ORIGINAL PAPER

Deep learning in medical image registration: a survey

Grant Haskins[1] · Uwe Kruger[1] · Pingkun Yan[1]

Abstract

The establishment of image correspondence through robust image registration is critical to many clinical tasks such as image fusion, organ atlas creation, and tumor growth monitoring and is a very challenging problem. Since the beginning of the recent deep learning renaissance, the medical imaging research community has developed deep learning-based approaches and achieved the state-of-the-art in many applications, including image registration. The rapid adoption of deep learning for image registration applications over the past few years necessitates a comprehensive summary and outlook, which is the main scope of this survey. This requires placing a focus on the different research areas as well as highlighting challenges that practitioners face. This survey, therefore, outlines the evolution of deep learning-based medical image registration in the context of both research challenges and relevant innovations in the past few years. Further, this survey highlights future research directions to show how this field may be possibly moved forward to the next level.

Keywords Image registration · Deep learning · Medical imaging · Convolutional neural networks

1 Introduction

Image registration is the process of transforming different image datasets into one coordinate system with matched imaging contents, which has significant applications in medicine. Registration may be necessary when analyzing a pair of images that were acquired from different viewpoints, at different times, or using different sensors/modalities [41,122]. Until recently, image registration was mostly performed manually by clinicians. However, many registration tasks can be quite challenging and the quality of manual alignments is highly dependent upon the expertise of the user, which can be clinically disadvantageous. To address the potential shortcomings of manual registration, automatic registration has been developed. Although other methods for automatic image registration have been extensively explored prior to (and during) the deep learning renaissance, deep learning has changed the landscape of image registration

research [4]. Ever since the success of AlexNet in the ImageNet challenge of 2012 [3], deep learning has allowed for state-of-the-art performance in many computer vision tasks including, but not limited to: object detection [84], feature extraction [37], segmentation [87], image classification [3], image denoising [112], and image reconstruction [115].

Initially, deep learning was successfully used to augment the performance of iterative, intensity-based registration [18,36,96]. Soon after this initial application, several groups investigated the intuitive application of reinforcement learning to registration [62,71,76]. Further, demand for faster registration methods later motivated the development of deep learning-based one-step transformation estimation techniques and challenges associated with procuring/generating ground truth data have recently motivated many groups to develop unsupervised frameworks for one-step transformation estimation [23,61]. One of the hurdles associated with this framework is the familiar challenge of image similarity quantification [38,106]. Recent efforts that use information theory-based similarity metrics [23], segmentations of anatomical structures [44], and generative adversarial network-like frameworks [30] to address this challenge have shown promising results.

Figure 1 shows the various categorizations of different deep learning-based registration methods. On the other hand, Fig. 2 shows the observed growing interest in deep
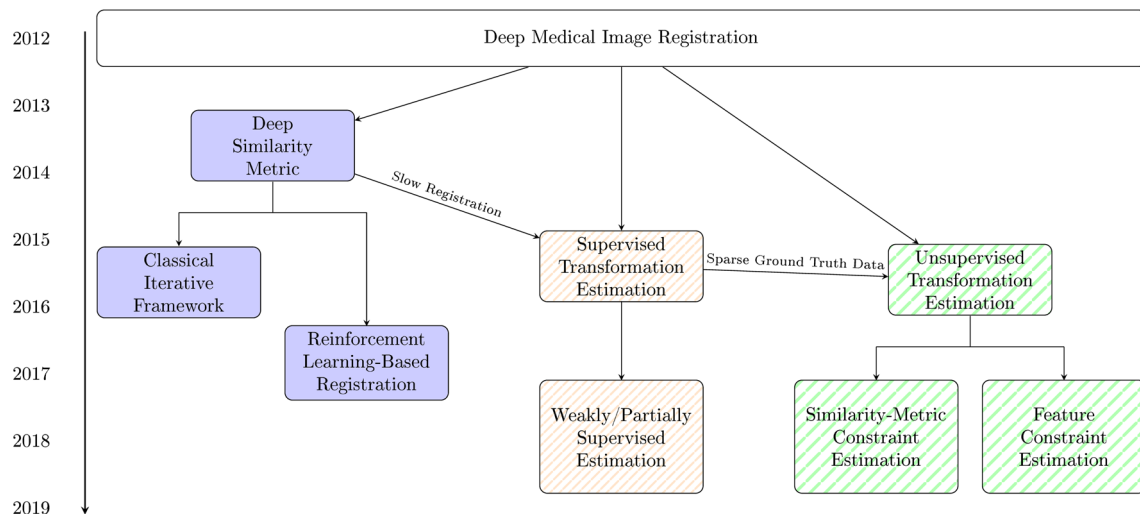
✉ Pingkun Yan
  yanp2@rpi.edu

