Deep learning in medical image registration

Progress in Biomedical Engineering

OPEN ACCESS

RECEIVED
11 August 2020

REVISED
15 October 2020

ACCEPTED FOR PUBLICATION
14 December 2020

PUBLISHED
5 February 2021

TOPICAL REVIEW

Deep learning in medical image registration

Xiang Chen[1], Andres Diaz-Pinto[1,2], Nishant Ravikumar[1,2] and Alejandro F Frangi[1,2,3,*]

[1] Centre for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), School of Computing, University of Leeds, Leeds, United Kingdom
[2] Leeds Institute of Cardiovascular and Metabolic Medicine (LICAMM), School of Medicine, University of Leeds, Leeds, United Kingdom
[3] Department of Electrical Engineering and Department of Cardiovascular Sciences, Katholieke Universiteit Leuven, Leuven, Belgium
[*] Author to whom any correspondence should be addressed.

E-mail: a.frangi@leeds.ac.uk

Keywords: deep learning, medical image registration, review

Abstract

Image registration is a fundamental task in multiple medical image analysis applications. With the advent of deep learning, there have been significant advances in algorithmic performance for various computer vision tasks in recent years, including medical image registration. The last couple of years have seen a dramatic increase in the development of deep learning-based medical image registration algorithms. Consequently, a comprehensive review of the current state-of-the-art algorithms in the field is timely, and necessary. This review is aimed at understanding the clinical applications and challenges that drove this innovation, analysing the functionality and limitations of existing approaches, and at providing insights to open challenges and as yet unmet clinical needs that could shape future research directions. To this end, the main contributions of this paper are: (a) discussion of all deep learning-based medical image registration papers published since 2013 with significant methodological and/or functional contributions to the field; (b) analysis of the development and evolution of deep learning-based image registration methods, summarising the current trends and challenges in the domain; and (c) overview of unmet clinical needs and potential directions for future research in deep learning-based medical image registration.

1. Introduction

Medical image registration has been a central component of various applications in medical image analysis over the last three decades. The field has evolved immensely with growth in computational resources, and algorithmic capabilities and complexities. Various clinical applications involving disease diagnosis and monitoring, image-guided treatment delivery, and post-operative assessment, utilise image registration. It is also widely used as a tool to preprocess data for subsequent tasks such as object detection, segmentation or classification, as variation in spatial resolution of medical images is very common. Consequently, the performance of the latter is heavily influenced by the quality of the image registration algorithm used to bring the images to a common co-ordinate frame, and fixed size and resolution.

1.1. Framework of registration

Image registration is the process of identifying a spatial transformation that maps two (pair-wise registration) or more (group-wise registration) images to a common co-ordinate frame such that corresponding anatomical structures are optimally aligned, or in other words, a voxel-wise 'correspondence' is established between the images. Depending on the degrees of freedom associated with the desired spatial transformation, image registration algorithms may be broadly grouped into rigid, affine or non-rigid/deformable. In the case of pair-wise image registration, this can be formally defined as follows: Let F and M denote the fixed and moving images, respectively, and T be the desired spatial transformation that maps the voxels of M to those of F. Registering the two images can be posed as an optimisation problem expressed as:

$$\widehat{T} = \arg\min_{T} \mathcal{S}(F, T(M)), \qquad (1)$$

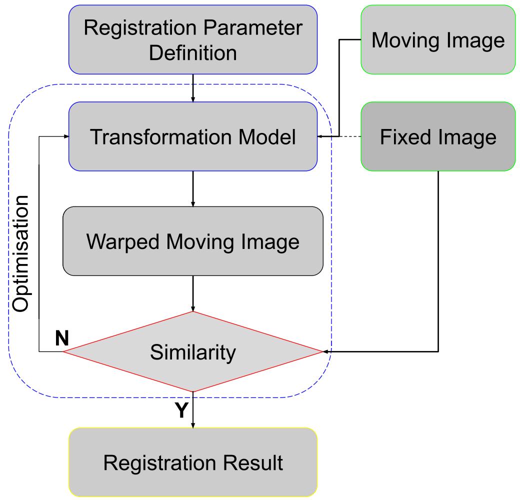

Figure 1. A flowchart of medical image registration framework.

where $\mathcal{S}()$ represents a measure of dissimilarity (or similarity depending on the formulation of the objective function) between the fixed image and the warped moving image. The images are registered by iteratively improving estimates for the desired T, such that the defined $S()$ in the cost function is either maximised or minimised.

Intuitively, non-rigid or deformable image registration is an ill-posed problem, which makes it fundamentally different from other computer vision tasks such as object localisation, segmentation or classification. For example, given two images as input, deformable image registration aims to find a spatial transformation that warps the moving image to match the fixed image as closely as possible. However, there is no ground-truth available for the desired deformation field and without enforcing any constraints on the properties of the spatial transformation, the resulting cost function is ill-conditioned and highly non-convex. In order to address the latter and ensure tractability, all image registration algorithms regularise the estimated deformation field, based on some prior assumptions on the properties of the underlying unknown deformation.

Conventionally, medical image registration algorithms comprise three distinct components: a transformation model, similarity metric, and an optimisation algorithm, as illustrated in figure 1.

The overall process of image registration involves: (1) design/choice of a suitable transformation model (rigid, affine, or non-rigid) and initialisation of its associated parameters, (2) use of the transformation model to warp the moving image, (3) evaluation of the dissimilarity between the warped moving image and the fixed image, and (4) update of the parameters in the transformation model by optimising the cost function formulated using the dissimilarity metric, using a suitable optimisation algorithm. The registration algorithm iterates between step (2) and step (4) until a suitable convergence criterion is satisfied (usually based on the change in the dissimilarity metric or the transformation parameters across iterations). As image registration using conventional algorithms is an iterative process, they are typically computationally intensive and time-consuming. The overall framework is generic and can be formulated within a deep learning (DL) setting, enabling significant acceleration, for registering a pair or group of unseen images using a trained registration network.

1.2. Basic deep learning networks

Although the theoretical concepts that underpin neural networks have existed for decades, early attempts to train such algorithms [1, 2] were constrained by the limited computational power available at the time. Recent years have witnessed an almost exponential growth in the development and use of DL algorithms, sustained thus far by rapid improvements in computational hardware (e.g. GPUs). Consequently, clinical applications requiring image classification, segmentation, registration, or object detection/localisation, have witnessed significant improvements in algorithmic performance, in terms of accuracy and/or efficiency. Although DL-based medical image registration algorithms are yet to achieve significant breakthroughs in terms of registration accuracy compared with traditional methods, they have provided the means to accelerate registration many-fold. To offer a basis to understand deep learning-based registration methods, we briefly introduce and discuss three fundamental and widely used components of image registration

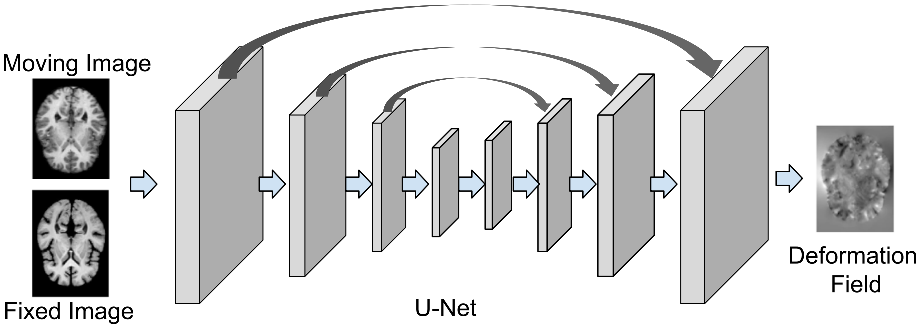

Figure 2. An example of the U-Net framework used for brain MRI image registration. The moving image and fixed image are concatenated at first. A U-Net takes it as input and predicts the deformation field. The U-Net is an encode-decoder network.

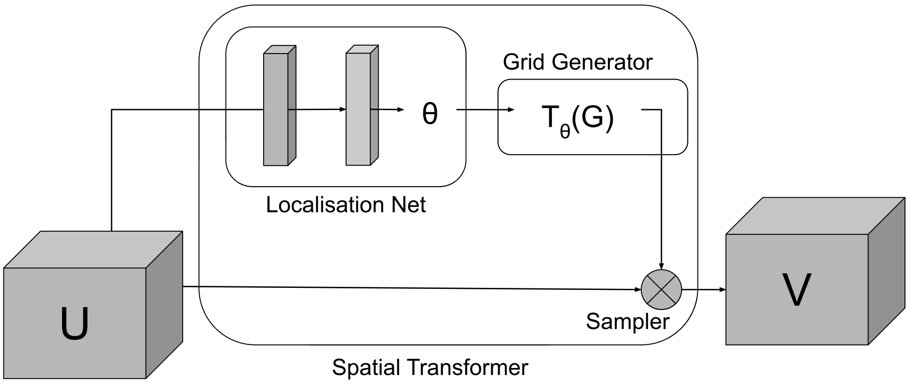

Figure 3. Illustration of spatial transfer network, which consists of three sub-blocks: localisation net, grid generator and sampler. U and V means the input and output respectively. Localisation net is used to learn the transformation parameter θ from U. With θ, the grid generator can generate transformation grid. Then a bilinear sampling is applied to sample coordinates from U to V. The whole network is differential so the back-propagation could update the parameters automatically.

networks, namely, an encoder–decoder Convolutional Neural Network (CNN), a Spatial Transformer Network (STN) [3], and a Generative Adversarial Network (GAN) [4].

The success of DL in visual recognition tasks can be attributed primarily to CNNs. This type of DL network comprises a hierarchical structure of replicated feature detectors, or in other words, successive 'convolution' layers that are used to learn task-specific multi-scale features automatically. Several CNN architectures have been proposed in recent years, each with specific architectural modifications to address the issue of vanishing/exploding gradients common to deep networks, such as AlexNet [5], VGG [6], ResNet [7], and DenseNet [8]. Among these, in medical image segmentation and registration, the most widely used architecture is the U-Net [9]—an encoder–decoder style network with skip connections between the encoding and decoding paths (as depicted in figure 2). The encoder contains several convolutional layers and pooling layers, which downsample the input image to a low resolution. While, the decoder is made up of deconvolution layers with a matching number of layers to the encoder. Through the decoder, the feature maps are reconstructed to the original size of the input images. The U-Net utilises several down- and up-sampling layers to learn features at different resolutions, at the limited expense of computational resources. It has been widely applied in various medical imaging applications (e.g. segmentation), and due to its flexibility, most state-of-the-art Deep Learning-based medical Image Registration (DLIR) methods utilise it as well in some component of the overall framework.

Another core component of most DLIR approaches is STN, proposed in 2015 [3], which learns to spatially transform feature maps in a manner beneficial to the task of interest. Although they were not explicitly designed for image registration, but rather to imbue networks with the means to learn features in a manner invariant to rigid and deformable transformations, they have become the basis for most unsupervised registration methods. As shown in figure 3, STN includes three components: a localisation network, a grid generator and a sampler. The localisation network is a CNN, which takes feature maps as input and outputs the parameters of a suitable/user-specified spatial transformation. The transformation parameters are subsequently used to generate a resampling grid by the grid generator, following which differentiable image sampling is performed by a linear sampler using the grid generated in the previous step.

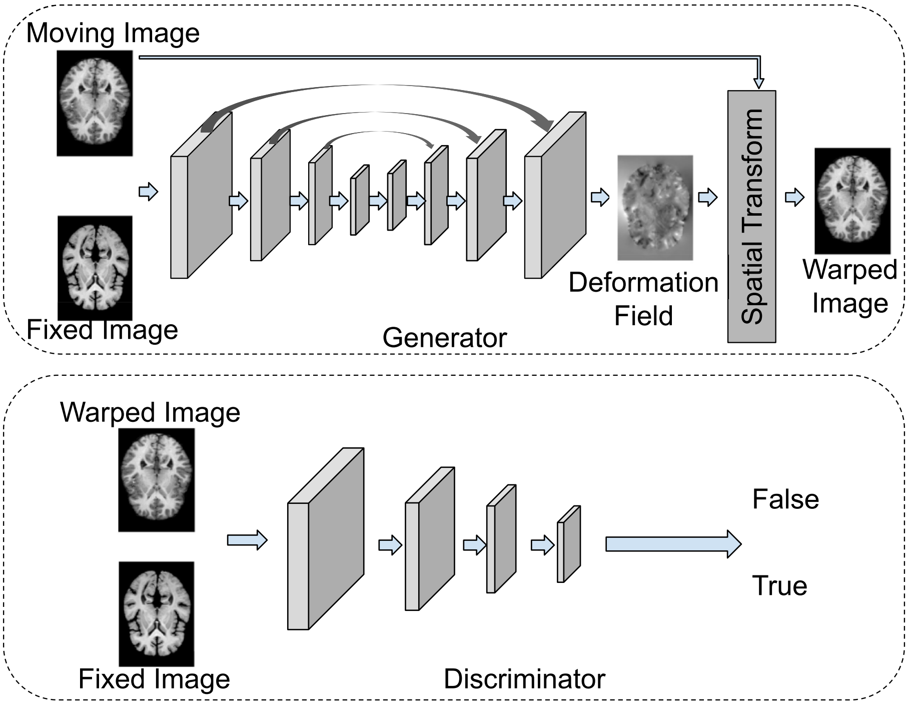

Figure 4. An example of GAN-based image registration. The generator combines a U-Net and an STN to synthesise the deformation field and the warped moving image simultaneously. The discriminator is used to discriminate the difference between warped moving image and fixed image, urging the generator to predict high-similarity warped moving image to the fixed image.

