR and CAR only used T1w images for training, which demonstrates their superior generalization ability. Removing NCC during the training of AC-CAR results in performance degradation, demonstrating the effectiveness of adding NCC as the similarity loss. The improvement in terms of the Dice score of AC-CAR compared to CAR shows the effectiveness of our proposed ACFM module. Both AC-CAR and CAR significantly outperform the other contrast-agnostic registration approaches, SM-shape. Because it relies on randomly generated labels for training, and may not register well in terms of fine-grained details of the human brain. Fig. 4 visualizes the registration error maps and deformation fields for a specific image pair, comparing AC-CAR with other baseline methods. We can observe from the figure that AC-CAR can achieve lower registration error and a smoother deformation field compared to other baselines.

2) *Generalization Experiments on 3D Brain MRI registration on CamCAN dataset:* To assess the generalizability of AC-CAR, we evaluated its performance on the CamCAN dataset across different contrast combinations. Specifically, we tested T2w-to-T1w, T1w-to-T1w, and T2w-to-T2w registrations. For

generalizability evaluation, we directly employed AC-CAR and other baseline models pretrained on the CamCAN dataset originally used for T1w-T2w registration. The results are shown in Table III. We can observe from the table that AC-CAR still performs the best among all the methods. For the T2w-T1w registration task, AC-CAR achieves registration accuracy improvements of 2.7% and 2.1% compared to OTMorph, UTSRMorph in terms of Dice score, respectively. For both the T1w-to-T1w and T2w-to-T2w registration tasks, all methods except AC-CAR, CAR, and SM-shape exhibited substantial performance degradation. The Dice score of most baseline methods is even lower than the unregistered image. This is likely because these methods were trained specifically for the T1w-to-T2w registration task. While they could maintain performance on the related T2w-to-T1w task, they failed on the other two tasks due to the distributional shift between the training and test data. For SM-Shape, the performance showed little variation across the three tasks since it was trained on randomly generated label maps. However, the registration accuracy of SM-Shape remained well below that of CAR and AC-CAR. Furthermore, AC-CAR consistently outperformed CAR across all three tasks with statistical significance, demonstrating that the proposed ACFM module effectively enhances model generalizability.

3) *3D Inter-patient Brain MRI Registration on IXI dataset:* Table II summarizes the quantitative results for the 3D inter-patient registration task on the IXI dataset. We reported the registration accuracy in terms of Dice score and deformation regularity in terms of folding ratio for T1w-T2w, T1w-PD, and T2w-PD registration tasks. It can be observed that AC-CAR still consistently achieved the best registration accuracy while maintaining a low folding ratio compared to other learning-based approaches for all three registration tasks. For the T1w-T2w and T1w-PD registration tasks, AC-CAR achieves Dice improvements of 2.3% and 2.8%, respectively, over the second-best baseline methods (excluding CAR). For the T2w-PD registration task, despite not being exposed to either contrast during training, AC-CAR demonstrates superior registration accuracy compared to all other methods, achieving Dice improvements of 2.7% over the second-best baseline method (excluding CAR), demonstrating the superior generalizability of AC-CAR. Compared to CAR, AC-CAR also shows

TABLE III

QUANTITATIVE RESULTS OF DIFFERENT CONTRAST COMBINATIONS ON THE 3D CAMCAN DATASET. * REPRESENTS FOR VALUES THAT AC-CAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Methods	T2w→T1w		T1w→T1w		T2w→T2w	
	Dice ↑	$J_{<0\%} \downarrow$	Dice ↑	$J_{<0\%} \downarrow$	Dice ↑	$J_{<0\%} \downarrow$
Unregistered	0.609*±0.078	-	0.609*±0.078	-	0.609*±0.078	-
SyN	0.764*±0.023	0.493±0.164%	0.823*±0.014	0.736*±0.091%	0.782*±0.019	0.695*±0.134%
VXM-LNCC	0.762*±0.036	0.468±0.168%	0.597*±0.029	0.392 *±0.091%	0.592*±0.028	0.327 *±0.089%
VXM-MIND	0.738*±0.038	0.389 *±0.255%	0.626*±0.045	0.445±0.281%	0.617*±0.033	0.489*±0.168%
MIDIR	0.720*±0.030	0.404±0.138%	0.557*±0.026	0.862*±0.146%	0.556*±0.024	0.973*±0.174%
SM-Shape	0.698*±0.059	0.484±0.150%	0.706*±0.063	0.413±0.161%	0.694*±0.066	0.524*±0.217%
CAR	0.803*±0.029	0.598*±0.164%	0.825*±0.036	0.544*±0.148%	0.788*±0.042	0.408±0.147%
OTMorph	0.787*±0.029	0.745*±0.169%	0.547*±0.028	1.451*±0.131%	0.543*±0.027	1.601*±0.125%
UTSRMorph	0.793*±0.026	0.463±0.100%	0.571*±0.025	0.857*±0.129%	0.558*±0.022	0.902*±0.159%
AC-CAR (Ours)	0.814 *±0.027	0.521±0.137%	0.844 *±0.033	0.488±0.128%	0.799 *±0.038	0.401±0.101%

statistically significant improvement over Dice score, showing the effectiveness of our proposed ACFM module.