1 Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Fig. 1 An overview of deep learning-based medical image registration broken down by approach type. The popular research directions are written in bold

Fig. 2 An overview of the number of deep learning-based image registration works and deep learning-based medical imaging works. The red line represents the trend line for medical imaging-based approaches, and the blue line represents the trend line for deep learning-based medical image registration approaches. The dotted line represents extrapolation (color figure online)
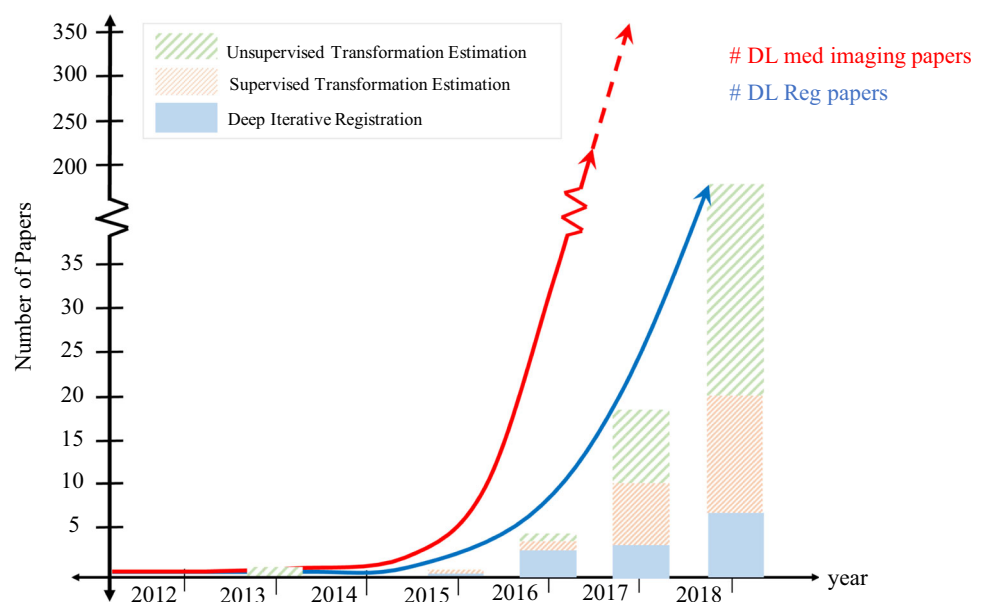
learning-based registration methods according to the number of published papers in recent years. As the trends visualized in Figs. 1 and 2 suggest, this field is moving very quickly to surmount the hurdles associated with deep learning-based medical image registration and several groups have already enjoyed significant successes for their applications [44,65,96].

Therefore, the purpose of this article is to comprehensively survey the field of deep learning-based medical image registration, highlight common challenges that practitioners face, and discuss future research directions that may address these challenges. Deep learning belongs to a class of machine learning that uses neural networks with a large number of layers to learn representations of data [34,91]. When discussing

neural networks, it is important to provide insight into the different types of neural networks that can be used for various applications, the notable architectures that were recently invented to tackle engineering problems, and the variety of strategies that are used for training neural networks. Therefore, this deep learning introduction section is divided into three sections: neural network types, network architectures, and training paradigms and strategies. Note that there are many publicly available libraries that can be used to build the networks described in the section, for example TensorFlow [1], MXNet [16], Keras [20], Caffe [49], and PyTorch [82]. Detailed discussion of deep learning-based medical image analysis and various deep learning research directions is outside of the scope of this article. Comprehensive

Table 1 Deep iterative registration methods overview

References	Learning	Transform	Modality	ROI	Model
[29]	Metric	Deformable	CT	Thorax	9-layer CNN
[10]	Metric	Deformable	CT	Lung	FCN
[96]	Metric	Deformable	MR	Brain	5-layer CNN
[18]	Metric	Deformable	CT/MR	Head	5-layer DNN
[92]	Metric	Rigid	MR/US	Abdominal	5-layer CNN
[36]	Metric	Rigid	MR/US	Prostate	14-layer CNN
[75]	Metric	Rigid	MR/US	Fetal brain	LSTM/STN
[55]	RL agent	Deformable	MR	Prostate	8-layer CNN
[62]	RL agent	Rigid	CT/CBCT	Spine/cardiac	8-layer CNN
[76]	Multiple RL agents	Rigid	X-ray/CT	Spine	Dilated FCN
[71]	RL agent	Rigid	MR/CT	Spine	Dueling network

RL denotes reinforcement learning

review articles that survey the application of deep learning to medical image analysis [59,63], reinforcement learning [51], and the application of GANs to medical image analysis [52] are recommended to the interested readers. In this article, the surveyed methods were divided into the following three categories: Deep Iterative Registration, Supervised Transformation Estimation, and Unsupervised Transformation Estimation. Following a discussion of the methods that belong to each of the aforementioned categories, future research directions and current trends are discussed in Sect. 6.