For 3D rigid registration, the spatial transformation is composed of just six parameters, namely, three rotation and three translation parameters. In the case of non-rigid registration, the localisation network estimates a deformation field represented in a parametric or non-parametric form, as defined by the user, of the same size as the input. Most DLIR methods could be seen as expanded STNs, which look to improve the performance of the localisation network to generate more accurate deformation fields for warping the moving image. As with conventional image registration algorithms, the objective function optimised in image registration networks is a similarity/dissimilarity metric computed between the warped moving image and the fixed image, in addition to suitable regularisation terms which ensure that the problem is suitably constrained and well-posed. The latter also controls the smoothness of the estimated deformation field.

As STN gives neural networks the ability to spatially transform feature maps, it has become the basis of most of the DLIR methods, especially unsupervised/weakly-supervised DLIR methods. The generator in figure 4 could be seen as a general DLIR framework, which consists of a CNN (U-Net) and spatial transform block (refer to STN). The CNN takes the moving image and fixed image as input and predicts a deformation field (deformable registration), then the spatial transform block deforms the moving image based on the predicted deformation field. The registration networks are thus formulated as end-to-end networks that utilise CNN and STN to jointly estimate the desired deformation fields and warp the moving images. The losses of similarity/dissimilarity (between warped moving images and fixed images) and regularisation (on deformation fields) would be computed to update the parameters in the CNN. Once the network is trained, the registration between new image pairs is just one forward prediction.

Generative adversarial networks [4] are also a common component of DLIR approaches. They are the most widely used generative models for image synthesis [10–12] and have found use in the medical domain as tools for data augmentation [13], and for applications requiring image-to-image translation [14], and segmentation [15], among others. It contains two parts, a generator and a discriminator, both of which are typically convolutional neural networks. The former constitutes the generative model in the network, which learns to sample from the data distribution and can be used to synthesise new instances. The latter on the other hand, is used to distinguish between synthesised and real samples, thereby competing with the generator, or in other words, acting as its 'adversary'. Essentially, GANs are trained in a minimax two-player game, where the generator looks to maximise the probability of the discriminator mistaking a synthesised sample as a real one from the training data. This leads to both networks learning hierarchical representations of the training data in an unsupervised fashion. A generic GAN-based registration framework is shown in figure 4. With the fixed image and moving image as input, the generator predicts the warped image. Then the

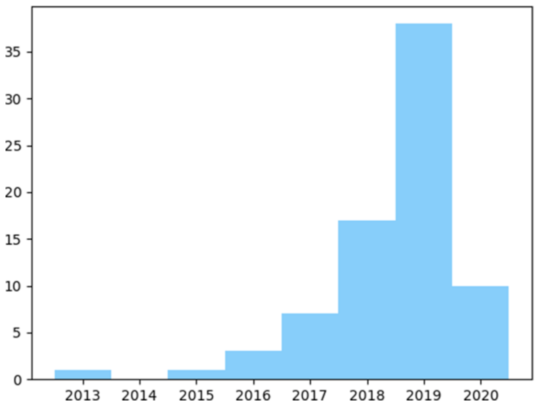

Figure 5. An overview of the number of papers published from 2013 to 2020 about DLIR methods.

discriminator evaluates how similar the warped image is to the fixed image. The discriminator in GANs offers a novel learnable mechanism to evaluate the similarity between two images. This property carries significant potential for building adaptable and learnable similarity metrics, especially relevant for multi-modal image registration. In numerous multi-modal registration approaches, GAN-based image translation networks (e.g. Cycle-GANs [11]) learn to map the appearance shift between domains, i.e. between images from different modalities. This simplifies the task of choosing a suitable similarity metric, by transforming the multi-modal registration problem to a monomodal one. Consequently, GAN-based networks are widely used in medical image registration, which we discuss in more detail in section 2.

The aim of this review is to provide a critical overview of existing literature on DL-based image registration, by highlighting innovations from a methodological and functional perspective, discussing current trends, challenges and limitations, and providing insights to the possible directions for future research. While there have been several review papers published recently on DL-based medical image registration [16–19], they primarily focus on the architecture of networks proposed for DL-based medical image registration, grouping and discussing them according to their design and learning paradigms (i.e. supervised, weakly-supervised or unsupervised, for example). Consequently, in this review, we provide an up-to-date detailed account of both the methodological and functional contributions of DLIR techniques proposed thus far. To facilitate benchmarking of existing DLIR approaches and provide future works with a frame of reference for comparison, we also present a comprehensive summary of publicly available datasets used to design and validate numerous DLIR methods and provide links for all methods with publicly available code. We include 77 papers focusing on DLIR in this review, with the majority published after 2016. The increasing adoption of DL for medical image registration is highlighted by the yearly count described by figure 5. Throughout the review we provide statistics of the number of DLIR papers published, grouped according to their methodological and functional characteristics. We restrict the focus of this review to publications concerned with medical image registration alone. To identify relevant publications, PubMed and Web of Science were queried for papers using combinations of terms such as—DL, medical image registration, deformable image registration, image fusion, multi-modal image registration, motion tracking, among others. In addition to these databases, other sources such as Google Scholar[4], ArXiv[5] and Semantic Scholar[6] were also searched in the same way, and publications with significant contributions to the community, were selected for review.

The remainder of this review is organised as follows: in section 2, we discuss how DL networks are applied in medical image registration. Section 3 describes those methods from the perspective of applications. Sections 4 and 5 discuss the development trends, main challenges/limitations, and possible directions of innovation for DL-based medical image registration.

2. Deep learning-based medical image registration methods

The fundamental building blocks of image registration are identical in both traditional and DL-based approaches, comprising, a similarity metric, transformation model and an optimiser. Neural networks have

[4] https://scholar.google.co.uk/.

[5] https://arxiv.org/.

[6] https://www.semanticscholar.org/.

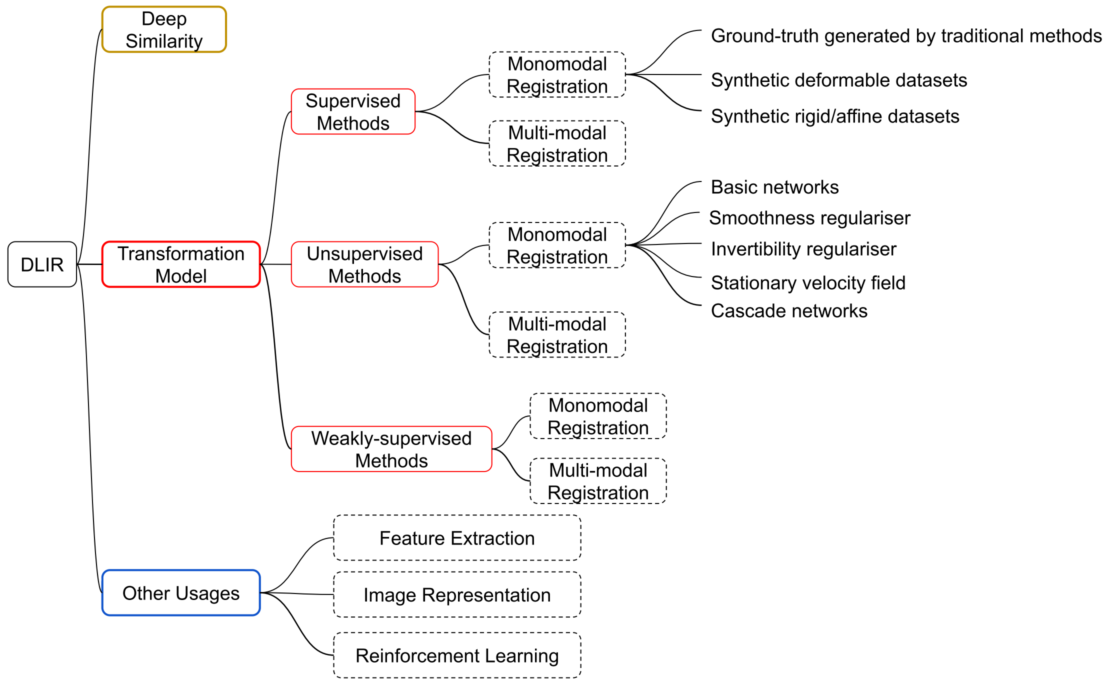

Figure 6. Categories of deep learning-based image registration methods.

been integrated into this framework replacing/enhancing the role played by one or several of these components. We categorise DLIR methods into three parent classes, namely, approaches that (a) use neural networks as a similarity metric (often called deep similarity); (b) parameterise the transformation model using neural networks; and (c) employ neural networks to facilitate other operations (such as feature extraction or learning new image representations, referred to as other usages in this paper) that improve registration quality. Each of these categories can be further divided into sub-groups as described by figure 6, and will be discussed in subsequent sections.

2.1. Deep learning for similarity metrics

In traditional medical image registration methods, studies often focus on improving the similarity metric to obtain higher registration accuracy. Various similarity metrics have been employed in previous studies, such as Cross-correlation (CC), Mutual Information (MI) and Dice Similarity Coefficient (DSC), corresponding to different scenarios, without sufficient justification for their choice in many cases. I.e. these similarity metrics were not application or image modality-specific as they were not learned from or designed for the images to be registered. Visual recognition and perception tasks have benefited substantially from the ability of deep neural networks (specifically, convolutional neural networks) to extract features and combine them across multiple scales, providing the possibility to evaluate the distance between images from different modalities, in a common feature space. Several studies [20, 21] have used neural networks as learnable, data-driven interpretations of similarity metrics, thereby providing a framework adaptable to different applications and image modalities.

DL-based similarity metrics are usually employed for multi-modal image registration, due to the substantial variation in the appearance and intensity distributions of the moving and fixed images. For instance, a similarity metric based on a regression CNN was proposed by Haskins *et al* [20] to register Magnetic Resonance Imaging (MRI) and Transrectal Ultrasound (TRUS) images, which demonstrated promising performance compared with MI, and several other conventional similarity metrics. Deep CNN-based similarity metrics have also been demonstrated to be useful for monomodal image registration. For instance, Zhu *et al* [21] used a pre-trained CNN as a similarity metric for Ultrasound (US) image registration, showing comparable or better performance than manual registration.

Additionally, the formulation of the discriminator in GANs, naturally lends itself to use as a similarity metric, as its role in distinguishing between generated and real images can easily be reformulated as one of computing the difference between the warped and fixed images. Such metrics are often referred to as adversarial similarity, and have been employed in several unsupervised image registration networks [22–25].

While deep neural networks used in this context provide improvements over conventional similarity metrics in terms of robustness and flexibility, the image registration process is still iterative. Consequently, although methods within this category achieve comparable or better registration accuracy than conventional approaches, they are still time-consuming during inference.

2.2. Deep learning for transformation models

In this section we discuss approaches that parameterise spatial transformations using deep neural networks. As described by figure 6, this category of approaches can be further divided into supervised, weakly-supervised and unsupervised approaches, based on the learning paradigm used to train the networks. The fundamental advantage of this group of techniques over conventional approaches and deep similarity networks is the substantial acceleration they afford during inference, enabling real-time rigid and non-rigid image registration.

2.2.1. Supervised registration

This sub-group of techniques employ deep neural networks to estimate the spatial transformation parameters necessary to register two (or a group of) images, in a 'supervised' fashion, i.e. using ground-truth/target values for the parameters to guide the learning process. As with other supervised learning approaches common to medical image analysis tasks such as segmentation or classification, such techniques depend on the availability of ground-truth/target values for the transformation parameters. In general, there are two methods to obtain these target parameters: (a) by estimating them using traditional registration methods; or (b) using simulated images with known ground-truth transformations. Supervised registration networks thus estimate the parameters associated with the transformation model adopted (rigid or non-rigid) to warp the moving image to the fixed image space, and subsequently, compute the loss between predicted parameters and ground-truth values. This loss over the transformation parameters, in turn, is used to compute its gradients with respect to the weights of the network, which parameterise the spatial transformations, and is used to guide the training of the network. Following training, registration of two or more images is achieved as a single forward pass through the network, substantially reducing the execution time relative to iterative approaches.

Table 1 summarises the most relevant supervised DL-based medical image registration methods we identified for this review. In order to provide readers with operationally useful information, we also provide links to repositories for all methods that have made their code publicly available. We further group supervised methods in to monomodal registration and multi-modal registration. Monomodal registration also called uni-modal registration, aims to register moving images and fixed images from the same modality such as MRI, Computed Tomography (CT) and x-ray. Multi-modal registration is applied to register images from different modalities (e.g. CT to MRI, x-ray to MRI). We found that a large proportion of existing supervised DLIR methods are monomodal (refer to table 1). As obtaining ground-truth transformations is a key problem for supervised registration methods, we further classify the monomodal registration methods in to three classes: (a) generating them using traditional registration methods; (b) using synthetic datasets with known ground-truth deformation fields (for non-rigid registration); and (c) generating synthetic datasets with rigid/affine transformations (for rigid/affine registration).

2.2.1.1. Ground-truth generated by traditional methods

In 2016, Yang *et al* [26] proposed a supervised encoder–decoder network for Large Deformation Diffeomorphic Metric Mapping (LDDMM) registration, which utilised PyCA[7] LDDMM to generate ground-truth deformations. Their approach was shown to substantially accelerate registration and achieve lower registration error, compared with traditional methods. Similarly, Cao *et al* [27] designed a 3D patch similarity-steered CNN regression network for brain MRI registration, which used Symmetric Normalisation (SyN) and diffeomorphic Demons to generate ground-truth deformation fields. Their final registration results obtained a higher DSC than SyN and Demons. They also proposed a key-point truncated-balanced sampling strategy and a cue-aware deep regression network to enhance registration generalisation, which tackled various registration tasks on different databases [28]. With ground-truth generated by Advanced Normalisation Tools (ANTs) [44] and LCC-Demons [45], Fan *et al* [29] proposed a dual-guidance network BIRNET which involved two losses to guide the training process: the distance between generated deformation fields and ground-truth, and the dissimilarity between fixed image and warped moving image.