4) 2D cardiac MRI registration for T1 Mapping Sequence:

The right part of Table. I shows the quantitative comparison results of the 2D cardiac registration task for the T1 mapping sequence in the CMRxRecon dataset. AC-CAR still achieves the best registration accuracy and good deformation regularity. Compared to VXM-LNCC and UTSRMorph, AC-CAR still achieves a 3.3% and 1.2% higher Dice score, respectively. AC-CAR also outperforms its counterpart of removing NCC. Noted that AC-CAR and CAR still used only single contrast images for training, which further demonstrated the generalizability of AC-CAR when applied to eight other unseen contrasts during inference. OTMorph achieves slightly higher Dice scores but also higher HD95 values than AC-CAR when trained on all available contrasts. Nonetheless, AC-CAR demonstrates superior performance in delineating structural boundaries, as reflected by its lower HD95 in this setting. Fig. 5 shows the qualitative registration results of a specific image pair of AC-CAR versus the baseline methods. It can be shown that AC-CAR can register more fine-grained details with smooth deformation fields. In addition, SM-Shape and OTMorph exhibit errors outside the cardiac region, despite the simulated deformation being confined within it. This may be because the leveraged image-to-image translation in OTMorph could introduce extra translation error, and the SM-shape could overlook details within the background region due to the Dice loss, resulting in background errors. On the contrary, AC-CAR and CAR use pre-augmented mono-contrast images to compute LNCC, which results in a zero dissimilarity difference for regions without deformation. This demonstrated that AC-CAR has a better capability to capture the deformation while using mono-contrast similarity for training.

5) Generalization Experiments on 2D cardiac MRI registration:

To further evaluate the generalizability of AC-CAR, we also trained our model and baseline methods on CMRxRecon Cine MRI data and evaluated model performance on the multi-contrast T1 Mapping data (also registering misaligned images of $T1_{i=2:9}$ to $T1_1$). The results are shown in Table. IV. AC-CAR remains to achieve the best registration accuracy and a 4% higher Dice score compared to SM-Shape. It should be noted that AC-CAR and CAR achieve similar performance compared to their counterpart trained using the T1 Map-

TABLE IV

QUANTITATIVE RESULTS OF THE GENERALIZABILITY EXPERIMENT. THE MODELS ARE TRAINED ON CMRxRECON CINE DATA WHILE TESTED ON CMRxRECON T1 MAPPING DATA. * REPRESENTS FOR VALUES THAT AC-CAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Methods	Dice ↑	$ \nabla J \downarrow$	$J_{<0\%} \downarrow$	HD95 ↓
Unregistered	0.783*±0.079	-	-	2.749*±0.546
VXM-LNCC	0.790*±0.129	0.089*±0.025	0.176*±0.168%	2.773*±1.200
VXM-MIND	0.835*±0.088	0.072*±0.021	0.010*±0.016%	2.233*±0.648
MIDIR	0.764*±0.047	0.013 *±0.047	0.153*±0.125%	2.926*±0.927
SM-Shape	0.836*±0.078	0.049*±0.005	0.006±0.016%	2.302*±0.576
OTMorph	0.865*±0.063	0.030*±0.008	0.004 *±0.006%	2.464*±0.732
UTSRMorph	0.848*±0.078	0.037*±0.017	0.005±0.009%	2.714*±1.032
CAR	0.866*±0.073	0.018±0.005	0.006±0.014%	1.915*±0.518
AC-CAR (Ours)	0.876 *±0.062	0.019±0.006	0.006±0.015%	1.762 *±0.506

ping sequence. This demonstrated that our proposed contrast-agnostic registration framework can generalize to unseen contrasts during inference without sacrificing performance. However, VXM-LNCC, VXM-MIND, MIDIR, OTMorph, and UTSRMorph suffered different degrees of performance degradation compared to using the T1 Mapping sequence for training. Compared with OTMorph and UTSRMorph, AC-CAR demonstrates a substantially greater improvement in HD95 than in Dice score. This demonstrated that AC-CAR can better deal with structural boundaries. SM-shape maintains the same performance since it was trained on randomly generated labels. Compared to CAR, AC-CAR still achieves a 1% higher Dice score, demonstrating that the proposed ACFM can also further improve the network's generalizability.

B. Evaluation on Registration Uncertainty

We evaluate the estimated variance map through our proposed contrast-agnostic uncertainty estimation framework using the sparsification error plots from Poggi et al. [46] following Zhang et al. [27]. The sparsification error plot can provide a quantitative measure of the accuracy of the estimated variance. The plots are calculated by removing one pixel (voxel) at a time from the largest to the smallest variance magnitudes and measuring the MSE of the remaining pixels. An ideal sparsification error plot should decrease monotonically, indicating that the estimated uncertainty map can correctly identify pixels with the largest errors. As shown in Fig. 6,

our proposed method can effectively estimate the registration uncertainty of the warped moving image as the sparsification plots are monotonically decreasing. The estimated variance is highly correlated with the registration error. This can be demonstrated by the high correlations between the areas with high variance in the variance map and the areas with high registration errors in the error map.