2 Deep iterative registration

Automatic intensity-based image registration requires both a metric that quantifies the similarity between a moving image and a fixed image and an optimization algorithm that updates the transformation parameters such that the similarity between the images is maximized. Prior to the deep learning renaissance, several manually crafted metrics were frequently used for such registration applications, including: sum of squared differences (SSD), cross-correlation (CC), mutual information (MI) [72,106], normalized cross-correlation (NCC), and normalized mutual information (NMI). Early applications of deep learning to medical image registration are direct extensions of the intensity-based registration framework [96,109,110]. Several groups later used a reinforcement learning paradigm to iteratively estimate a transformation [55,62,71,76] because this application is more consistent with how practitioners perform registration.

A description of both types of methods is given in Table 1. We will survey earlier methods that used deep similarity-based registration in Sect. 2.1 and then some more recently developed methods that use deep reinforcement learning-based registration in Sect. 2.2.

2.1 Deep similarity-based registration

In this section, methods that use deep learning to learn a similarity metric are surveyed. This similarity metric is inserted into a classical intensity-based registration framework with a defined interpolation strategy, transformation model, and optimization algorithm. A visualization of this overall framework is given in Fig. 3. The solid lines represent data flows that are required during training and testing, while the dashed lines represent data flows that are required only during training. Note that this is the case for the remainder of the figures in this article as well.

2.1.1 Overview of works

Although manually crafted similarity metrics perform reasonably well in the unimodal registration case, deep learning has been used to learn superior metrics. This section will first discuss approaches that use deep learning to augment the performance of unimodal intensity-based registration pipelines before multimodal registration.

2.1.1.1 Unimodal registration Wu et al. [109,110] were the first to use deep learning to obtain an application-specific similarity metric for registration. They extracted the features that are used for unimodal, deformable registration of 3D brain MR volumes using a convolutional-stacked autoencoder (CAE). They subsequently performed the registration using gradient descent to optimize the NCC of the two sets of features. This method outperformed diffeomorphic demons [104] and HAMMER [94]-based registration techniques.

Recently, Eppenhof et al. [29] estimated registration error for the deformable registration of 3D thoracic CT scans (inhale–exhale) in an end-to-end capacity. They used a 3D CNN to estimate the error map for inputted inhale–exhale

Fig. 3 A visualization of the registration pipeline for works that use deep learning to quantify image similarity in an intensity-based registration framework
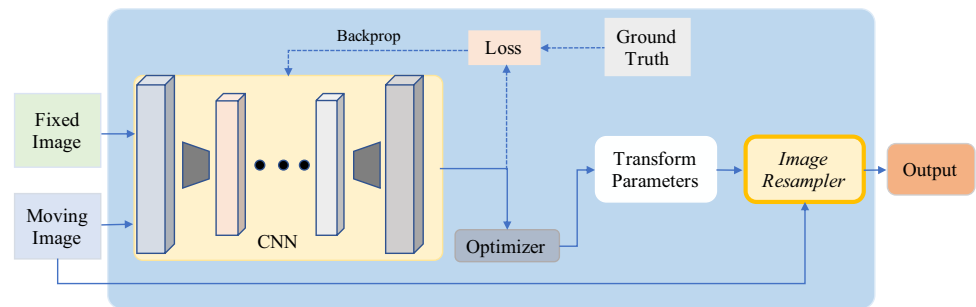
pairs of thoracic CT scans. Like the above method, only learned features were used in this work.

Instead, Blendowski et al. [10] proposed the combined use of both CNN-based descriptors and manually crafted MRF-based self-similarity descriptors for lung CT registration. Although the manually crafted descriptors outperformed the CNN-based descriptors, optimal performance was achieved using both sets of descriptors. This indicates that, in the unimodal registration case, deep learning may not outperform manually crafted methods. However, it can be used to obtain complementary information.

2.1.1.2 Multimodal registration The advantages of the application of deep learning to intensity-based registration are more obvious in the multimodal case, where manually crafted similarity metrics have had very little success.

Cheng et al. [17,18] recently used a stacked denoising autoencoder to learn a similarity metric that assesses the quality of the rigid alignment of CT and MR images. They showed that their metric outperformed NMI-optimization-based and local cross-correlation (LCC)-optimization-based for their application.