[7] https://bitbucket.org/scicompanat/pyca.

Table 1. A summary of supervised DL-based registration methods. All the supervised methods are firstly classified into two general classes: monomodal registration and multi-modal registration. Then monomodal registration methods are categorised further according to the methodology of obtaining ground-truth. Hyperlinks are given for those works with code publicly available.

Registration	Reference	Network	Modality	Dimension	Organ	Code
	Ground-truth generated by traditional methods					
Monomodal	Yang *et al* [26]	Encoder–decoder	MRI	2D, 3D patch	Brain	link
	Cao *et al* [27]	Similarity-steered CNN regression	MRI	2D	Brain	—
	Cao *et al* [28]	Cue-aware deep regression network	MRI	3D patch	Brain	—
	Fan *et al* [29]	U-Net+hierarchical dual-supervision	MRI	3D	Brain	—
	Synthetic deformable datasets					
	Rohe *et al* [30]	SVF-Net (U-Net)	MRI	3D	Heart	—
	Eppenhof *et al* [31]	U-Net	CT	3D	Lung	—
	Eppenhof *et al* [32]	Progressive U-Net network	CT	3D	Lung	—
	Sokooti *et al* [33]	RegNet	CT	3D patch	Chest	link
	Synthetic rigid/affine datasets					
	Mohseni *et al* [34]	18-layer residual CNN	MRI	3D	Foetal brain	—
	Xia *et al* [35]	Cascaded CNN	Plantar pressure image (PPI)	2D	Plantar	—
	Zhao *et al* [36]	10-layer CNN	MRI, CT	2D, 3D	Brain, Lung	—
Multi-modal	Yang *et al* [37]	Bayesian encoder–decoder network	MRI	3D patch	Brain	—
	Yang *et al* [38]	Encoder–decoder	MRI	3D patch	Brain	link
	Yan *et al* [39]	GAN	MRI, TRUS	3D	Prostate	—
	Sedghi *et al* [40]	3D classification CNN	MRI	3D	Brain	—
	Yao *et al* [41]	CIR	CT, CBCT	3D	Head, Abdomen, Chest, Pelvic	—
	Liao *et al* [42]	POINT	x-ray, CBCT	2D, 3D	Whole body	—
	Liao *et al* [43]	MSReg	MRI, TRUS	3D	Prostate	—

2.2.1.2. Synthetic deformable datasets

Instead of generating ground-truth deformations using traditional registration methods, Sokooti *et al* [33] utilised artificially generated Displacement Vector Fields (DVF) as ground-truth, and designed a network ‘RegNet’ for chest CT image registration. They proved that the trained model could be applied to real data and obtained registration results on par with a conventional B-spline registration approach. Eppenhof *et al* [31] proposed a U-Net based registration network trained on synthetically deformed clinical images, with augmentation transformations to aid in generalisation. Similarly, they generated a large number of ground-truth data by applying random synthetic transformations to a training set of images and proposed a progressive learning network, which enabled training on large and small transformations within the same CNN [32]. Rohe *et al* [30] proposed to derive a reference Stationary Velocity Field (SVF) deformation using segmented shapes. Using the obtained reference SVF as the ground-truth, they designed a 3D U-Net based network SVF-Net for cardiac MRI image registration.

2.2.1.3. Synthetic rigid/affine datasets

The ground-truth for rigid/affine registration is much easier to synthesise as random combinations of operations such as rotation, translation and scaling would be sufficient to generate data required to train a network. Besides, unlike the non-rigid transformations, most rigid transformation parameters could be obtained manually. Though this task is much easier than non-rigid registration, a few studies have investigated the use of DLIR for rigid registration. For instance, Salehi *et al* [34] proposed an 18-layer residual CNN regression model for 3D pose estimation, and rigidly registered reconstructed foetal brain MRI images to a standard (atlas) space. While, based on images generated by the four transformations (i.e. scaling, horizontal or vertical shift and rotation), Xia *et al* [35] proposed a two-level cascade CNN for plantar pressure image registration. To capture large and complex deformations, Zhao *et al* [36] proposed a 10-layer CNN to estimate the rotation parameters (360 classes) and initialise the subsequent registration step. They utilised the Demons algorithm for non-rigid registration, and achieved substantial improvements in registration accuracy over previous approaches.

2.2.1.4. Multi-modal registration

Supervised DL networks have also been employed for multi-modal image registration. As in their previous study [26], Yang *et al* [37] utilised PyCA to obtain ground-truth deformation fields and proposed a 3D Bayesian encoder–decoder network to estimate the momentum fields for brain MRI multi-modal image registration. Furthermore, they developed an approach applicable to both monomodal and multi-modal registration called ‘Quicksilver’ [38], which combined a registration and correction network for LDDMM registration. Using images aligned manually by experts as ground-truth, Yan *et al* [39] proposed a GAN-based multi-modal image registration method called ‘AIR-Net’, which estimated the transformation parameters directly with an efficient forward pass of the generator and additionally evaluated the quality of registration using the discriminator. Different from general DL methods predicting displacement field directly, Sedghi *et al* [40] used a deep multi-class classifier to predict a collection of discrete displacements between patches. They obtained the final registration results by iterations.

A few approaches have also focused on rigid multi-modal image registration, for example—Yao *et al* [41] utilised a regression CNN for coarse rigid registration, which subsequently initialised a conventional intensity-based registration method for fine-grained registration. This approach combined CNNs with conventional methods to align 3D CT and CBCT images. Liao *et al* [42] proposed a novel learning-based multiview 2D–3D rigid registration method that directly measured the 3D misalignment using a Point-Of-Interest Network for Tracking (POINT), and found the point-to-point correspondence between two images. To tackle the task of rigid MRI-TRUS registration on prostate images, Guo *et al* [43] proposed a new strategy to generate augmented datasets, and designed a coarse-to-fine multi-stage network, which significantly reduced the registration error than previous methods.

2.2.2. Unsupervised learning methods

Although supervised DLIR methods have been shown to substantially accelerate registration, and achieve comparable accuracy to traditional methods, the difficulty in obtaining plausible ground-truth transformations is a fundamental challenge and limitation of this group of methods. Methods used to obtain ground-truth transformations typically result in implausible or over-simplified transformations, or are constrained by the performance of the traditional registration methods used to estimate the same. Consequently, in either scenario, the performance of DLIR methods on real data may be limited by the quality of ground-truth transformations available for training. Therefore, researchers have explored unsupervised learning and weakly-supervised learning methods to ameliorate the need for ground-truth. Unsupervised registration networks require only the moving and fixed images for training, while, weakly-supervised approaches (discussed in subsection 2.2.3) require some additional information such as segmentation masks or landmarks, which are much easier to obtain than ground-truth transformations.

Currently, unsupervised methods are the hot topic in medical image registration, as they can predict the deformation fields and warped moving images in just one forward pass, and do not require ground-truth transformations for training. Similar to supervised methods, table 2 gives a summary of the most relevant unsupervised medical image registration methods. As before, we first classify all methods as monomodal or multi-modal. The monomodal methods are further categorised according to the type of regularisation used. Without ground-truth deformation fields, it is difficult for DLIR methods to guarantee diffeomorphic transformations. Therefore, several approaches have been proposed to constrain the estimation of deformation fields and improve their smoothness. To provide an overview of the types of regularisation techniques employed thus far, we group the monomodal unsupervised methods into several sub-classes: (a) basic networks, (b) smoothness regulariser, (c) invertibility regulariser, (d) SVF, and (e) cascade networks.

2.2.2.1. Basic networks

As no ground-truth data is available/used, the first problem to tackle with training unsupervised registration networks, is to formulate a loss function that can be optimised to train the network. Using STN, DL networks can generate deformation fields to warp the moving image. The dissimilarity between the warped moving image(s) and fixed image(s) can subsequently be used to calculate the loss function for back-propagation. This measure of dissimilarity (or similarity) is typically estimated using metrics such as Mean Square Error (MSE) and MI, in traditional registration approaches, and can be employed for DLIR methods as well. This group of networks, which we refer to as 'CNN+STN', form the basis for most DL-based image registration networks.

In 2017, De Vos *et al* [46] were the first to propose an unsupervised end-to-end network, based on CNN and STN, to register 2D cardiac cine MRI images. The registration accuracy of their approach was demonstrated to be comparable to SimpleElastix[8]. Similarly, Jun *et al* [47] proposed a 'CNN+STN' network for 2D abdomen MRI registration, which was the first CNN-based registration method for abdominal images.

2.2.2.2. Smoothness regulariser

Although similarity metrics can guide the training of unsupervised registration networks, previous studies have demonstrated that the estimated deformation fields may contain several regions with 'folds', where the determinant of the Jacobian (of the deformation field) is negative. The proportion of voxels with negative values for the determinant of the Jacobian (or number of folds) is an important criterion used in most DLIR methods to evaluate the smoothness of the predicted deformation fields. Ideally, deformation fields should be diffeomorphic and hence smooth, and invertible. To enforce the estimated deformation fields to be spatially smooth, several researchers [48, 53] have employed various forms of regularisation within the loss function during training. Li *et al* [48] employed the total variation (TV) loss as a smoothness regulariser and designed a multi-resolution FCN to estimate dense deformation fields. Instead of the TV loss, Stergios *et al* [53] proposed a network similar to 'CNN+STN' with L1 regularisation for 3D lung MRI image registration.

Regularisation using L2-norm derivatives of the deformation fields have also been proposed previously [49, 50]. Here, the proposed approach (called 'Voxelmorph') was based on a 'U-Net+STN' framework with different traditional similarity metrics (MSE and CC) for 3D brain MRI image registration. The approach was shown to outperform several traditional registration methods such as SyN [74] and NiftyReg[9]. Following Voxelmorph, Hu *et al* [56] designed a two-stream 3D encoder–decoder network which computed two convolutional feature pyramids separately, and included a pyramid registration module to predict multi-scale registration fields. Similarly, Ali *et al* [55] proposed a novel end-to-end CNN that comprised sequential linear and deformable convolutions along with a learned non-linear sampler. With the same smoothness regulariser, Fan *et al* [22] proposed an adversarial similarity network (combining a registration network and a discrimination network) for brain MRI registration. They further learned a meaningful metric for effective training of the registration network, using the discrimination network. Using a similar smooth loss, Zhu *et al* [51] designed an end-to-end network comprising affine alignment subnetwork and deformable subnetwork, which did not require an additional preprocessing of affine registration before registration. Similarly, Fu *et al* [52] proposed a LungRegNet based on two GAN-based networks to register lung CT images from coarse to fine, where the adversarial network in GANs was used to enforce additional DVF regularisation. Kuang *et al* [54] designed a fast image registration network (FAIM), with two explicit anti-folding regularisation terms, which forced the generated deformation field to be

[8] https://simpleelastix.github.io/.

[9] https://cmiclab.cs.ucl.ac.uk/mmodat/niftyreg.

Table 2. A summary of unsupervised deep learning-based registration methods. Methods are fist classified as monomodal or multi-modal. Monomodal approaches are then further classified into several sub-classes.

Registration	Reference	Networks	Modality	Dimension	Organ	Code
	Basic networks					
Monomodal	De Vos *et al* [46]	DIRNet	Cine MRI	2D	Heart	link
	Jun *et al* [47]	CNN+STN	MRI	2D patch, 2D	Abdomen	—
	Smoothness regulariser					
	Li *et al* [48]	Multi-resolution FCN	x-ray, MRI	3D	Brain	link
	Balakrishnana *et al* [49], [50]	VoxelMorph	MRI	3D	Brain	link
	Fan *et al* [22]	GAN-based registration network	MRI	3D	Brain	—
	Zhu *et al* [51]	Affine subnetwork+Deformable subnetwork	MRI	3D	Brain	—
	Fu *et al* [52]	LungRegNet	CT	3D patch	Lung	—
	Stergios *et al* [53]	CNN+STN	MRI	3D	Lung	link
	Kuang *et al* [54]	FAIM	MRI	3D	Brain	link
	Ali *et al* [55]	Conv2Warp	CT, MRI	3D,4D patch	Lung, Brain	—
	Hu *et al* [56]	Dual-PRNet	MRI	3D	Brain	—
	Bhalodia *et al* [57]	U-Net+Cooperative auto-encoder (CAE)	MRI	2D,3D	Brain, Heart	—
	Sang *et al* [58]	CNN+Convolution auto-encoder	MRI	2D,3D	Heart	—
	Invertibility regulariser					
	Fechter *et al* [59]	U-Net+STN	CT, MRI	3D,4D	Lung, Heart	link
	Mahapatra *et al* [25]	SARNet	x-ray, MRI	2D,3D	Chest, Brain	—
	Gu *et al* [60]	SCC-Net	MRI	3D	Brain	—
	Kim *et al* [61]	Cycle-Consistent CNN	CT	3D	Liver	—
	SVF					
	Dalca *et al* [62], [63]	Voxelmorph-diff (Probabilistic Model)+SVF	MRI	3D	Brain	link
	Krebs *et al* [64]	CVAE+SVF	Cine MRI	3D	Heart	—
	Liu *et al* [65]	CNN (Feature-level Probabilistic Model)	MRI	3D	Brain	—
	Shen *et al* [66]	AVSM	MRI	3D	Knee, Femoral, Tibial cartilage	link
	shen *et al* [67]	3D U-Net+SVF	MRI, CT	2D,3D	Knee, Lung	link
	niethammer *et al* [68]	CNN+vSVF+CNN regulariser	MRI	2D,3D	Brain	link
	Cascade networks					
	De Vos *et al* [69]	DLIR (multi-stage ConvNets)	Cine MRI, CT	3D patch	Heart, Chest	—
	Zhao *et al* [70]	Recursive cascade architecture	CT, MRI	3D	Liver, Brain	link
	Zhao *et al* [71]	Cascading VTN	CT, MRI	3D	Liver, Brain	link
Multi-modal	Cao *et al* [72]	U-Net+STN	CT, MRI	3D patch	Prostate, Bladder, Rectum	—
	Qin *et al* [24]	UMDIR+cross-cycle reconstruction	CT, MRI	3D	Lung, Brain	—
	Fan *et al* [23]	GAN-based registration network	MRI, CT	3D	Brain, Prostate, Bladder, Rectum	—
	Jiang *et al* [73]	MJ-CNN	CT, CBCT	3D	Lung	—

smooth: regularisation for overall smoothness of the predicted displacements, and regularisation for negative Jacobian determinants in the transformation.