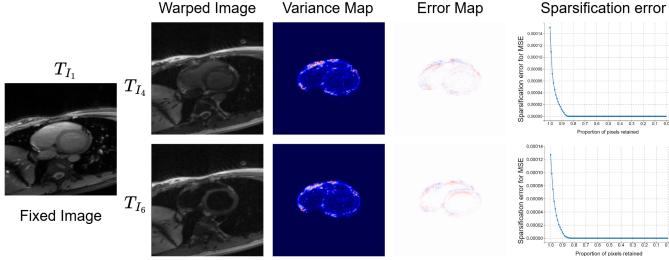


Fig. 6. Illustration of the estimated variance map and corresponding sparsification plot on the CMRRecon dataset.

C. Evaluation on Feature Invariance

To evaluate whether the features learned by AC-CAR are genuinely contrast-agnostic, we quantify the differences between features extracted from input image pairs that share the same anatomy but differ in imaging contrast. For each subject, we first extract features from eight input image pairs obtained from the T1 Mapping sequence without applying random spatial deformations ($TI_1 - TI_{i=2:9}$). Since the proposed CLR is applied at the end of the encoder, the learned features are highly abstract at that stage. Therefore, we utilize features from the final layer of the decoder, which maintains the same spatial resolution as the input images, to compute feature differences, following [13]. Specifically, we generate all 56 unique pairs via permutation of the eight extracted features and calculate the root-mean-square difference (RMSD) for each pair, following [13]. The mean RMSD across the 56 feature pairs is then computed to quantify the model's feature invariance across the nine different imaging contrasts of a single subject. We compared AC-CAR with VXM-LNCC, VXM-MIND, and CAR, as all of them adopted a U-Net architecture. Figure 8 presents the quantitative results as a boxplot of the mean RMSD of the 56 feature pairs across all test subjects, while Figure 7 shows qualitative feature visualizations from the final decoder layer for four contrast pairs of the same subject. From Figure 7, we observe that the features extracted by AC-CAR and CAR remain consistent across different imaging contrasts of the same subject, whereas those produced by VXM-LNCC and VXM-MIND exhibit substantial variations in intensity and texture. However, compared with AC-CAR, the features from CAR appear oversimplified and lack anatomical detail, whereas AC-CAR preserves fine-grained anatomical boundaries and detail. The quantitative results in Figure 8 further support these observations. AC-CAR achieves substantially lower mean RMSD compared to VXM-LNCC and VXM-MIND, demonstrating superior invariance to contrast variations. While CAR also achieves low RMSD values, it falls short of AC-CAR in registration

accuracy according to previous experiments in Section V-A, highlighting the advantage of incorporating the proposed ACFM module, which enables the network to adaptively learn feature representations that are both anatomically informative and robust to variations in imaging contrast.

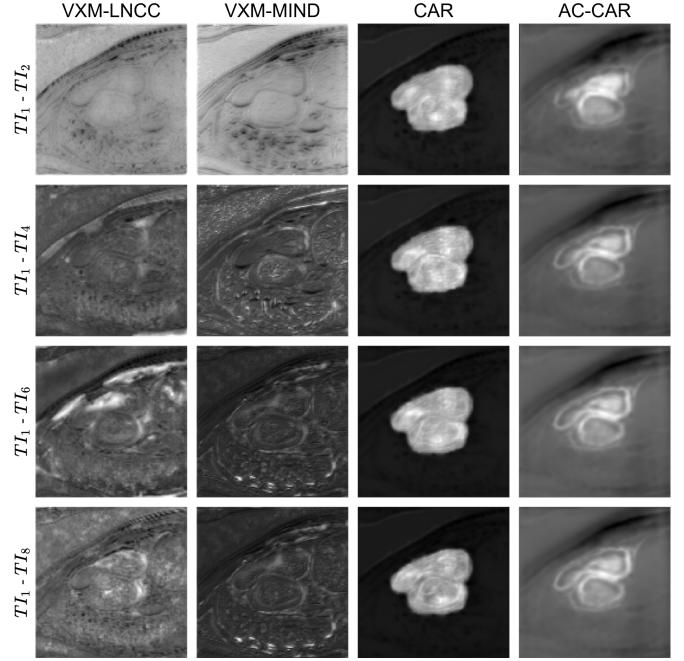


Fig. 7. Feature visualization from the final layer of the decoder of 4 feature pairs of the same subject on the CMRRecon dataset.

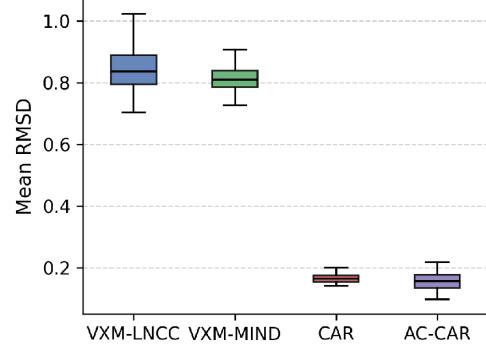


Fig. 8. Boxplot of Mean RMSD of 56 feature pairs across all subjects in the test set on the CMRRecon dataset.