In an effort to explicitly estimate image similarity in the multimodal case, Simonovsky et al. [96] used a CNN to learn the dissimilarity between aligned 3D T1 and T2 weighted brain MR volumes. Given this similarity metric, gradient descent was used in order to iteratively update the parameters that define a deformation field. This method was able to outperform MI-optimization-based registration and set the stage for deep intensity-based multimodal registration.

Additionally, Sedghi et al. [92] performed the rigid registration of 3D US/MR (modalities with an even greater appearance difference than MR/CT) abdominal scans by using a 5-layer neural network to learn a similarity metric that is then optimized by Powell's method. This approach also outperformed MI-optimization-based registration.

Haskins et al. [36] learned a similarity metric for multimodal rigid registration of MR and transrectal US (TRUS) volumes by using a CNN to predict target registration error (TRE). Instead of using a traditional optimizer like the above methods, they used an evolutionary algorithm to explore the solution space prior to using a traditional optimization algo-

rithm because of the learned metric's lack of convexity. This registration framework outperformed MIND-optimization-based [38] and MI-optimization-based registration.

In stark contrast to the above methods, Wright et al. [75] used LSTM spatial co-transformer networks to iteratively register MR and US volumes group-wise. The recurrent spatial co-transformation occurred in three steps: image warping, residual parameter prediction, parameter composition. They demonstrated that their method is more capable of quantifying image similarity than a previous multimodal image similarity quantification method that uses self-similarity context descriptors [39].

2.1.2 Discussion and assessment

Recent works have confirmed the ability of neural networks to assess image similarity in multimodal medical image registration. The results achieved by the approaches described in this section demonstrate that deep learning can be successfully applied to challenging registration tasks. However, the findings from [10] suggest that learned image similarity metrics may be best suited to complement existing similarity metrics in the unimodal case. Further, it is difficult to use these iterative techniques for real-time registration.

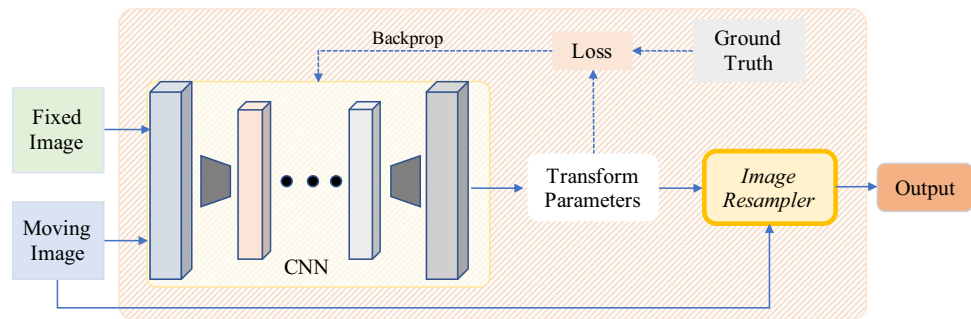## 2.2 Reinforcement learning-based registration

In this section, methods that use reinforcement learning for their registration applications are surveyed. Here, a trained agent is used to perform the registration as opposed to a pre-defined optimization algorithm. A visualization of this framework is given in Fig. 4. Reinforcement learning-based registration typically involves a rigid transformation model. However, it is possible to use a deformable transformation model.

Liao et al. [62] were the first to use reinforcement learning-based registration to perform the rigid registration of cardiac and abdominal 3D CT images and cone-beam CT (CBCT) images. They used a greedy supervised approach for end-to-end training with an attention-driven hierarchical strategy. Their method outperformed MI-based registration and semantic registration using probability maps.

Fig. 4 A visualization of the registration pipeline for works that use deep reinforcement learning to implicitly quantify image similarity for image registration. Here, an agent learns to map states to actions based on rewards that it receives from the environment
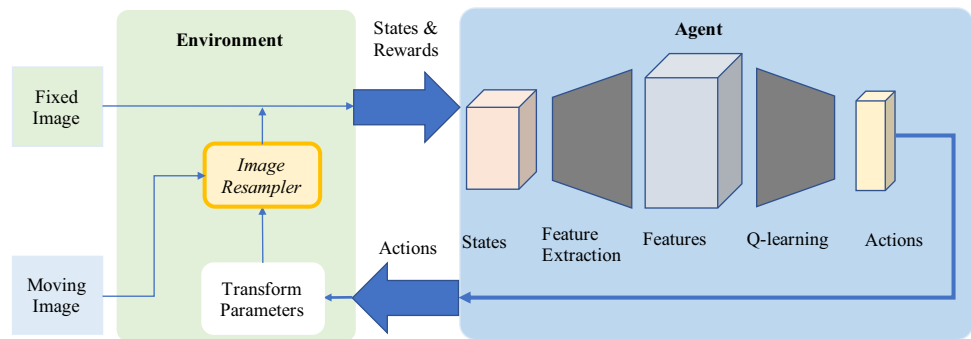
Fig. 5 A visualization of supervised single-step registration