In addition to adopting a smoothness enforcing loss, Bhalodia *et al* [57] proposed to simultaneously learn and use the population-level statistics of the spatial transformations to regularise the neural networks. To do this task, they employed a Cooperative Auto-encoder (CAE) on the predicted deformation fields to urge them to lie in the vicinity of a low-dimensional manifold, then the reconstruction loss of the CAE was used as a regulariser term. Similarly, Sang [58] pre-trained a convolutional auto-encoder on 3000 DVF samples obtained by SimpleElastix, and applied it as the regulariser, which improved the physical and physiological feasibility of the DVF.

2.2.2.3. Invertibility regulariser

Although the aforementioned smooth losses contribute to improving the smoothness of deformation fields, they are unable to guarantee an invertible deformation. Hence, several studies have focused on designing invertible frameworks and appropriate losses to tackle this issue. Using a cyclic constraint in the loss, Fechter *et al* [59] presented an approach to calculate DVF for periodic motion tracking in 3D and 4D medical image datasets. This approach was able to calculate the forward and inverse transformation simultaneously. Similarly, using a cycle-consistency loss, Mahapatra *et al* [25] proposed a GAN-based registration network in combination with segmentation information (learned automatically), which could directly transfer the registration model trained on one type of images to another type of images (for example, training on lung x-ray images while registering brain MRI on testing). To improve registration consistency, Gu *et al* [60] designed a Symmetric Cycle Consistency Network (SCC-Net), which introduced pair-wise and group-wise deformation consistency constraints by losses on inverse-consistency and cycle-consistency. Some researchers also proposed to improve the invertibility by network design. Kim *et al* [61] designed a novel registration framework containing two invertible registration networks, where fixed image and moving image were both deformed/warped to match each other, and subsequently, deformed back to the original fixed and moving images.

2.2.2.4. SVF

Smoothness and invertibility regularisation enhance the diffeomorphic properties of spatial transformations. However, they cannot guarantee the prediction of diffeomorphic transformation fields. In theory, SVF and LDDMM can guarantee diffeomorphism [64]. Therefore, instead of predicting regular dense displacement fields, previous studies have opted to predict SVF to guarantee diffeomorphic transformations. Krebs *et al* [64] designed a multi-scale Conditional Variational Auto-encoder (CVAE) to estimate stationary velocity fields, which enabled accurate registration of two images and the analysis of deformations. Similarly, Dalca *et al* [62, 63] proposed a network Voxelmorph-diff, combining diffeomorphic transformations with DL networks, and provided a framework for quantifying registration uncertainty. Following the structure in Voxelmorph-diff to estimate SVF, Liu *et al* [65] developed feature-level probabilistic models to estimate the deformation fields for feature maps/images from multiple layers of two convolutional neural networks, which provided direct regularisation for hidden CNN layers. Shen *et al* [66] developed an end-to-end registration method Affine-vSVF-Mapping (AVSM), using a multi-step Affine-Net to obtain an initial transformation map and a U-Net like network to generate initial momentum. These two outputs were subsequently used as input to the registration component, vSVF, to obtain the final registration fields. Experiments showed that their method achieved higher accuracy and smoother (fewer foldings) fields than Voxelmorph-diff. Based on a vector momentum SVF model, Niethammer *et al* [68] were the first to proposed a CNN-based local regulariser for registration, generating deformation fields with no foldings. The initial momentum could be obtained using various methods, including DLIR methods. For simplicity, we categorised it as an unsupervised DL method. Similar to the method proposed in [66] to obtain the deformation fields, Shen *et al* [67] proposed a region-specific diffeomorphic metric mapping registration technique. They obtained large diffeomorphic deformations with a spatio-temporal regulariser, and achieved higher accuracy than AVSM [66]. Rather than estimating displacement fields, these methods predict SVF/LDDMM and generate smoother fields than previous methods. The benefit of such approaches are that the estimated deformation fields contain only a few foldings, or in some cases are perfectly smooth.

2.2.2.5. Cascade networks

Cascade networks combine several registration networks to obtain the final registration results, often obtaining higher accuracy following several rounds of registration. However, these networks do not guarantee a diffeomorphic transformations. De Vos *et al* [69] proposed a novel registration framework comprising several ConvNets to solve the problem of unsupervised affine and deformable registration. They demonstrated that stacking multiple ConvNets into a more extensive architecture facilitated a coarse-to-fine

Table 3. A summary of weakly-supervised DL methods (categorised as monomodal and multi-modal registration).

Registration	Reference	Network	Modality	Dimension	Organ	Code
Monomodal	Hering *et al* [75]	CNN	Cine MRI	2D	Heart	—
	Balakrishnana *et al* [50]	Voxelmorph	MRI	3D	Brain	link
	Dalca *et al* [63]	Voxelmorph-diff	MRI	3D	Brain	link
	Heinrich *et al* [76]	PDD-Net	CT	3D	Abdominal	link
	Xu *et al* [77]	DeepAtlas (segmentation and registration CNN)	MRI	3D	Knee, Brain	link
	Chen *et al* [78]	Segmentation + Two-stage registration network	CT	3D	Lung	—
	Ha *et al* [79]	U-Net+Two-stage registration network	MRI	3D	Heart	link
	Mansilla *et al* [80]	AC-RegNet	x-ray	2D	Chest	link
Multi-modal	Hu *et al* [81, 82]	Global-Net, Local-Net CNN	MRI, TRUS	3D	Prostate Gland	link
	Hering *et al* [83]	U-Net	MRI, CT	3D	Heart	—

image registration. Zhao *et al* [70] presented a deep recursive cascade architecture for deformable image registration, which could be used to cascade other state-of-the-art networks to improve registration quality. In addition, they further designed a registration framework called Volume Tweening Network (VTN) and incorporated an additional invertibility loss into the training process [71]. They showed that cascaded registration sub-networks improved performance for registering images with large deformations, with minimal increase in computational cost.

2.2.2.6. Multi-modal registration

Unsupervised registration methods, especially GAN-based methods are also widely employed for multi-modal image registration. A common problem in multi-modal image registration is to choose/formulate a suitable metric to evaluate the dissimilarity between images from different modalities. Cao *et al* [72] designed a 'CNN+STN' network for image registration between CT and MRI image. With a prealigned CT and MR dataset (both fixed and moving images are CT-MRI pairs), they proposed an intra-modality similarity metric, turning the dissimilarity between MRI and CT images into a combination of two intra-modality dissimilarity in MRI and CT. Qin *et al* [24] presented a multi-modal deformable image registration method (UMDIR), which learned a bi-directional registration function based on disentangled shape representation. They pre-trained a image-to-image translation network with unpaired data, then used it to train multi-modal registration network and GAN discriminator (to calculate the dissimilarity between images). This method reduced multi-modal image registration to monomodal image registration. Fan *et al* [23] designed a GAN-based network for multi-modal and monomodal image registration between 3D MRI and CT images, designing an adversarial similarity network to learn a meaningful metric for the network training. Focusing on pulmonary CT-CBCT and CBCT-CBCT registration, Jiang *et al* [73] proposed a multi-scale framework called 'MJ-CNN' to prevent the registration network from being trapped in a local minimum, which contained three sub-networks at different scale level (from coarse to fine). They trained these three sub-networks separately first, then jointly trained them in a whole framework.

Compared with the traditional registration methods, unsupervised DLIR methods are significantly faster. Additionally, unsupervised registration networks do not need ground-truth transformations for training, addressing a fundamental limitation of supervised image registration methods. Moreover, numerous approaches [62, 63] have shown that unsupervised methods achieve similar or sometimes better registration performance than traditional state-of-the-art registration methods. Consequently, current research in the field is predominantly focused on improving the performance and expanding the capabilities of unsupervised image registration techniques.

2.2.3. Weakly-supervised learning methods

As discussed previously, supervised image registration methods require ground-truth deformation fields, which are generally difficult to obtain. In contrast, unsupervised image registration methods disregard all available information and utilise just the fixed and moving images. Consequently, useful information that may help guide image registration is not exploited. To utilise such information (typically encoded as anatomical cues) and improve image registration performance of unsupervised approaches, several weakly-supervised learning methods have been proposed. Table 3 summarises all deep learning-based weakly-supervised registration methods published to date. It is relevant to note that, several studies have proposed both unsupervised registration networks and their weakly-supervised counterparts simultaneously [50, 63]. As done previously for supervised and unsupervised methods, we categorise this group of approaches into monomodal and multi-modal registration, and discuss them accordingly.

2.2.3.1. Monomodal registration

Most weakly-supervised registration networks are similar to unsupervised networks, with the exception that additional information is utilised during training. This additional information is typically encoded as region-wise labels/masks or landmarks, and is only utilised during training. The labels are spatially aligned jointly with the images, by minimising a loss function of the warped moving label and the fixed label. The intuition here is that the labels help preserve anatomical coherence between tissue/organ boundaries, by acting as attention maps that guide the estimation of spatial transformations. These label pairs for the fixed and moving images might include solid organs, ducts, vessels, point landmarks and other ad hoc structures that are deemed relevant to guiding registration. In the reviewed literature, there are mainly two types of labels utilised to guide registration—segmentation masks and landmarks. Both types of labels are used to construct a combined loss that is optimised to match both labels and images, and estimate the desired deformation field. Hering *et al* [75] advanced the state-of-the-art in CNN-based deformable registration by combining a square difference loss between fixed segmentation and warped moving segmentation with similarity between fixed and warped moving images. Following Voxelmorph, Balakrishnan *et al* [50] proposed an extension that incorporated a segmentation loss during training, calculated as the Dice score between the fixed and warped moving segmentation masks. Similarly, Dalca *et al* [63] also built a weakly-supervised version of Voxelmorph-diff by incorporating the surface distance between segmentation results. With an MSE loss on segmentation, Heinrich *et al* [76] designed PDD-Net for monomodal abdominal CT image registration, which combined probabilistic dense displacements with differentiable mean-field regularisation. This approach was shown to outperform previous DL approaches, achieving an improvement of 15% in Dice overlap.

Instead of using segmentation masks as just additional terms to match in the loss function, Xu *et al* [77] proposed the first approach to jointly learn two deep neural networks for simultaneous image registration and segmentation. The registration network and segmentation network can guide each other's training on unlabelled data based on anatomy similarity loss, therefore the proposed method only required a few manual segmentation samples. With a similar idea, Chen *et al* [78] proposed to use semantic information (lung lobules and airway masks obtained from a pre-trained segmentation network) to guide the registration. They designed a two-stage registration network, where the first predicted a coarse deformation filed on segmentation masks while the second was fine registration on vessel structures. Instead of registering images directly, Ha *et al* [79] proposed a semantically guided registration network, which applied a U-Net to extracted semantic features and used a two-stage registration network to predict the final deformation fields based on the extracted semantic features, under the guidance of two losses on segmentation. As applying Dice score on segmentation results does not consider the global context of the anatomical structures, to tackle this issue, Mansilla *et al* [80] proposed to use an auto-encoder to extract the global anatomical features from fixed and warped moving masks, then computed the squared Euclidean distance on them as an additional global loss, which helped to predict more realistic and accurate results.

2.2.3.2. Multi-modal registration

Weakly-supervised registration methods have also been employed for multi-modal registration. Hu *et al* [81] introduced a flexible framework that could utilise all types of anatomical labels for T2W-TRUS multi-modal registration. They proposed a network combined global-net (affine registration) and local-net (deformable registration), significantly outperformed a separate global-net or local-net. Based on the reviewed literature, this is the first DLIR method to utilise weak labels to guide image registration. Using segmentation masks for the whole heart in CT and MRI scans, Hering *et al* [83] combined three 2D networks to construct a 2.5D registration approach, for cardiac MRI-CT registration. They demonstrated that their approach achieved a higher Dice score than previous state-of-the-art unsupervised registration methods.

2.3. Other usages

Besides predicting similarity metrics and transformation fields, deep neural networks have been used in other ways to facilitate image registration, such as: feature extraction, learning new image representations, reinforcement learning, among others. Table 4 summarises these other usages of DL networks for medical image registration. The majority of approaches thus far have employed DL networks to either: (a) learn feature maps for the input moving images and fixed images or (b) learn a new image representations (transfer the original images to new images which are more convenient for registration, for example, learn a clean image from noisy image, or transfer fixed and moving images to same modality in multi-modal registration) for the original fixed images and moving images. We discuss the details of these methods in subsequent sections.