D. Ablation Studies

1) Ablation study on network components: We assess the effectiveness and computational cost of each module in AC-CAR by systematically removing them and evaluating the resulting model performance along with the associated training and inference times on the CamCAN dataset. The results are summarized in Table. V. CA denotes the contrast augmentation scheme in Section III-A. UE denotes the uncertainty estimation module in Section III-D. By removing the proposed ACFM, we remove the conditioning mechanism in Fig.2 while the two convolution operations per layer are retained.

TABLE V

ABLATION STUDY ON EACH MODULE OF AC-CAR ON THE 3D CAMCAN DATASET AND CORRESPONDING TRAINING AND INFERENCE TIME ON GPU. * REPRESENTS THAT AC-CAR (WITH 4 CHECKMARKS) SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

CA	ACFM	CLR	UE	Dice \uparrow	$J_{<0\%} \downarrow$	Training	Inference
✓	✓			0.793* \pm 0.028	0.516 \pm 0.126%	31hr	3.277s
✓	✓			0.798* \pm 0.024	0.524 \pm 0.134%	30hr	3.292s
✓	✓	✓		0.807 \pm 0.027	0.517 \pm 0.126%	34hr	3.324s
✓	✓	✓	✓	0.808 \pm 0.026	0.494 \pm 0.101%	46hr	3.421s

Table. V shows that removing either ACFM or CLR leads to a performance drop, which demonstrates that ACFM and CLR mutually reinforce contrast-invariant feature learning. Notably, removing ACFM results in a more significant drop than removing CLR. This suggests that CLR alone cannot adequately handle distribution shifts across arbitrary contrasts, as the network is unable to modulate itself and therefore tends to overfit to each contrast seen at training. Moreover, adding CLR on top of ACFM can further improve the model performance, which shows that CLR can be regarded as a guided constraint for the modulation in ACFM to further enhance contrast-invariant feature learning. Table. V also shows that incorporating UE into AC-CAR does not yield a statistically significant change in model performance. The UE module primarily serves to provide trustworthiness to the AC-CAR.

For training time, the addition of the proposed ACFM results in only a marginal increase, extending the total from 31 hours to 34 hours. The exclusion of CLR yields the shortest training time, as each iteration involves only a single forward pass. By contrast, the most substantial increase arises from the UE module, which requires training an auxiliary variance network to estimate the uncertainty map. At the inference stage, however, the runtimes of all model variants remain virtually identical, indicating that both the ACFM and the UE module impose negligible additional computational overhead.

2) *Ablation study on the configuration of ACFM*: We first conduct the ablation study on using different frequency components as the condition for CIN. The ablation study in this section is conducted on the 2D slices of the CamCAN dataset for efficiency. The results are summarized in Table. VI. **Haar-LL** means that we use the low-frequency decomposed image only as we proposed in Section III-B. **Haar-H** means we used the other three decomposed images (LH, HL, HH) as the condition. **Haar-all** means we used all four decomposed images as a condition. It can be demonstrated that **Haar-all** suffers performance degradation by introducing other high-frequency components of the image for CIN. **Haar-H** experiences further performance degradation by discarding the low-frequency component of the image. This is because high-frequency components are more correlated with structural information than contrast information. This shows that the low-frequency component is more suitable for feature modulations to inversely remove contrast-related information through CIN.

In addition, we also performed an ablation study on different approaches for extracting low-frequency information

TABLE VI

ABLATION STUDY ON USING DIFFERENT COMPONENTS OF DWT DECOMPOSED IMAGES FOR ACFM ON 2D SLICES ON THE CAMCAN DATASET. * REPRESENTS FOR VALUES THAT HAAR-LL SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Variants	Dice \uparrow	$ \nabla_J \downarrow$	$J_{<0\%} \downarrow$	HD95 \downarrow
Haar-LL (Ours)	0.751 \pm 0.053	0.082 \pm 0.010	0.498 \pm 0.183%	2.949 \pm 0.642
Haar-H	0.733* \pm 0.054	0.083 \pm 0.010	0.530* \pm 0.186%	3.105* \pm 0.659
Haar-all	0.745* \pm 0.057	0.084 \pm 0.010	0.551* \pm 0.191%	3.015* \pm 0.642