Shortly after, Kai et al. [71] used a reinforcement learning approach to perform the rigid registration of MR/CT chest volumes. This approach is derived from Q-learning and leverages contextual information to determine the depth of the projected images. The network used in this method is derived from the dueling network architecture [108]. Notably, this work also differentiates between terminal and non-terminal rewards. This method outperforms registration methods that are based on iterative closest points (ICP), landmarks, Hausdorff distance, Deep Q Networks, and the Dueling Network [108].

Instead of training a single agent like the above methods, Miao et al. [76] used a multi-agent system in a reinforcement learning paradigm to rigidly register X-ray and CT images of the spine. They used an auto-attention mechanism to observe multiple regions and demonstrate the efficacy of a multi-agent system. They were able to significantly outperform registration approaches that used a state-of-the-art similarity metric given by [22].

As opposed to the above rigid registration-based works, Krebs et al. [55] used a reinforcement learning-based approach to perform the deformable registration of 2D and 3D prostate MR volumes. They used a low-resolution deformation model for the registration and fuzzy action control to influence the stochastic action selection. The low-resolution deformation model is necessary to restrict the dimensionality of the action space. This approach outperformed registration performed using the Elastix toolbox [53] and LCC Demons [69]-based registration techniques.

The use of reinforcement learning is intuitive for medical image registration applications. One of the principle challenges for reinforcement learning-based registration is the ability to handle high-resolution deformation fields. There are no such challenges for rigid registration. Because of the intuitive nature and recency of these methods, we expect that such approaches will receive more attention from the research community in the next few years.

3 Supervised transformation estimation

Despite the early success of the previously described approaches, the transformation estimation in these methods is iterative, which can lead to slow registration. [36]. This is especially true in the deformable registration case where the solution space is high dimensional [59]. This motivated the development of networks that could estimate the transformation that corresponds to optimal similarity in one step. However, fully supervised transformation estimation (the exclusive use of ground truth data to define the loss function) has several challenges that are highlighted in this section.

A visualization of supervised transformation estimation is given in Fig. 5, and a description of notable works is given in Table 2. This section first discusses methods that use fully supervised approaches in Sect. 3.1 and then discusses methods that use dual/weakly supervised approaches in Sect. 3.2.

3.1 Fully supervised transformation estimation

In this section, methods that used full supervision for single-step registration are surveyed. Using a neural network to

Table 2 Supervised transformation estimation methods

References	Supervision	Transform	Modality	ROI	Model
[114]	**Real transforms**	**Deformable**	**MR**	**Brain**	**FCN**
[13]	Real transforms	Deformable	MR	Brain	9-layer CNN
[70]	Real transforms	Deformable	MR	Abdominal	CNN
[86]	Real transforms	Deformable	MR	Cardiac	SVF-net
[99]	Synthetic transforms	Deformable	CT	Chest	RegNet
[28]	Synthetic transforms	Deformable	CT	Lung	U-net
[103]	Synthetic transforms	Deformable	MR	Brain/Cardiac	FlowNet
[47]	Synthetic transforms	Deformable	MR	Brain	GoogleNet
[102]	Synthetic transforms	Deformable	CT/US	Liver	DVFNet
[113]	Real + synthetic transforms	Deformable	MR	Brain	FCN
[97]	Synthetic transforms	Rigid	MR	Brain	6-layer CNN 10-layer FCN
[90]	Synthetic transforms	Rigid	MR	Brain	11-layer CNN ResNet-18
[119]	Synthetic transforms	Rigid	X-ray	Bone	17-layer CNN PDA module
[78]	Synthetic transforms	Rigid	X-ray/DDR	Bone	6-layer CNN
[15]	Synthetic transforms	Rigid	MR	Brain	AIRNet
[44]	Segmentations	Deformable	MR/US	Prostate	30-layer FCN
[40]	Segmentations + similarity metric	Deformable	MR/US	Prostate	U-net GAN
[42]	Segmentations + adversarial loss	Deformable	MR/US	Prostate	GAN
[31]	Real transforms + similarity metric	Deformable	MR	Brain	U-net
[111]	Synthetic transforms + adversarial loss	Rigid	MR/US	Prostate	GAN

Bold values rows use Diffeomorphisms

perform registration as opposed to an iterative optimizer significantly speeds up the registration process.