2.3.1. Feature extraction

As DL networks have proven to be efficient at feature extraction, a few early studies [84, 85] first utilised DL networks for feature extraction, and subsequently applied traditional registration methods using the obtained features. Wu *et al* [84] built a stacked convolutional independent subspace analysis network to learn the hierarchical basis filters from several image patches in the brain MRI. They applied the HAMMER [99] for registration, achieving better registration performance than other HAMMER-based methods. Based on a similar idea, they also designed a stacked auto-encoder to learn latent feature representations for 3D medical image patches [85]. Kearney *et al* [86] proposed a Deep Convolutional Inverse Graphics Network (DCIGN) to extract hierarchical features as input channels to a sparse Deformable Image Registration (DIR) algorithm for registering CBCT to CT images. Blendowski *et al* [88] proposed a CNN-based approach for learning discriminative 3D binary descriptors. Focusing on multi-modal registration, Zhu *et al* [87] designed a novel structural representation method based on PCANet [100] to learn intrinsic image features automatically. Subsequently, the spline-based Free-form Deformation (FFD) was applied to register the images, obtaining lower Target Registration Error (TRE) than traditional state-of-the-art methods. Besides, Canalini *et al* [90] firstly proposed a segmentation-based registration method, combining a 3D U-Net for segmentation and a traditional registration method, which registered US volumes acquired at different surgical stages. To transfer the model trained on source domain (i.e. synthetic data) to target domain (i.e. clinical data), Zheng *et al* [89] proposed a pair-wise domain adaptation (PDA) module to tackle the domain shifting problem for CNN-based 2D–3D registration, which learned domain invariant features using only a few paired real and synthetic data. Experiments showed that they obtain better performance than fine-tuning, using the same pre-trained registration model.

2.3.2. Image representation

Given the fixed and moving images, most previous studies focus on improving the performance of a component in the registration algorithm, and often overlook the quality of the given images. However, even in several well-curated publicly available datasets, low quality images resulting from tissue, motion or scanner-related artefacts are prevalent. This adversely affects the accuracy of the final registration, unless addressed adequately. Consequently, given such low quality images, generating new image representations with prominent distinguishable anatomical features is essential to ensure high registration accuracy. Additionally, in the context of multi-modal registration, shifting the domain of the fixed and moving images to a single modality, would simplify the registration task. To this end, several studies have proposed to utilise DL networks for learning new representations of the images to be registered. Yang *et al* [91] proposed an encoder–decoder network to learn a mapping from pathological images to quasi-normal images. Subsequently, they utilised NiftyReg for registration and demonstrated superior registration performance compared with other state-of-the-art approaches. Lee *et al* [94] proposed an image-and-spatial transformer network to learn a new image representation for the downstream registration task (using STNs). They showed that their approach outperformed both unsupervised and supervised STNs.

Using DL networks to learn new image representations also attracts much attention in multi-modal registration. Liu *et al* [92] designed a 10 layers FCN for image synthesis, which learned a direct image-to-image/patch-to-patch mapping between different modalities and turned multi-modal image registration into mono-modal registration. With a similar idea, Liu *et al* [93] presented a novel modality synthesis approach IB-cGAN to synthesise Kilovoltage Digital Reconstructed Radiographs (KV-DRRs) images from Megavoltage Digital Radiographs (MV-DRs), and built a multi-modal image registration method combining the IB-cGAN with a traditional registration approach. Rather than converting images (generally the fixed image) from one modality to another, Blendowski *et al* [95] built a shared space for images from different modalities. In contrast, Tang *et al* [14] designed a multi-atlas registration framework, using a Cycle-GAN to synthesise multi-modal average atlases.

2.3.3. Reinforcement learning

Reinforcement learning networks are also explored in medical image registration, where the key idea is to provide a reward for every registration action. This class of approaches are mainly employed for rigid registration, mimicking a manual registration process. In 2017, Liao *et al* [96] firstly decomposed the 3D rigid registration task into a sequence of classification problems. They trained the intelligent agent in a

Table 4. A summary of reviewed other usages of DL networks for medical image registration, including 3 main classes and several interesting works which are not included in former classes.

Reference	Network	Modality	Dimension	Organ	Usage	Code
Wu *et al* [84]	2-layer ISA	MRI	3D patch	Brain	Feature extraction	—
Wu *et al* [85]	SAE	MRI	3D patch	Brain		—
Kearney *et al* [86]	DCIGN	CBCT, CT	3D patch	Head, neck		—
Zhu *et al* [87]	PCANet	CT, MRI	2D patch	Brain		—
Blendowsk *et al* [88]	CNN	CT	3D	Lung		—
Zheng *et al* [89]	PDA module	x-ray, DRR	2D,3D	Spine		—
Canalini *et al* [90]	3D U-Net	US	3D	Brain		—
Yang *et al* [91]	Encoder–decoder	MRI	2D	Brain	Image representation	—
Liu *et al* [92]	10-layer FCN	MRI	2D patch	Brain		—
Liu *et al* [93]	IB-cGAN	MV-DRs, KV-DRRs	2D	Head, neck, chest, pelvis		—
Lee *et al*,2019 [94]	ISTN	MRI	3D	Brain		link
Tang *et al* [14]	Cycle-GAN	MRI	3D	Brain		—
Blendowski *et al* [95]	Shape encoder–decoder	CT, MRI	3D	Heart		—
Liao *et al* [96]	3D classification CNN	CT, CBCT	3D	Spine, heart	Reinforcement learning	—
Toth *et al* [97]	CNN	CT, MRI, x-ray	2D,3D	Heart		—
Miao *et al* [98]	FCN+MDP	x-ray, CBCT	2D,3D	Spine		—

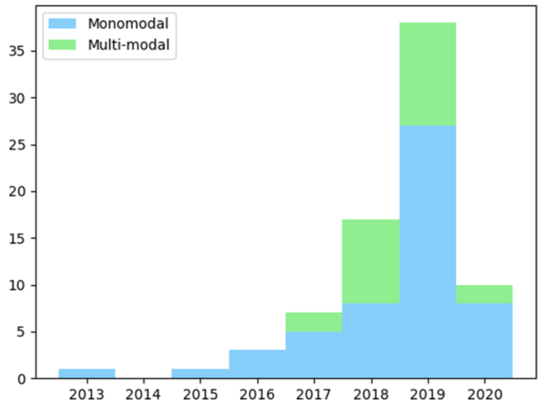

Figure 7. The number of papers for monomodal and multi-modal image registration methods in recent years.

greedy supervised fashion and proposed a hierarchical registration framework relying on the trained networks. Subsequent studies also explored a multi-agent system [98] and multi-modal registration [97]. Miao *et al* [98] formulated 2D–3D registration as a Markov Decision Process (MDP) with observations, actions, and rewards defined according to x-ray imaging systems, and proposed a multi-agent system to solve this challenging problem. Similarly, Toth *et al* [97] proposed a novel solution to register 3D preoperative models to 2D intraoperative images. They used a CNN to predict the optimal action with the highest reward, and demonstrated clinical feasibility through the robustness and efficiency of their framework.

In summary, DLIR methods have been demonstrated to outperform traditional registration methods in two main aspects, the registration speed and accuracy. After training, the registration of DLIR methods (supervised/unsupervised/weakly-supervised methods) is just one forward prediction, generally less than 1 s for an image pair. It is significantly faster than traditional methods, because several iterations are necessary for traditional registration methods. Additionally, most studies have demonstrated that DLIR methods are able to achieve higher registration accuracy than traditional methods, by utilising large training datasets. The introduction of deep neural networks has significantly improved image registration technologies, from their use for deriving novel representations of transformation models, to augmenting the execution of existing, traditional image registration methods. In the next section, we further introduce DLIR methods from the view of application.

3. Applications

In this section, we discuss DLIR methods from a different perspective, analysing them based on their applications. Medical image registration is essential for various clinical applications, such as, disease diagnosis and treatment planning, image-guided therapy and surgical interventions, treatment evaluation and patient prognostication, among others. The primary advantage of DLIR methods is their ability to compensate for soft tissue and patient motion in real-time, setting them apart from iterative traditional registration approaches. For instance, Krebs *et al* [101] designed an unsupervised generative deformation model within a temporal convolutional network to learn a probabilistic motion model from a sequence of images, which could be applied for both cardiac cine MRI spatio-temporal registration and motion analysis. Such an approach could be used for real-time cardiac motion analysis, providing the basis for discovery of novel motion-based disease biomarkers. DLIR methods can also be applied to estimate population-averaged atlases medical images. Dalca *et al* [102] described a probabilistic spatial deformation model based on diffeomorphisms, which enabled generation of atlases conditioned on several attributes of interest, such as, age and gender. Such approaches could be employed to generate virtual populations of anatomical structures of interest, useful for conducting in-silico clinical trials of medical devices. Furthermore, they provide a structured framework for assessing anatomical variability across populations, conditioned on relevant covariates. Image registration can also be used to directly facilitate image segmentation. By transforming images from a labelled atlas, Dalca *et al* [103] proposed a Bayesian segmentation method for 3D brain MRI, based on an unsupervised DLIR framework (removing the need laborious manual segmentation of numerous images). These studies highlight the versatility in application of DLIR methods, and present several promising directions for future research.

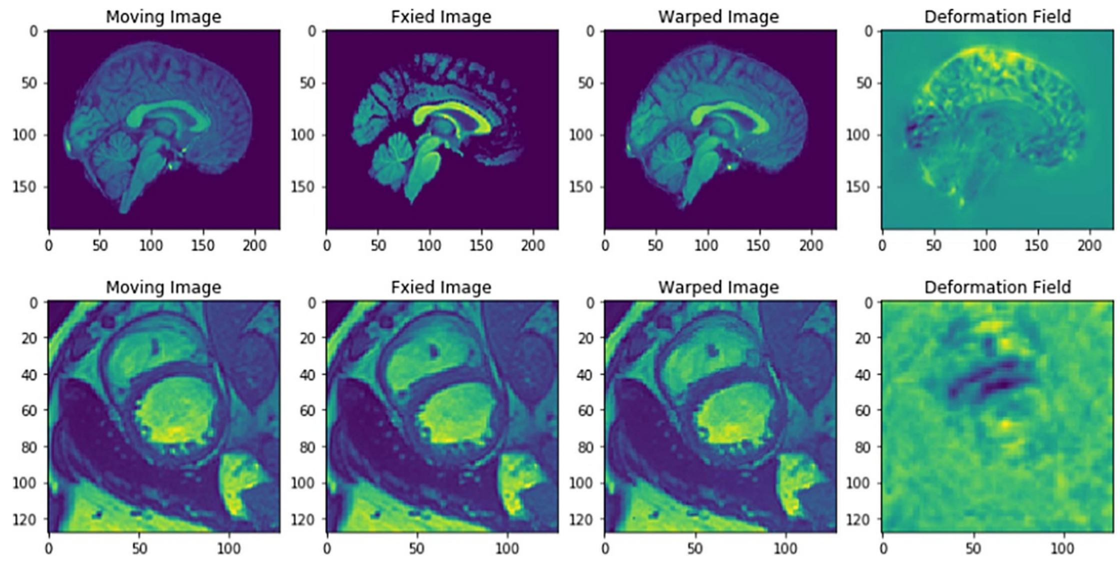

Figure 8. An example of brain and cardiac MRI image registration with Voxelmorph-diff [62, 63]. The first and second rows are brain MRI registration and cardiac MRI registration, respectively.

3.1. Monomodal registration

To facilitate and enhance future research on DLIR, we summarise all the publicly available datasets used for developing the registration method in table 5, with hyperlinks to each. Figure 7 summarises the number of articles published on monomodal and multi-modal registration methods in recent years. We observe that most studies thus far have focused on monomodal registration, with a substantial increase over the past year. The rate of development of DL-based multi-modal registration techniques is relatively slow in comparison, but the observed trend indicates that it is likely to increase substantially over the next couple of years. In this section, we review monomodal DLIR methods, focusing the most common image modalities used in the clinic, namely, MRI, CT, US and x-ray.

3.1.1. MRI registration

MRI is the most widely used modality for developing image registration techniques, with a special focus on brain MRIs, due to the availability of numerous large-scale public datasets (an example of brain and cardiac MRI registration is shown in figure 8). A large proportion of recent DLIR methods are thus validated on brain MRIs, in order to compare performance against previous state-of-the-art methods, such as Voxelmorph [49, 50], VTN [71], and Conv2warp [55]. Several brain MRI datasets are also utilised for developing multi-modal image registration methods [14, 38], with T1W and T2W modalities available in most brain MRI datasets. Apart from neuroimaging, cine MRI is the primary modality used for cardiac image registration and cardiac motion estimation [64, 101], with two available public datasets, Sunnybrook Cardiac Data (SCD) [104] and Automatic Cardiac Diagnosis Challenge (ACDC) [105].

3.1.2. CT registration

CT images are widely used to scan organs in the chest (lungs, heart) and abdomen (liver, kidneys, and pancreas). Specifically, as shown in table 5, there are four liver CT images datasets (MICCAI 2007 Grand Challenge [106], MSD, SLIVER [107], LiTS) and eight thoracic CT datasets (LIDC-IDRI [108], POPI [109], Empire 10 lung datasets, COPDGen [110], NLST [111], DIR-Lab-COPDgen [112], DIR-Lab-4DCT [113]). Besides, there are also several multi-modal datasets containing CT images, VISCERAL Anatomy3 [114], MM-WHS [115] and RIRE respectively. We found that CT image registration is the second largest domain used for developing medical image registration methods, with numerous recent studies on the topic [31, 32, 41, 55, 59, 69]. Compared with brain MRI registration, CT image registration is more challenging to some extent, due to limited soft-tissue contrast, and greater variability in image quality.