TABLE VII

ABLATION STUDY ON DIFFERENT DISCRETE WAVELET TRANSFORMS FOR ACFM ON 2D SLICES ON THE CAMCAN DATASET. * REPRESENTS FOR VALUES THAT HAAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Variants	Dice \uparrow	$ \nabla_J \downarrow$	$J_{<0\%} \downarrow$	HD95 \downarrow
Daubechies	0.747 \pm 0.051	0.085* \pm 0.008	0.548* \pm 0.158%	2.992* \pm 0.641
Biorthogonal	0.746* \pm 0.054	0.082 \pm 0.010	0.484 \pm 0.186%	3.129* \pm 0.659
Low-pass FFT	0.746* \pm 0.057	0.086* \pm 0.010	0.556* \pm 0.191%	2.998* \pm 0.642
Haar (Ours)	0.751 \pm 0.053	0.082 \pm 0.010	0.498 \pm 0.183%	2.949 \pm 0.642
Haar-2nd	0.748 \pm 0.053	0.083 \pm 0.010	0.528* \pm 0.191%	2.969 \pm 0.665
Haar-3rd	0.747 \pm 0.053	0.081 \pm 0.009	0.515* \pm 0.182%	2.979 \pm 0.658
Haar-4th	0.740* \pm 0.061	0.080 \pm 0.010	0.484 \pm 0.164%	2.996* \pm 0.654

from the original input image in ACFM, including various types of DWT or using FFT directly for low-pass filtering. **Daubechies** and **Biorthogonal** represent using Daubechies and Biorthogonal wavelets for extracting the low-frequency component of the image. **Haar-2nd**, **Haar-3rd**, and **Haar-4th** represent using the low-frequency image of the second, third, and fourth level of Haar DWT for all encoder layers, respectively. Table. VII summarizes the results of different variants of our proposed method. It can be observed that all 3 types of DWT perform better than low-pass FFT. Among the three types of DWT, the Haar wavelet achieves the best performance while being the simplest and efficient window. Table. VII also demonstrated that using a progressively increasing level of DWT to generate low-frequency images for each encoder layer outperforms using a fixed level of DWT for all encoder layers.

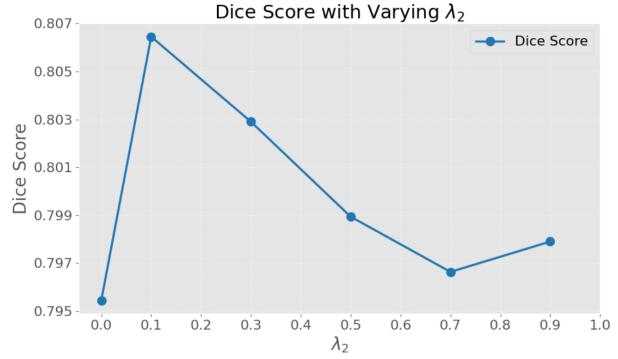


Fig. 9. Dice score on the 3D CamCAN dataset for AC-CAR with varied contrast-invariance latent regularization parameters.

3) *Ablation study on Contrast-Invariance Latent Regularization*: Fig. 9 shows the registration accuracy in terms of the Dice score for different values of the contrast-invariance loss parameter λ_2 on the 3D brain MRI inter-patient registration

TABLE VIII

COMPUTATIONAL COSTS ANALYSIS FOR AC-CAR AND BASELINE METHODS ON THE IXI DATASET, REPORTING FLOPS, MODEL PARAMETERS, AND PEAK GPU MEMORY MEASURED DURING THE INFERENCE STAGE, ALONG WITH TRAINING AND INFERENCE TIMES.

Methods	# Params	FLOPs	Memory	Training	Inference
VXM-LNCC	1.56M	2204.49G	9.18GB	26hr	2.478s
VXM-MIND	1.56M	2204.49G	9.18GB	18hr	2.568s
MIDIR	1.16M	116.02G	5.91GB	20hr	2.445s
SM-Shape	1.56M	2204.49G	9.00GB	16hr	2.613s
CAR	1.31M	2161.97G	8.87GB	26hr	2.517s
OTMorph	111.64M	7405.47G	10.84GB	72hr	2.308s
UTSRMorph	421.50M	979.22G	8.11GB	34hr	2.691s
AC-CAR (w/o UE)	2.83M	2442.21G	9.74GB	31hr	2.538s
AC-CAR (Ours)	4.05M	4546.37G	11.07GB	43hr	2.579s

task on the CamCAN dataset. It can be seen from the figure that when the CLR is removed at λ_2 equals 0, the model performance suffers a severe degradation. The Dice score increases as the CLR parameter λ_2 increases to 0.1, showing the effectiveness of our proposed Contrast-Invariant Latent Regularization (CLR). But when the CLR parameter λ_2 continues to increase, the Dice score experiences some degree of degradation and fluctuation. This demonstrated that excessive CLR can cause the network to overemphasize learning contrast-invariant features at the expense of information relevant to the registration task being ignored.