3.1.1 Overview of works

Several registration application requires deformable transformation models that often prohibit the use of traditional convolutional neural networks because of the computational expense associated with using FC-layers to make predictions in highly dimensional solution spaces [55]. Because the networks that are used to predict deformation fields are fully convolutional, the dimensionality of the solution space associated with a deformation field does not introduce additional computational constraints [114]. This section will first discuss approaches that use a rigid transformation model and then discuss approaches that use a deformable transformation model.

3.1.1.1 Rigid registration Miao et al. [77,78] were the first to use deep learning to predict rigid transformation parameters. They used a CNN to predict the transformation matrix associated with the rigid registration of 2D/3D X-ray attenuation maps and 2D X-ray images. Hierarchical regression is proposed in which the six transformation parameters are partitioned into three groups. Ground truth data were synthesized in this approach by transforming aligned data. This

is the case for the next three approaches that are described as well. This approach outperformed MI, CC, and gradient correlation (GC)-optimization-based registration approaches with respect to both accuracy and computational efficiency. The improved computational efficiency is due to the use of a forward pass through a neural network instead of an optimization algorithm to perform the registration.

Recently, Chee et al. [15] used a CNN to predict the transformation parameters used to rigidly register 3D brain MR volumes. In their framework, affine image registration network (AIRNet), the MSE between the predicted and ground truth affine transforms is used to train the network. They were able to outperform MI-optimization-based registration for both the unimodal and multimodal cases.

That same year, Salehi et al. [90] used a deep residual regression network, a correction network, and a bivariant geodesic distance-based loss function to rigidly register T1 and T2 weighted 3D fetal brain MRs for atlas construction. The use of the residual network to initially register the image volumes prior to the forward pass through the correction network allowed for an enhancement of the capture range of the registration. This approach was evaluated for both slice-to-volume registration and volume-to-volume registration. They validated the efficacy of their geodesic loss term and outperformed NCC-optimization-based registration.

Additionally, Zheng et al. [119] proposed the integration of a pairwise domain adaptation module (PDA) into a pre-trained CNN that performs the rigid registration of preoperative 3D X-ray images and intraoperative 2D X-ray images using a limited amount of training data. Domain adaptation was used to address the discrepancy between synthetic data that was used to train the deep model and real data.

Sloan et al. [97] used a CNN that is used to regress the rigid transformation parameters for the registration of T1 and T2 weighted brain MRs. Both unimodal and multimodal registration were investigated in this work. The parameters that constitute the convolutional layers that were used to extract low-level features in each image were only shared in the unimodal case. In the multimodal case, these parameters were learned separately. This approach also outperformed MI-optimization-based image registration.

3.1.1.2 Deformable registration

Unlike the previous section, methods that use both real and synthesized ground truth labels will be discussed. Methods that use clinical/publicly available ground truth labels for training are discussed first. This ordering is reflective of the fact that simulating realistic deformable transformations is more difficult than simulating realistic rigid transformations.

First, Yang et al. [114] predicted the deformation field with an FCN that is used to register 2D/3D intersubject brain MR volumes in a single step. A U-net-like architecture [87] was used in this approach. Further, they used large diffeomorphic metric mapping to provide a basis, used the initial momentum values of the pixels of the image volumes as the network input, and evolved these values to obtain the predicted deformation field. This method outperformed semi-coupled dictionary learning-based registration [11].

The following year, Rohe et al. [86] also used a U-net [87] inspired network to estimate the deformation field used to register 3D cardiac MR volumes. Mesh segmentations are used to compute the reference transformation for a given image pair, and SSD between the prediction and ground truth is used as the loss function. This method outperformed LCC Demons-based registration [69].

That same year, Cao et al. [13] used a CNN to map input image patches of a pair of 3D brain MR volumes to their respective displacement vector. The totality of these displacement vectors for a given image constitutes the deformation field that is used to perform the registration. Additionally, they used the similarity between inputted image patches to guide the learning process. Further, they used equalized active-point-guided sampling strategy that makes it so that patches with higher gradient magnitudes and displacement values are more likely to be sampled for training. This method outperforms SyN [6] and Demons [104]-based registration methods.

Recently, Jun et al. [70] used a CNN to perform the deformable registration of abdominal MR images to compensate for the deformation that is caused by respiration. This approach achieved registration results that are superior to those obtained using non-motion corrected registrations and local affine registration. Recently, unlike many of the other approaches discussed in this paper, Yang et al. [113] quantified the uncertainty associated with the deformable registration of 3D T1 and T2 weighted brain MRs using a low-rank Hessian approximation of the variational Gaussian distribution of the transformation parameters. This method was evaluated on both real and synthetic data.