3.1.3. Ultrasound registration and x-ray registration

In contrast to the modalities discussed thus far, there are few publicly available datasets for US and x-ray images. Correspondingly, the number of papers focusing on the registration of US and x-ray images is also limited. There are two brain datasets, RESECT and BITE, containing US images, and just one paper focusing on monomodal US image registration [90] using publicly available datasets. As for x-ray images, there are six publicly available datasets, NLST [111], NIH ChestXray14 [116], OAI, JSRT [117], Montgomery County x-ray database [118] and Shenzhen Hospital x-ray database [118]. However, there are relatively few studies on x-ray image registration [25, 80], compared with MRI and CT.

3.2. Multi-modal registration

With the ability to calculate the dissimilarity between images from different modalities, DL has been widely applied in multi-modal registration. However, in contrast to monomodal registration, there is limited availability of public datasets for multi-modal registration. Based on the reviewed literature, we found just three publicly available multi-modal datasets for developing registration approaches, namely, RIRE, VISCERAL Anatomy3 benchmark [114] and Multi-modality Whole Heart Segmentation dataset (MM-WHS) [115] respectively. Although there are numerous studies focusing on multi-modal registration, most of them collect and use independent, private datasets to develop and validate their algorithms. In this section, we discuss several typical multi-modal registration applications: T1W-T2W registration, CT-MRI registration, CT-CBCT registration, 2D–3D registration.

3.2.1. T1W-T2W registration

T1W-T2W registration aims to learn a mapping between T1-weighted MRI images and T2-weighted MRI images. It is a common multi-modal registration task in neuroimaging, with many publicly available brain MRI datasets. Yang *et al* [37] proposed a 3D Bayesian encoder–decoder network for T1W-T2W multi-modal registration based on IBIS 3D Autism Brain image dataset. Qin *et al* [24] proposed a GAN-based network UMDIR for this task based on BraTS2017 dataset. Liu *et al* [92] tested their methods on several multi-modal registration tasks, T2W vs proton density (PD), T1W vs PD, and T1W vs T2W respectively. Tang *et al* [14] utilised a Cycle-GAN to synthesise multi-modal atlases (T1W, T1 contrast-enhanced, T2W, FLAIR), building a bridge between different modalities.

3.2.2. CT-MRI registration

CT-MRI matching is another common multi-modal registration application. The three public multi-modal registration datasets we mentioned previously, all contain both CT and MRI images for the same subjects, useful for developing multi-modal registration approaches. Zhu *et al* [87] proposed a PCANet to learn structural representations for FFD on the RIRE dataset. Using a private dataset, Cao *et al* [72] proposed a 'CNN+STN' network for registering CT and MRI images. Besides these, GAN-based networks have also been employed for pelvic [23], while other studies have proposed approaches to register cardiac CT and MRI images based on the MM-WHS dataset [83].

3.2.3. MRI-TRUS registration

Several papers have also explored registering MRI and TRUS images. From our reviewed research, there are two datasets RESECT [135] and BITE [134] publicly available for this registration task, and several methods proposed were developed based on them [136, 137]. However, most of these studies are based on private datasets. Guo *et al* [43] proposed a supervised network to tackle rigid MRI-TRUS registration on prostate images. Hu *et al* [81, 82] proposed a global sub-network, for affine registration, with a local sub-network for deformable registration of T2W MRI and TRUS images. Yan *et al* [39] designed a GAN-based adversarial image registration network (AIR-Net) to address this task. Haskins *et al* [17] utilised CNN to calculate the similarity between MRI and TRUS images.

3.2.4. CT-CBCT registration

Recently, image registration between CBCT and CT images has also drawn some attention [41, 73, 86]. Focusing on CT-CBCT deformable registration on head and neck images, Kearney *et al* [86] proposed DCIGN to learn hierarchical features, which outperformed intensity corrected Demons and landmark-guided DIR. To achieve CT-CBCT rigid registration in Image-guided Radiotherapy (IGRT), Yao [41] proposed a CNN to predict an initial rough transformation, then utilised a traditional intensity-based registration to refine the registration. This shortened the prediction time while ensuring high registration accuracy.

3.2.5. 2D–3D registration

In most multi-modal registration applications discussed thus far, the dimension of the fixed and moving images are identical. Publicly available datasets provide 3D image volumes, which can also be employed for slice-wise 2D–2D registration. Therefore, studies thus far have primarily focused on 2D–2D and 3D–3D image registration. In addition to these, 2D–3D image registration is also useful for a variety of clinical applications and forms a major part of ongoing research in DL-based multi-modal image registration. This

Table 5. Overview of datasets used for medical image registration. We list some basic information (organ, registration type, name and image modality) of every dataset and the correspondent link and references which exemplify their methods on it. Note that some brain MRI datasets containing various modalities (e.g. T1W, T2W) could also be applied for multi-modal registration.

Organ	Registration	Datasets	Modality	Reference
Brain	Monomodal	ADNI [119]	MRI	[25, 27, 28, 48, 49, 60, 62, 63, 70, 71, 84, 85, 102]
		IXI	MRI	[28, 29, 40, 84, 92]
		OASIS [120]	MRI	[26, 38, 49–51, 57, 62, 63, 91, 102, 103]
		BRATS2015 [121]	MRI	[91, 14]
		LPBA40 [122]	MRI	[14, 22, 23, 27–29, 38, 48, 51, 55, 56, 60, 65, 68, 70, 71, 85]
		IBIS [123]	MRI	[38, 37].
		IBSR18 [122]	MRI	[22, 23, 29, 38, 60, 68]
		MGH10 [122]	MRI	[22, 23, 29, 38, 55, 60, 68]
		CUMC12 [122]	MRI	[22, 23, 29, 38, 55, 60, 68]
		ABIDE [124]	MRI	[49, 50, 62, 63, 70, 71, 102, 103]
		ADHD200 [125]	MRI	[49, 50, 62, 63, 70, 71, 102, 103]
		MCIC [126]	MRI	[49, 50, 62, 63, 102, 103]
		PPMI [127]	MRI	[49, 50, 62, 63, 102, 103]
		HABS [128]	MRI	[49, 50, 62, 63, 102, 103]
		Harvard GSP [129]	MRI	[49, 50, 62, 63, 102, 103]
		FreeSurfer Buckner40 [130]	MRI	[50]
		Mindboggle101 [131]	MRI	[51, 54, 56, 65, 77]
		BraTS2017 [132]	MRI	[24]
		BrainWeb [133]	Simulated MRI	[36, 87, 92]
	Multi-modal	RIRE	CT, MRI	[87]
		BITE [134]	US, MRI	[90]
		RESECT [135]	US, MRI	[90, 136, 137]
Heart	Monomodal	Sunnybrook [104]	Cine MRI	[46, 58, 59, 69]
		ACDC [105]	Cine MRI	[64, 75, 79, 101]
	Multi-modal	MM-WHS [115]	CT, MRI	[83, 95]
Knee	Multi-modal	OAI	MRI, x-ray	[66, 67, 77]

(Continued)

Table 5. (Continued).

Organ	Registration	Datasets	Modality	Reference
Liver	Monomodal	MICCAI 2007 Grand Challenge [106]	CT	[71]
		MSD	CT	[70]
		SLIVER [107]	CT	[70]
		LiTS	CT	[70, 71]
Chest	Monomodal	COPDGen [110]	CT	[24]
		NLST [111]	CT, x-ray	[69]
		DIR-Lab-COPDgen [112]	CT	[88]
		DIR-Lab-4DCT [113]	CT	[31, 32, 52, 55, 59, 69, 73]
		SPARE [138]	CT, CBCT	[73]
		POPI [109]	CT	[31, 32, 55, 59]
		LIDC-IDRI [108]	CT	[97, 32]
		Empire 10 lung datasets	CT	[36]
		NIH ChestXray14 dataset [116]	x-ray	[25]
		JSRT [117]	x-ray	[80]
		Montgomery County x-ray database [118]	x-ray	[80]
		Shenzhen Hospital x-ray database [118]	x-ray	[80]
Several organs	Multi-modal	UK Biobank Imaging Study	MRI	[94]
Whole body	Multi-modal	VISCERAL Anatomy3 [114]	CT, MRI	[76]

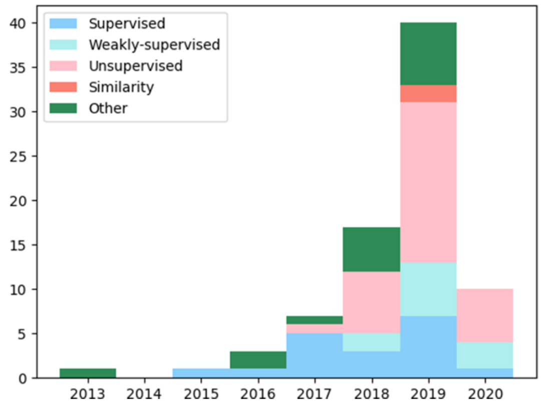

Figure 9. Histogram depicts the number of DLIR papers published to date, grouped according to the categories defined in section 2). 'Similarity' refers to the category of deep similarity.

task is even more challenging, due to the difference in dimensionality and the issue of overlapping tissues and contrast common to 2D images such as x-rays. Studies on 2D–3D registration have mainly focused on registering x-ray images to other 3D modality images, such as MR/US [97], CT [89], and CBCT [42, 98]. Additionally, slice-to-volume registration has also received some attention in recent years [34].

4. Discussion

Previous sections have introduced and discussed the most relevant DLIR published to date. In this section, we present current trends in the development of DLIR methods, and discuss the main challenges that are yet to be addressed. Finally, a summary of the possible directions for future research in the field, are outlined.

4.1. Development trends

As discussed previously, recent years have witnessed a dramatic increase in the number of papers published on DLIR methods. Unsurprisingly, this follows wider trends in the use of DL for various tasks in medical image analysis and computer vision. The development of DL experienced a boom after 2015, with the release of several open-source deep learning software libraries (e.g. Tensorflow, Keras and Pytorch). This offered a convenient and easy-to-use environment for quick prototyping of DL networks. We found that the development of DLIR began in 2015. The first two methods proposed in 2013 and 2015 just applied CNNs for feature extraction. DLIR methods with high impact in this domain were first proposed in 2016, where, DL networks were used to predict deformation fields. Subsequent years have seen continuous increase in the number of DLIR papers, with several significant and innovative contributions making a strong case for their superiority over traditional, iterative registration approaches.

Although it has been just a few years since DL networks were applied to medical image registration, the use of DL for medical image registration has seen several changes. The evolution in the development of DLIR methods is described by the histogram plot shown in figure 9. We characterise this evolution over four stages. The first stage attempted to use deep neural networks for feature extraction, which in turn were used to guide traditional registration algorithms, by providing more discriminative information than the original images. Next, studies focused on addressing a crucial limitation of iterative traditional registration approaches, viz. long execution times. By learning the space of desired spatial transformations, given suitable training data, the aim of several supervised networks proposed in this stage was primarily to speed up registration during inference. Models trained in this fashion on suitable image pairs are many-folds faster than iterative registration approaches during testing/inference. However, supervised methods require ground-truth spatial transformations to be available for training samples, which are difficult to obtain in most real world applications, thereby limiting their applicability.

To circumvent the need for ground-truth deformation fields, at the third stage, unsupervised and weakly-supervised methods were proposed. These approaches demonstrated comparable registration accuracy and speed with supervised methods, while requiring just weak labels or no labels at all. Specifically, weakly-supervised registration methods were proposed a little later than unsupervised methods. In this stage, there was no noticeable improvement in accuracy. In contrast, the deformation fields generated by DL networks were sometimes non-smooth and unrealistic. The final stage aimed to improve the accuracy of registration and make deformation fields smoother. Several additional types of information (e.g. segmentation masks) were incorporated into networks using weakly-supervised learning frameworks, and various forms of regularisation were introduced during training. These four stages are not strictly separated. However, we could see a clear line in the development of DLIR methods, as evidenced by the plot shown in figure 9.

We note that the dimensionality of images used to train DLIR networks is gradually tending towards the natural space of deformations of organs of interest, as powerful computing hardware becomes available to handle the high computational and memory requirements. Initially, the input data used to train DL registration networks were mostly 2D images [26, 35, 46, 75, 87, 91, 93, 101] or 2D image patches [21, 47, 92]. They gradually became 3D image patches [26, 27, 29, 37, 38, 69, 72, 84–86], and finally, whole 3D image volumes and 4D images/patches [55, 59]. In fact, it is natural to perform 3D registration for most medical images, as most organ motions take place in 3D. For most medical image registration applications, 3D is enough for registration tasks. However, for some special applications such as cardiac motion estimation, researchers are exploring 3D+t or 4D image registration techniques, which is less common in other computer vision applications.

4.2. Main challenges

Though DLIR methods have addressed many challenging problems in medical image registration and achieved faster and more accurate registration than traditional methods, there are several challenges that are yet to be tackled in this domain.