E. Computational Cost Analysis

To evaluate the computational cost of AC-CAR against the baseline learning-based approaches, we also reported the training and inference time, FLOPs, model parameters and peak GPU memory consumption during the inference stage in Table VIII. We can observe that VXM-based approaches (VXM-LNCC, VXM-MIND, SM-Shape) share similar architectural complexity, each using around 1.56M parameters, over 2200 GFLOPs, and approximately 9GB of GPU memory during inference. MIDIR is the most lightweight baseline in terms of FLOPs and memory consumption, although its inference speed remains similar to other learning-based methods. In contrast, OTMorph and UTSRMorph impose substantially higher computational burdens, with OTMorph exhibiting the highest FLOPs and longest training time, and UTSRMorph employing the largest parameter count. Against these baselines, AC-CAR(w/o UE) represents AC-CAR without the uncertainty estimation module, which stands for the pure registration cost of our framework. With 2.83M parameters and 2442.21 GFLOPs, it remains comparable to VXM-based methods while maintaining a competitive inference time of 2.538s and modest memory usage (9.74GB). Compared with CAR, AC-CAR(w/o UE) exhibits an increased parameter count due to the ACFM module, yet it introduces only marginal increases in FLOPs, memory usage, and inference time. AC-CAR increases the computational cost due to the uncertainty estimation module, yet its inference time (2.579s) remains nearly identical to that of other deep learning approaches and comfortably within the narrow 2-3s range shared across all learning-based models. This indicates that the proposed ACFM and uncertainty estimation modules are lightweight,

scalable addition that preserves the model's deployment efficiency. AC-CAR therefore demonstrates superior accuracy and reliability while maintaining efficient inference, achieving a better trade-off between computational cost and performance.

TABLE IX

QUANTITATIVE RESULTS ON THE LEARN2REG DATASETS. * REPRESENTS FOR VALUES THAT AC-CAR SIGNIFICANTLY OUTPERFORMED WITH p -VALUE < 0.01 IN A PAIRED t -TEST.

Methods	Dice \uparrow	$ \nabla J \downarrow$	$J_{<0\%} \downarrow$	HD95 \downarrow
Unregistered	0.498* \pm 0.127	-	-	13.817* \pm 7.649
VXM-MIND	0.704 \pm 0.120	0.058* \pm 0.004	0.367 \pm 0.109%	7.778 \pm 3.088
MIDIR	0.635* \pm 0.098	0.059* \pm 0.005	0.589* \pm 0.353%	11.656* \pm 3.798
SM-Shape	0.605* \pm 0.181	0.092* \pm 0.010	0.398 \pm 0.065%	12.364* \pm 8.647
AC-CAR (Ours)	0.644 \pm 0.108	0.047 \pm 0.005	0.463 \pm 0.160%	9.207 \pm 3.709

F. Preliminary Evaluation on Multi-modal Registration

To evaluate the generalizability of the proposed contrast-agnostic representation beyond multi-contrast scenarios, we further conduct preliminary experiments on a multi-modal abdominal CT-MR registration task using the Learn2Reg dataset [41]. Table. IX summarizes the quantitative results. Although AC-CAR was not originally designed for multi-modal registration, the quantitative results indicate that it attains competitive performance across all evaluation metrics on the Learn2Reg dataset. Specifically, AC-CAR achieves the second-highest registration accuracy among the baseline methods, with a Dice score of 0.644 and an HD95 of 9.207, outperforming both MIDIR and SM-Shape. Notably, AC-CAR also yields the lowest gradient of Jacobian and a comparable folding ratio. These findings suggest that, despite the absence of an explicit multi-modal design, AC-CAR exhibits a substantial degree of intrinsic robustness to modality variability. This capability implies that the proposed contrast-agnostic representation has the potential to be extended toward multi-modal deformable registration. Figure 10 presents qualitative CT-MR registration results and the corresponding uncertainty maps for two test pairs. Note that no subject has paired and perfectly aligned CT-MR volumes, making dense error evaluation through error map or sparcification plot infeasible. We therefore only show the uncertainty maps. AC-CAR achieves visually accurate alignment, and the highlighted regions in the uncertainty maps generally correspond to areas of misregistration in the warped images. However, some misaligned regions are not fully captured by the uncertainty estimates, likely because the current contrast augmentation scheme can deal with modality variation to some extent but still has inherent limitations. This limitation will be further investigated in our future work.

VI. DISCUSSION AND CONCLUSION

We present a novel adaptive conditional contrast-agnostic deformable image registration framework (AC-CAR) that can register arbitrary contrasts of images without observing them during training. To achieve this, we propose a random convolution-based contrast augmentation scheme to simulate images with varying contrast as the input to our network. This prevents the network from overfitting a certain distribution

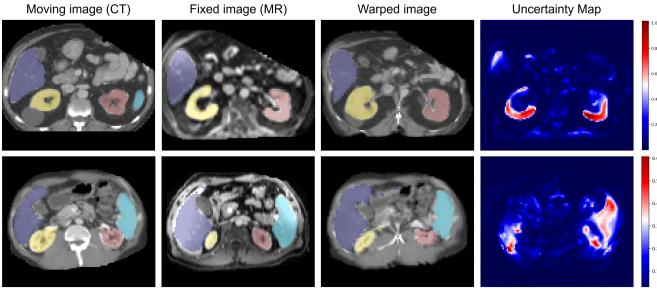


Fig. 10. Qualitative results of AC-CAR on CT–MR abdominal registration and the corresponding uncertainty maps for two test pairs from the Learn2Reg dataset.