Just as deep learning practitioners use random transformations to enhance the diversity of their dataset, Sokooti et al. [99] used random DVFs to augment their dataset. They used a multi-scale CNN to predict a deformation field. This deformation is used to perform intra-subject registration of 3D chest CT images. This method used late fusion as opposed to early fusion, in which the patches are concatenated and used as the input to the network. The performance of their method is competitive with B-Spline-based registration [99].

Such approaches have notable, but also limited ability to enhance the size and diversity of datasets. These limitations motivated the development of more sophisticated ground truth generation. The rest of the approaches described in this section use simulated ground truth data for their applications.

For example, Eppenhof et al. [28] used a 3D CNN to perform the deformable registration of inhale–exhale 3D lung CT image volumes. A series of multi-scale, random transformations of aligned image pairs eliminate the need for manually annotated ground truth data while also maintaining realistic image appearance. Further, as is the case with other methods that generate ground truth data, the CNN can be trained using relatively few medical images in a supervised capacity.

Unlike the above works, Uzunova et al. [103] generated ground truth data using statistical appearance models (SAMs). They used a CNN to estimate the deformation field for the registration of 2D brain MRs and 2D cardiac MRs and adapt FlowNet [26] for their application. They demonstrated that training FlowNet using SAM-generated ground truth data resulted in superior performance to CNNs trained using either randomly generated ground truth data or ground truth data obtained using the registration method described in [27].

Unlike the other methods in this section that use random transformations or manually crafted methods to generate ground truth data, Ito et al. [47] used a CNN to learn plausible deformations for ground truth data generation. They evaluated their approach on the 3D brain MR volumes in the ADNI dataset and outperformed the MI-optimization-based approach proposed in [45].

Fig. 6 A visualization of deep single-step registration where the agent is trained using dual supervision. The loss function is determined using both a metric that quantifies image similarity and ground truth data
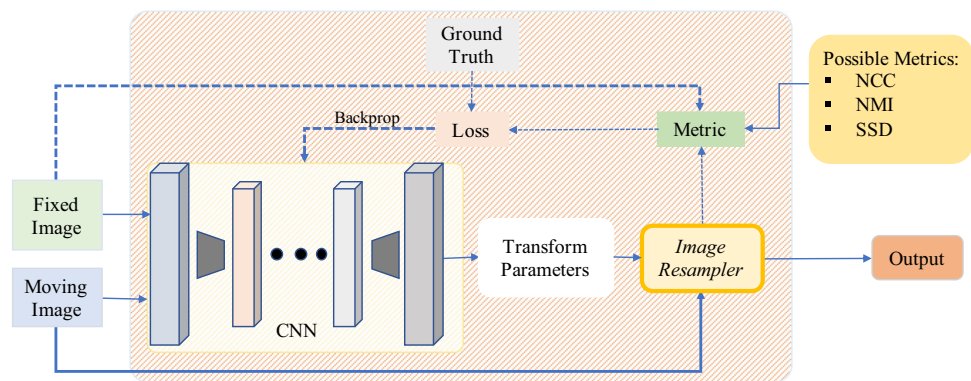
Fig. 7 A visualization of an adversarial image registration framework. Here, the generator is trained using output from the discriminator. The discriminator takes the form of a learned metric here
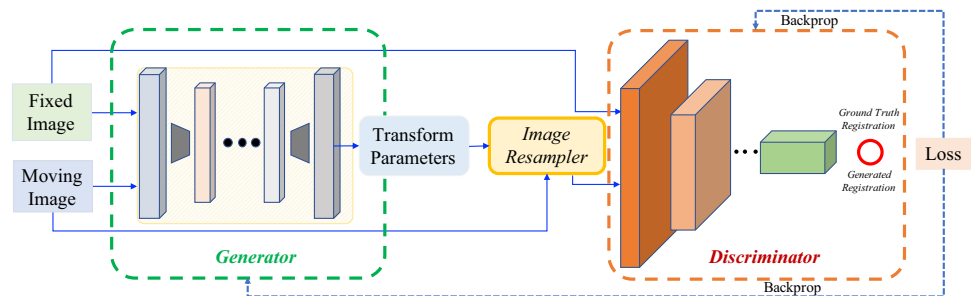
3.1.2 Discussion and assessment

Supervised transformation estimation has allowed for real-time, robust registration across applications. However, such works are not without their limitations. Firstly, the quality of the registrations using this framework is dependent on the quality of the ground truth registrations. The quality of these labels is, of course, dependent upon the expertise of the practitioner. Furthermore, these labels are fairly difficult to obtain because there are relatively few individuals with the expertise necessary to perform such registrations. Transformations of training data and the generation of synthetic ground truth data can address such limitations. However, it is important to ensure that simulated data is sufficiently similar to clinical data. These challenges motivated the development of partially supervised/unsupervised approaches, which will be discussed next.