4.2.1. Preprocessing

Preprocessing is an integral part of image registration, which generally consists of several operations geared towards simplifying the data to be registered. Different preprocessing steps may lead to different registration results, even using the same datasets. In other computer vision tasks such as image classification and image segmentation, researchers demonstrate their methods on public datasets, where the prepossessing is easy to realise and shared by all researchers. However, in medical image registration, although there are many publicly available datasets, the preprocessing steps tend to vary across studies. For example, in brain MRI image registration, there are many publicly available datasets, such as OASIS [120], ADNI [119], IXI and MGH10 [122]. Furthermore, there are several well-acknowledged preprocessing steps, such as skull-stripping, affine registration, spatial resampling, image enhancement, intensity normalisation and cropping. However, studies often use different datasets for training and testing, employ different preprocessing procedures with adapted parameters for each step (e.g. voxel size, smoothing factor, etc). Therefore, in some earlier DLIR studies, specifically, prior to Voxelmorph, methods were usually only compared with traditional state-of-the-art registration approaches (e.g. ANTs [44], Elastix [139], Demons [140, 141]).

4.2.2. Clinical applications

Clinical applications are the final destination for all the medical image processing and analysis methods. Until now, numerous DL-based image registration methods have proved their efficiency and superiority compared to classical methods. However, we are yet to see a DL-based tool deployed in a clinical setting, like ANTs and Elastix in classical methods. With no well-adapted tools, it is difficult for clinicians and clinical researchers to use DL networks in clinical applications. Besides, as DL networks are challenging to interpret, even though a trained model shows high accuracy in the test datasets, clinicians are still wary of employing them regularly to analyse patient data. There have been a few studies that have attempted to quantify the uncertainty of the predicted registration results, with a view to providing clinicians with useful information regarding the validity of the registration [26, 37, 62, 63, 91]. However, further research and systematic means for assessing registration uncertainty are necessary to build trust in the community and accelerate the adoption of DLIR methods in clinical settings.

4.2.3. Limited data

Lack of suitable public datasets is another fundamental problem limiting the development of DLIR methods. To obtain accurate and robust models, DL networks need to be trained on large-scale datasets. Although unsupervised learning registration methods do not require any ground-truth data, currently, the primary publicly available datasets are focused solely on brain MRI images, with just a few datasets containing other organs/modality images. Besides, for supervised methods and weakly-supervised methods, sourcing high-quality ground-truth data remains a challenge. We also observe that several studies only exemplify their method on their private datasets due to a lack of publicly available datasets, which is not convenient for benchmarking and comparing state-of-the-art methods. With the increase of datasets, a fairer comparison will be possible, facilitating greater innovation in DLIR.

4.3. Possible directions

In this section, we outline possible directions for future research in DLIR, to address the challenges discussed thus far. The first step towards identifying these is to consider the aims of DLIR. Accuracy, robustness and speed are common goals for all registration methods. DLIR methods trained to predict the spatial transformation matching a pair or group of images shown no significant difference in registration speed. Therefore, the obvious focus of future approaches on DLIR should be on improving the accuracy and generalisation capability of the networks, and ensuring that the estimated deformation fields more realistic and smooth.

4.3.1. Combining the superiority of traditional methods with DL

A possible direction is to combine the advantages of traditional methods with deep learning networks. Though DLIR methods have significantly improved registration speed and accuracy compared with classical methods, the superiority of classical methods (e.g. diffeomorphic attributes and robust registration) can not be overlooked. The trend to make deformation fields smoother is just combining the diffeomorphic transformation in traditional methods with DL networks.

4.3.2. Boosting performance with priors

As discussed previously, medical image registration differs greatly from other medical image analysis tasks. Future research should introduce more registration priors to DL networks, making DL networks more specific to image registration, and more application specific. To improve registration performance DLIR networks could be imbued with prior information related to the expected type of deformation, spatial relationship between anatomical structures, and the topology and morphology of anatomical structures. For example, although ground-truth spatial transformations are seldom available, other labels could serve as the ground-truth to guide the training process. Several methods on weakly-supervised image registration have been proposed, which generally achieve better performance than its corresponding unsupervised variant (at no additional cost in terms of execution speed). More informative priors combined with training data that is synthetically modified, such as, blackening pixels in the moving image, or generating adversarial examples [142], could enhance the ability of networks to generalise to unseen data, while remaining robust to variable image quality. Consequently, combining different types of spatial and temporal priors with DL networks is a promising direction for future research in the field.

5. Conclusion

In this review, we comprehensively summarised the evolution of deep learning-based medical image registration. We discussed the existing challenges and potential directions for future research. We present a thorough summary for publicly available datasets and links to code of published papers, to facilitate benchmarking of algorithms and enhance future research. The development of deep learning-based image registration methods have experienced a similar trend to the development of DL. Image registration networks increasingly operate in the natural space of the organs or deformations of interest, i.e. gradually evolving from processing 2D images to 3D/4D (dynamic) volumes. Recent contributions range from speeding up registration in higher dimensions to reducing the need for ground-truth during training, or advanced regularisation constraints to retrieve plausible deformation fields and preserve anatomical topology. Due to the difficulty in obtaining ground-truth data for training, DLIR networks gradually turned to unsupervised learning from supervised learning.

Currently, the lack of suitable, publicly available data is a fundamental obstacle to the development of innovative DLIR techniques. Additionally, the varied preprocessing steps employed across different studies makes it difficult to compare state-of-the-art approaches and undertake comprehensive benchmarking studies. Although DLIR networks have achieved significant improvements in speed and accuracy of registration for most tasks, some tasks remain, in which their accuracy is at best comparable to traditional approaches. Additionally, studies demonstrating the clinical viability of DLIR methods, as done previously for several traditional registration tools (e.g. ANTs, Demons), are still lacking. We still have not seen such a trend in DLIR methods, but expect this to be the next frontier of research in the field.

Accuracy, generalisation, realistic and smooth deformation will likely remain the main research focus for medical image registration in the near future. Alongside an increased availability of multi-modal datasets, we expect an increased focus on multi-modal registration using DL approaches.

Acknowledgments

The Royal Academy of Engineering supports the work of A F F through a Chair in Emerging Technologies (CiET1819\19) and the MedIAN Network (EP/N026993/1) funded by the Engineering and Physical Sciences Research Council (EPSRC).

ORCID iDs

Xiang Chen https://orcid.org/0000-0003-4203-4578
Andres Diaz-Pinto https://orcid.org/0000-0002-4865-8296
Nishant Ravikumar https://orcid.org/0000-0003-0134-107X
Alejandro F Frangi https://orcid.org/0000-0002-2675-528X

References

[1] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 *Proc. IEEE* **86** 2278–324
[2] Hinton G E, Osindero S and Teh Y W 2006 *Neural Comput.* **18** 1527–54
[3] Jaderberg M *et al* 2015 Spatial transformer networks *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 2017–25
[4] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 2672–80
[5] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 1097–105
[6] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition *Int. Conf. Learning Representations*
[7] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. Conf. Computer Vision and Pattern Recognition* pp 770–8
[8] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *Proc. Conf. Computer Vision and Pattern Recognition* pp 4700–8
[9] Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 234–41
[10] Isola P, Zhu J Y, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks *Proc. Conf. Computer Vision and Pattern Recognition* pp 1125–34
[11] Zhu J Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *Proc. IEEE Int. Conf. Computer Vision* pp 2223–32
[12] Chen X, Qing L, He X, Su J and Peng Y 2018 *IEEE Access* **6** 14567–75
[13] Frid-Adar M, Klang E, Amitai M, Goldberger J and Greenspan H 2018 Synthetic data augmentation using GAN for improved liver lesion classification *IEEE Int. Symp. Biomedical Imaging* (IEEE) pp 289–93
[14] Tang Z, Yap P T and Shen D 2018 *IEEE Trans. Image Process.* **28** 2293–304
[15] Han Z, Wei B, Mercado A, Leung S and Li S 2018 *Med. Image Anal.* **50** 23–35
[16] Andrade N, Faria F A and Cappabianco F A M 2018 A practical review on medical image registration: from rigid to deep learning based approaches *Conf. Graphics, Patterns and Images* (IEEE) pp 463–70
[17] Haskins G, Kruger U and Yan P 2020 *Mach. Vis. Appl.* **31** 8
[18] Fu Y, Lei Y, Wang T, Curran W J, Liu T and Yang X 2020 *Phys. Med. Biol.* **65** 20
[19] Tustison N J, Avants B B and Gee J C 2019 *Magn. Reson. Imaging* **64** 142–53
[20] Haskins G, Kruecker J, Kruger U, Xu S, Pinto P A, Wood B J and Yan P 2019 *Int. J. Comput. Assist. Radiol. Surg.* **14** 417–25
[21] Zhu N, Najafi M, Han B, Hancock S and Hristov D 2019 *Technol. Cancer Res. Treat.* **18** 1–11
[22] Fan J, Cao X, Xue Z, Yap P T and Shen D 2018 Adversarial similarity network for evaluating image alignment in deep learning based registration *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 739–46
[23] Fan J, Cao X, Wang Q, Yap P T and Shen D 2019 *Med. Image Anal.* **58** 101545
[24] Qin C, Shi B, Liao R, Mansi T, Rueckert D and Kamen A 2019 Unsupervised deformable registration for multi-modal images via disentangled representations *Int. Conf. Information Processing in Medical Imaging* (Berlin: Springer) pp 249–61
[25] Mahapatra D and Ge Z 2019 Training data independent image registration with GANs using transfer learning and segmentation information *IEEE Int. Symp. Biomedical Imaging* (IEEE) pp 709–13
[26] Yang X, Kwitt R and Niethammer M 2016 Fast predictive image registration *Deep Learning and Data Labeling for Medical Applications* (Berlin: Springer) pp 48–57
[27] Cao X, Yang J, Zhang J, Nie D, Kim M, Wang Q and Shen D 2017 Deformable image registration based on similarity-steered CNN regression *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 300–8
[28] Cao X, Yang J, Zhang J, Wang Q, Yap P T and Shen D 2018 *IEEE Trans. Biomed. Eng.* **65** 1900–11
[29] Fan J, Cao X, Yap P T and Shen D 2019 *Med. Image Anal.* **54** 193–206
[30] Rohé M M, Datar M, Heimann T, Sermesant M and Pennec X 2017 SVF-Net: learning deformable image registration using shape matching *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 266–74
[31] Eppenhof K A and Pluim J P 2018 *IEEE Trans. Med. Imaging* **38** 1097–105
[32] Eppenhof K, Lafarge M, Veta M and Pluim J 2019 *IEEE Trans. Med. Imaging* **39** 1594–604
[33] Sokooti H, de Vos B, Berendsen F, Lelieveldt B P, Išgum I and Staring M 2017 Nonrigid image registration using multi-scale 3D convolutional neural networks *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 232–9
[34] Salehi S S M, Khan S, Erdogmus D and Gholipour A 2018 *IEEE Trans. Med. Imaging* **38** 470–81
[35] Xia Y, Li Y, Xun L, Yan Q and Zhang D 2019 *Gait Posture* **68** 403–8