of contrast to improve the network’s generalizability. This has been demonstrated by the generalizability experiment in Table. IV, where AC-CAR and CAR perform significantly better than all other baseline approaches. To learn contrast-invariant features, we propose an ACFM to enable the network to modulate the feature adaptively and a CLR to regularize the feature in the latent space. ACFM and CLR can mutually reinforce each other for contrast-invariant feature learning. Meanwhile, we also showed that AC-CAR can provide uncertainty estimation for multi-contrast image registration. Experiments demonstrated that on both 3D inter-subject brain MRI registration and 2D intra-subject cardiac MRI registration tasks, AC-CAR outperformed state-of-the-art methods and exhibited superior generalizability. The estimated uncertainty is highly correlated with the registration error, improving the reliability of AC-CAR.

However, one limitation of this work is that ACFM uses 2D slices for 3D conditioning to improve computational efficiency. Although this design achieves promising results, it may not fully capture the complete volumetric contrast variations present in 3D images. The other limitation of this work relies on the existing contrast augmentation scheme, which only simulates variations between different contrasts and may not capture variations across multiple modalities (e.g., MR-CT or US-CT registration tasks). In future work, we aim to adapt our framework for multi-modal registration.

REFERENCES

- [1] F. Alam, S. U. Rahman, S. Ullah, and K. Gulati, “Medical image registration in image guided surgery: Issues, challenges and research opportunities,” *Biocybernetics and Biomedical Engineering*, vol. 38, no. 1, pp. 71–89, 2018.
- [2] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable medical image registration: A survey,” *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [3] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [4] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [5] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, “Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration,” *Medical image analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.
- [6] M. P. Heinrich, M. Jenkinson, B. W. Papiez, S. M. Brady, and J. A. Schnabel, “Towards realtime multimodal fusion for image-guided interventions using self-similarities,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22–26, 2013, Proceedings, Part I 16*, pp. 187–194, Springer, 2013.
- [7] X. Cheng, X. Jia, W. Lu, Q. Li, L. Shen, A. Krull, and J. Duan, “Winet: Wavelet-based incremental learning for efficient medical image registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 761–771, Springer, 2024.
- [8] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, “Joint learning of motion estimation and segmentation for cardiac mr image sequences,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11*, pp. 472–480, Springer, 2018.
- [9] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen, “Unsupervised deformable registration for multi-modal images via disentangled representations,” in *International Conference on Information Processing in Medical Imaging*, pp. 249–261, Springer, 2019.
- [10] N. Dey, J. Schlemper, S. S. M. Salehi, B. Zhou, G. Gerig, and M. Sofka, “Contrareg: Contrastive learning of multi-modality unsupervised deformable image registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 66–77, Springer, 2022.
- [11] H. Qiu, C. Qin, A. Schuh, K. Hammernik, and D. Rueckert, “Learning diffeomorphic and modality-invariant registration using b-splines,” in *Medical Imaging with Deep Learning*, 2021.
- [12] M. Arar, Y. Ginger, D. Danon, A. H. Bermano, and D. Cohen-Or, “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 (CVPR)*, June 2020.
- [13] M. Hoffmann, B. Billot, D. N. Greve, J. E. Iglesias, B. Fischl, and A. V. Dalca, “Synthmorph: learning contrast-invariant registration without acquired images,” *IEEE transactions on medical imaging*, vol. 41, no. 3, pp. 543–558, 2021.
- [14] V. Sideri-Lampretsa, V. A. Zimmer, H. Qiu, G. Kaassis, and D. Rueckert, “Mad: Modality agnostic distance measure for image registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 147–156, Springer, 2023.
- [15] M. Ronchetti, W. Wein, N. Navab, O. Zettini, and R. Prevost, “Disa: Differentiable similarity approximation for universal multimodal registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 761–770, Springer, 2023.
- [16] D. Rueckert, M. J. Clarkson, D. L. Hill, and D. J. Hawkes, “Non-rigid registration using higher-order mutual information,” in *Medical Imaging 2000: Image Processing*, vol. 3979, pp. 438–447, SPIE, 2000.
- [17] J. P. Pluim, J. A. Maintz, and M. A. Viergever, “Image registration by maximization of combined mutual information and gradient information,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000: Third International Conference, Pittsburgh, PA, USA, October 11–14, 2000. Proceedings 3*, pp. 452–461, Springer, 2000.
- [18] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, “A nonrigid registration framework using spatially encoded mutual information and free-form deformations,” *IEEE transactions on medical imaging*, vol. 30, no. 10, pp. 1819–1828, 2011.
- [19] T. C. Mok, Z. Li, Y. Bai, J. Zhang, W. Liu, Y.-J. Zhou, K. Yan, D. Jin, Y. Shi, X. Yin, et al., “Modality-agnostic structural image representation learning for deformable multi-modality medical image registration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11215–11225, 2024.
- [20] X. Deng, E. Liu, S. Li, Y. Duan, and M. Xu, “Interpretable multi-modal image registration network based on disentangled convolutional sparse coding,” *IEEE Transactions on Image Processing*, vol. 32, pp. 1078–1091, 2023.
- [21] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [22] Y. Wang, S. Du, S. Zheng, X. Luo, and C. Qin, “Car: Contrast-agnostic deformable medical image registration with contrast-invariant latent regularization,” in *International Workshop on Biomedical Image Registration*, pp. 308–318, Springer, 2024.
- [23] P. Risholm, F. Janoos, I. Norton, A. J. Golby, and W. M. Wells III, “Bayesian characterization of uncertainty in intra-subject non-rigid registration,” *Medical image analysis*, vol. 17, no. 5, pp. 538–555, 2013.
- [24] A. Sedghi, T. Kapur, J. Luo, P. Mousavi, and W. M. Wells, “Probabilistic image registration via deep multi-class classification: characterizing

- uncertainty,” in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures: First International Workshop, UNSURE 2019, and 8th International Workshop, CLIP 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings* 8, pp. 12–22, Springer, 2019.
- [25] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, “Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces,” *Medical image analysis*, vol. 57, pp. 226–236, 2019.
- [26] J. Chen, E. C. Frey, Y. He, W. P. Segars, Y. Li, and Y. Du, “Transmorph: Transformer for unsupervised medical image registration,” *Medical image analysis*, vol. 82, p. 102615, 2022.
- [27] X. Zhang, D. H. Pak, S. S. Ahn, X. Li, C. You, L. H. Staib, A. J. Sinusas, A. Wong, and J. S. Duncan, “Heteroscedastic uncertainty estimation framework for unsupervised registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 651–661, Springer, 2024.
- [28] V. Dumoulin, J. Shlens, and M. Kudlur, “A learned representation for artistic style,” in *International Conference on Learning Representations*, 2017.
- [29] A. Achille and S. Soatto, “Emergence of invariance and disentanglement in deep representations,” *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 1947–1980, 2018.
- [30] M. Seitzer, A. Tavakoli, D. Antic, and G. Martius, “On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks,” in *International Conference on Learning Representations*, 2022.
- [31] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [33] M. A. Shafto, L. K. Tyler, M. Dixon, J. R. Taylor, J. B. Rowe, R. Cusack, A. J. Calder, W. D. Marslen-Wilson, J. Duncan, T. Dalgleish, et al., “The cambridge centre for ageing and neuroscience (cam-can) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing,” *BMC neurology*, vol. 14, pp. 1–25, 2014.
- [34] J. R. Taylor, N. Williams, R. Cusack, T. Auer, M. A. Shafto, M. Dixon, L. K. Tyler, R. N. Henson, et al., “The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample,” *neuroimage*, vol. 144, pp. 262–269, 2017.
- [35] B. Fischl, “Freesurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
- [36] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of simpleitk,” *Frontiers in neuroinformatics*, vol. 7, p. 45, 2013.
- [37] C. Ledig, R. A. Heckemann, A. Hammers, J. C. Lopez, V. F. Newcombe, A. Makropoulos, J. Lötzönen, D. K. Menon, and D. Rueckert, “Robust whole-brain segmentation: application to traumatic brain injury,” *Medical image analysis*, vol. 21, no. 1, pp. 40–58, 2015.
- [38] C. Wang, Y. Li, J. Lv, J. Jin, X. Hu, X. Kuang, W. Chen, and H. Wang, “Recommendation for cardiac magnetic resonance imaging-based phenotypic study: imaging part,” *Phenomics*, vol. 1, pp. 151–170, 2021.
- [39] C. Wang, J. Lyu, S. Wang, C. Qin, K. Guo, X. Zhang, X. Yu, Y. Li, F. Wang, J. Jin, et al., “Cmrxecon: an open cardiac mri dataset for the competition of accelerated image reconstruction,” *arXiv preprint arXiv:2309.10836*, 2023.
- [40] X. Luo and X. Zhuang, “ \mathcal{X} -metric: An n-dimensional information-theoretic framework for groupwise registration and deep combined computing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 9206–9224, 2023.
- [41] A. Hering, L. Hansen, T. C. Mok, A. C. Chung, H. Siebert, S. Häger, A. Lange, S. Kuckertz, S. Heldmann, W. Shao, et al., “Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 3, pp. 697–712, 2022.
- [42] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [43] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.
- [44] B. Kim, Y. Zhuang, T. S. Mathai, and R. M. Summers, “Otmorph: Unsupervised multi-domain abdominal medical image registration using neural optimal transport,” *IEEE Transactions on Medical Imaging*, vol. 44, no. 1, pp. 165–179, 2025.
- [45] R. Zhang, H. Mo, J. Wang, B. Jie, Y. He, N. Jin, and L. Zhu, “Utsrmorph: A unified transformer and superresolution network for unsupervised medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 44, no. 2, pp. 891–902, 2025.
- [46] M. Poggi, F. Aleotti, F. Tosi, and S. Mattoccia, “On the uncertainty of self-supervised monocular depth estimation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3227–3237, 2020.