[36] Zhao L and Jia K 2015 *Comput. Math. Methods Med.* **2015** 836202:1–836202:16
[37] Yang X, Kwitt R, Styner M and Niethammer M 2017 Fast predictive multimodal image registration *IEEE Int. Symp. Biomedical Imaging* (IEEE) pp 858–62
[38] Yang X, Kwitt R, Styner M and Niethammer M 2017 *Neuroimage* **158** 378–96
[39] Yan P, Xu S, Rastinehad A R and Wood B J 2018 Adversarial image registration with application for MR and TRUS image fusion *Int. Workshop Machine Learning in Medical Imaging* (Berlin: Springer) pp 197–204
[40] Sedghi A, Kapur T, Luo J, Mousavi P and Wells W M 2019 Probabilistic image registration via deep multi-class classification: characterizing uncertainty *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures* (Berlin: Springer) pp 12–22
[41] Yao Z, Feng H, Song Y, Li S, Yang Y, Liu L and Liu C 2019 *J. Med. Syst.* **43** 194:1–194:8
[42] Liao H, Lin W A, Zhang J, Zhang J, Luo J and Zhou S K 2019 Multiview 2D/3D rigid registration via a point-of-interest network for tracking and triangulation *Proc. Conf. Computer Vision and Pattern Recognition* pp 12638–47
[43] Guo H, Kruger M, Xu S, Wood B J and Yan P 2020 *Comput. Med. Imaging Graph.* **84** 101769
[44] Avants B B, Tustison N J, Song G, Cook P A, Klein A and Gee J C 2011 *Neuroimage* **54** 2033–44
[45] Lorenzi M *et al* 2013 *Neuroimage* **81** 470–83
[46] de Vos B D, Berendsen F F, Viergever M A, Staring M and Isgum I 2017 End-to-end unsupervised deformable image registration with a convolutional neural network *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support—3rd Int. Workshop, Dlmia 2017 and 7th Int. Workshop, ML-CDS 2017, Held in Conjunction With Miccai* (*Lecture Notes in Computer Science* vol 10553) (Berlin: Springer) pp 204–12
[47] Lv J, Yang M, Zhang J and Wang X 2018 *Br. J. Radiol.* **91** 20170788
[48] Li H and Fan Y 2018 Non-rigid image registration using self-supervised fully convolutional networks without training data *IEEE Int. Symp. Biomedical Imaging* (IEEE) pp 1075–8
[49] Balakrishnan G, Zhao A, Sabuncu M R, Guttag J and Dalca A V 2018 An unsupervised learning model for deformable medical image registration *Proc. Conf. Computer Vision and Pattern Recognition* pp 9252–60
[50] Balakrishnan G, Zhao A, Sabuncu M R, Guttag J and Dalca A V 2019 *IEEE Trans. Med. Imaging* **38** 1788–800
[51] Zhu Z, Cao Y, Qin C, Rao Y, Ni D and Wang Y 2020 Unsupervised 3D end-to-end deformable network for brain MRI registration *2020 42nd Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 1355–9
[52] Fu Y, Lei Y, Wang T, Higgins K, Bradley J D, Curran W J, Liu T and Yang X 2020 *Med. Phys.* **47** 1763–74
[53] Stergios C, Mihir S, Maria V, Guillaume C, Marie-Pierre R, Stavroula M and Nikos P 2018 Linear and deformable image registration with 3D convolutional neural networks *Image Analysis for Moving Organ, Breast and Thoracic Images* (Berlin: Springer) pp 13–22
[54] Kuang D and Schmah T 2019 Faim—a convnet method for unsupervised 3D medical image registration *Int. Workshop Machine Learning in Medical Imaging* (Berlin: Springer) pp 646–54
[55] Ali S and Rittscher J 2019 Conv2Warp: an unsupervised deformable image registration with continuous convolution and warping *Int. Workshop Machine Learning in Medical Imaging* (Berlin: Springer) pp 489–97
[56] Hu X, Kang M, Huang W, Scott M R, Wiest R and Reyes M 2019 Dual-stream pyramid registration network *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 382–90
[57] Bhalodia R, Elhabian S Y, Kavan L and Whitaker R T 2019 A cooperative autoencoder for population-based regularization of CNN image registration *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 391–400
[58] Sang Y and Ruan D 2020 Enhanced image registration with a network paradigm and incorporation of a deformation representation model *2020 IEEE 17th Int. Symp. Biomedical Imaging (ISBI)* (IEEE) pp 91–4
[59] Fechter T and Baltas D 2020 *IEEE Trans. Med. Imaging* **39** 2506–17
[60] Gu D, Cao X, Ma S, Chen L, Liu G, Shen D and Xue Z 2020 Pair-wise and group-wise deformation consistency in deep registration network *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 171–80
[61] Kim B, Kim J, Lee J G, Kim D H, Park S H and Ye J C 2019 Unsupervised deformable image registration using cycle-consistent CNN *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 166–74
[62] Dalca A V, Balakrishnan G, Guttag J and Sabuncu M R 2018 Unsupervised learning for fast probabilistic diffeomorphic registration *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 729–38
[63] Dalca A V, Balakrishnan G, Guttag J and Sabuncu M R 2019 *Med. Image Anal.* **57** 226–36
[64] Krebs J, Delingette H E, Mailhé B, Ayache N and Mansi T 2019 *IEEE Trans. Med. Imaging* **38** 2165–76
[65] Liu L, Hu X, Zhu L and Heng P A 2019 Probabilistic multilayer regularization network for unsupervised 3D brain image registration *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 346–54
[66] Shen Z, Han X, Xu Z and Niethammer M 2019 Networks for joint affine and non-parametric image registration *Proc. Conf. Computer Vision and Pattern Recognition* pp 4224–33
[67] Shen Z, Vialard F X and Niethammer M 2019 Region-specific diffeomorphic metric mapping *Advances in Neural Information Processing Systems* pp 1098–108
[68] Niethammer M, Kwitt R and Vialard F X 2019 Metric learning for image registration *Proc. Conf. Computer Vision and Pattern Recognition* pp 8463–72
[69] de Vos B D, Berendsen F F, Viergever M A, Sokooti H, Staring M and Išgum I 2019 *Med. Image Anal.* **52** 128–43
[70] Zhao S *et al* 2019 Recursive cascaded networks for unsupervised medical image registration *Proc. IEEE Int. Conf. Computer Vision* pp 10600–10
[71] Zhao S, Lau T, Luo J, Eric I, Chang C and Xu Y 2019 *IEEE J. Biomed. Health Inform.* **24** 1394–404
[72] Cao X, Yang J, Wang L, Xue Z, Wang Q and Shen D 2018 Deep learning based inter-modality image registration supervised by intra-modality similarity *Int. Workshop Machine Learning in Medical Imaging* (Berlin: Springer) pp 55–63
[73] Jiang Z, Yin F F, Ge Y and Ren L 2020 *Phys. Med. Biol.* **65** 015011
[74] Avants B B, Epstein C L, Grossman M and Gee J C 2008 *Med. Image Anal.* **12** 26–41
[75] Hering A, Kuckertz S, Heldmann S and Heinrich M P 2019 Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking *Bildverarbeitung für die Medizin 2019* (Berlin: Springer) pp 309–14
[76] Heinrich M P 2019 Closing the gap between deep and conventional image registration using probabilistic dense displacement networks *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 50–8
[77] Xu Z and Niethammer M 2019 Deepatlas: joint semi-supervised learning of image registration and segmentation *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 420–9

[78] Chen L, Cao X, Chen L, Gao Y, Shen D, Wang Q and Xue Z 2020 Semantic hierarchy guided registration networks for intra-subject pulmonary CT image alignment *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 181–9
[79] Ha I Y, Wilms M and Heinrich M 2020 *Sensors* **20** 1392
[80] Mansilla L, Milone D H and Ferrante E 2020 *Neural Netw.* **124** 269–79
[81] Hu Y *et al* 2018 Label-driven weakly-supervised learning for multimodal deformable image registration *IEEE Int. Symp. Biomedical Imaging* (IEEE) pp 1070–4
[82] Hu Y *et al* 2018 *Med. Image Anal.* **49** 1–13
[83] Hering A, Kuckertz S, Heldmann S and Heinrich M P 2019 *Int. J. Comput. Assist. Radiol. Surg.* **14** 1901–12
[84] Wu G, Kim M, Wang Q, Gao Y, Liao S and Shen D 2013 Unsupervised deep feature learning for deformable registration of MR brain images *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 649–56
[85] Wu G, Kim M, Wang Q, Munsell B C and Shen D 2016 *IEEE Trans. Biomed. Eng.* **63** 1505–16
[86] Kearney V, Haaf S, Sudhyadhom A, Valdes G and Solberg T D 2018 *Phys. Med. Biol.* **63** 185017
[87] Zhu X, Ding M, Huang T, Jin X and Zhang X 2018 *Sensors* **18** 1477
[88] Blendowski M and Heinrich M P 2019 *Int. J. Comput. Assist. Radiol. Surg.* **14** 43–52
[89] Zheng J, Miao S, Wang Z J and Liao R 2018 *J. Med. Imaging* **5** 021204
[90] Canalini L, Klein J, Miller D and Kikinis R 2019 *Int. J. Comput. Assist. Radiol. Surg.* **14** 1697–1713
[91] Yang X, Han X, Park E, Aylward S, Kwitt R and Niethammer M 2016 Registration of pathological images *Int. Workshop Simulation and Synthesis in Medical Imaging* (Berlin: Springer) pp 97–107
[92] Liu X, Jiang D, Wang M and Song Z 2019 *Med. Biol. Eng. Comput.* **57** 1037–48
[93] Liu C, Lu Z, Ma L, Wang L, Jin X and Si W 2019 *Med. Phys.* **46** 4575–87
[94] Lee M C, Oktay O, Schuh A, Schaap M and Glocker B 2019 Image-and-spatial transformer networks for structure-guided image registration *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 337–45
[95] Blendowski M, Bouteldja N and Heinrich M P 2020 *Int. J. Comput. Assist. Radiol. Surg.* **15** 269–76
[96] Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T and Comaniciu D 2017 An artificial agent for robust image registration *AAAI Conf. Artificial Intelligence*
[97] Toth D, Miao S, Kurzendorfer T, Rinaldi C A, Liao R, Mansi T, Rhode K and Mountney P 2018 *Int. J. Comput. Assist. Radiol. Surg.* **13** 1141–9
[98] Miao S, Piat S, Fischer P, Tuysuzoglu A, Mewes P, Mansi T and Liao R 2018 Dilated FCN for multi-agent 2D/3D medical image registration *AAAI Conf. Artificial Intelligence* pp 4694–701
[99] Shen D 2007 *Pattern Recognit.* **40** 1161–72
[100] Chan T H, Jia K, Gao S, Lu J, Zeng Z and Ma Y 2015 *IEEE Trans. Image Process.* **24** 5017–32
[101] Krebs J, Mansi T, Ayache N and Delingette H 2019 Probabilistic motion modeling from medical image sequences: application to cardiac cine-MRI *Int. Workshop Statistical Atlases and Computational Models of the Heart* (Berlin: Springer) pp 176–85
[102] Dalca A, Rakic M, Guttag J and Sabuncu M 2019 Learning conditional deformable templates with convolutional networks *Advances in Neural Information Processing Systems* pp 804–16
[103] Dalca A V, Yu E, Golland P, Fischl B, Sabuncu M R and Iglesias J E 2019 Unsupervised deep learning for bayesian brain MRI segmentation *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 356–65
[104] Radau P, Lu Y, Connelly K, Paul G, Dick A and Wright G 2009 Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI 49 (http://hdl.handle.net/10380/3070)
[105] Bernard O *et al* 2018 *IEEE Trans. Med. Imaging* **37** 2514–25
[106] Murphy K *et al* 2011 *IEEE Trans. Med. Imaging* **30** 1901–20
[107] Heimann T *et al* 2009 *IEEE Trans. Med. Imaging* **28** 1251–65
[108] Armato S G III *et al* 2011 *Med. Phys.* **38** 915–31
[109] Vandemeulebroucke J, Rit S, Kybic J, Clarysse P and Sarrut D 2011 *Med. Phys.* **38** 166–78
[110] Regan E A, Hokanson J E, Murphy J R, Make B, Lynch D A, Beaty T H, Curran-Everett D, Silverman E K and Crapo J D 2011 *COPD* **7** 32–43
[111] National Lung Screening Trial Research Team 2011 *New Engl. J. Med.* **365** 395–409
[112] Castillo R, Castillo E, Fuentes D, Ahmad M, Wood A M, Ludwig M S and Guerrero T 2013 *Phys. Med. Biol.* **58** 2861
[113] Castillo R, Castillo E, Guerra R, Johnson V E, McPhail T, Garg A K and Guerrero T 2009 *Phys. Med. Biol.* **54** 1849
[114] Jimenez-del Toro O *et al* 2016 *IEEE Trans. Med. Imaging* **35** 2459–75
[115] Zhuang X and Shen J 2016 *Med. Image Anal.* **31** 77–87
[116] Wang X, Peng Y, Lu L, Lu Z, Bagheri M and Summers R M 2017 Chest-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases *Proc. Conf. Computer Vision and Pattern Recognition* pp 2097–106
[117] Shiraishi J *et al* 2000 *Am. J. Roentgenol.* **174** 71–4
[118] Jaeger S, Candemir S, Antani S, Wáng Y X J, Lu P X and Thoma G 2014 *Quant. Imaging Med. Surg.* **4** 475–7
[119] Mueller S G, Weiner M W, Thal L J, Petersen R C, Jack C R, Jagust W, Trojanowski J Q, Toga A W and Beckett L 2005 *Alzheimer's dementia* **1** 55–66
[120] Marcus D S, Wang T H, Parker J, Csernansky J G, Morris J C and Buckner R L 2007 *J. Cogn. Neurosci.* **19** 1498–507
[121] Menze B H *et al* 2014 *IEEE Trans. Med. Imaging* **34** 1993–2024
[122] Klein A *et al* 2009 *Neuroimage* **46** 786–802
[123] Hazlett H C *et al* 2017 *Nature* **542** 348
[124] Di Martino A *et al* 2014 *Mol. Psychiatry* **19** 659
[125] Milham M P *et al* 2012 *Front. Syst. Neurosci.* **6** 62
[126] Gollub R L *et al* 2013 *Neuroinformatics* **11** 367–88
[127] Marek K *et al* 2011 *Prog. Neurobiol.* **95** 629–35
[128] Dagley A *et al* 2017 *Neuroimage* **144** 255–8
[129] Holmes A J *et al* 2015 *Sci. Data* **2** 150031
[130] Fischl B 2012 *Neuroimage* **62** 774–81
[131] Klein A and Tourville J 2012 *Front. Neurosci.* **6** 171
[132] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J S, Freymann J B, Farahani K and Davatzikos C 2017 *Sci. Data* **4** 170117
[133] Cocosco C A, Kollokian V, Kwan R K S, Pike G B and Evans A C 1997 *NeuroImage* **5** 425

[134] Mercier L, Del Maestro R F, Petrecca K, Araujo D, Haegelen C and Collins D L 2012 *Med. Phys.* **39** 3253–61

[135] Xiao Y, Fortin M, Unsgård G, Rivaz H and Reinertsen I 2017 *Med. Phys.* **44** 3875–82

[136] Sun L and Zhang S 2018 Deformable MRI-ultrasound registration using 3D convolutional neural network *Simulation, Image Processing and Ultrasound Systems for Assisted Diagnosis and Navigation* (Berlin: Springer) pp 152–8

[137] Hong J and Park H 2018 Non-linear approach for MRI to intra-operative US registration using structural skeleton *Simulation, Image Processing and Ultrasound Systems for Assisted Diagnosis and Navigation* (Berlin: Springer) pp 138–45

[138] Shieh C C *et al* 2019 *Med. Phys.* **46** 3799–3811

[139] Klein S, Staring M, Murphy K, Viergever M A and Pluim J P 2009 *IEEE Trans. Med. Imaging* **29** 196–205

[140] Vercauteren T, Pennec X, Perchant A and Ayache N 2008 Symmetric log-domain diffeomorphic registration: a demons-based approach *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 754–61

[141] Vercauteren T, Pennec X, Perchant A and Ayache N 2009 *Neuroimage* **45** S61–S72

[142] Ma X, Niu Y, Gu L, Wang Y, Zhao Y, Bailey J and Lu F 2020 *Pattern Recognit.* **110** 107332