3.2 Dual/weakly supervised transformation estimation

Dual supervision refers to the use of both ground truth data and some metric that quantifies image similarity to train a model. On the other hand, weak supervision refers to using the overlap of segmentations of corresponding anatomical structures to design the loss function. This section will discuss the contributions of such works in Sect. 3.2.1 and then discuss the overall state of this research direction in Sect. 3.2.2.

3.2.1 Overview of works

First, this section will discuss methods that use dual supervised and then will discuss methods that use weak supervision. Recently, Fan et al. [31] used hierarchical, dual-supervised learning to predict the deformation field for 3D brain MR registration. They amend the traditional U-Net architecture [87] by using "gap-filling" (i.e., inserting convolutional layers after the U-type ends or the architecture) and coarse-to-fine guidance. This approach leveraged both the similarity between the predicted and ground truth transformations, and the similarity between the warped and fixed images to train the network. The architecture detailed in this method outperformed the traditional U-Net architecture, and the dual supervision strategy is verified by ablating the image similarity loss function term. A visualization of dual-supervised transformation estimation is given in Fig. 6.

On the other hand, Yan et al. [111] used a framework that is inspired by the GAN [35] to perform the rigid registration of 3D MR and TRUS volumes. In this work, the generator was trained to estimate a rigid transformation. While, the discriminator was trained to discern between images that were aligned using the ground truth transformations and images that were aligned using the predicted transformations. Both Euclidean distance to ground truth and an adversarial loss term are used to construct the loss function in this method. Note that the adversarial supervision strategy that was used in this approach is similar to the ones that are used in a number of unsupervised works that will be described in the next sec-

Fig. 8 A visualization of deep single-step registration where the agent is trained using label similarity (i.e., weak supervision). Manually annotated data (segmentations) are used to define the loss function used to train the network
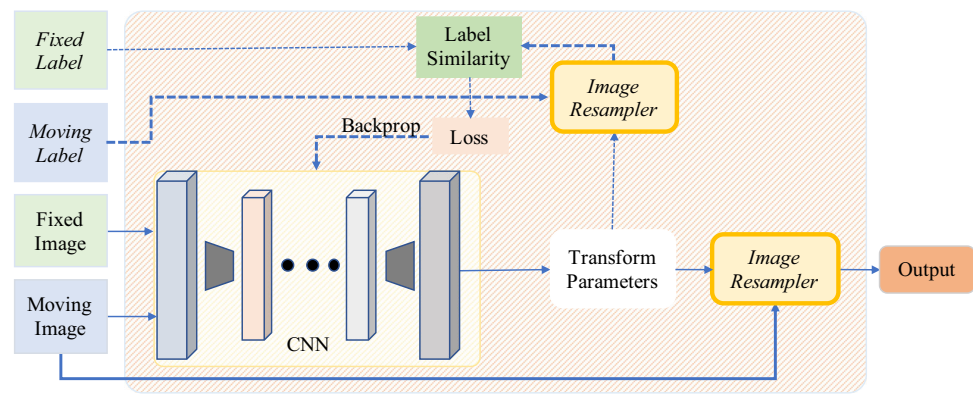
tion. A visualization of adversarial transformation estimation is given in Fig. 7.

Unlike the above methods that used dual supervision, Hu et al. [43,44] recently used label similarity to train their network to perform MR-TRUS registration. In their initial work, they used two neural networks: local-net and global-net to estimate the global affine transformation with 12 degrees of freedom and the local dense deformation field, respectively [43]. The local-net uses the concatenation of the transformation of the moving image given by the global-net and the fixed image as its input. However, in their later work [44], they combine these networks in an end-to-end framework. This method outperformed NMI-optimization-based and NCC-based registration. A visualization of weakly supervised transformation estimation is given in Fig. 8. In another work, Hu et al. [42] simultaneously maximized label similarity and minimized an adversarial loss term to predict the deformation for MR-TRUS registration. This regularization term forces the predicted transformation to result in the generation of a realistic image. Using the adversarial loss as a regularization term is likely to successfully force the transformation to be realistic given proper hyper parameter selection. The performance of this registration framework was inferior to the performance of their previous registration framework described above. However, they showed that adversarial regularization is superior to standard bending energy-based regularization. Similar to the above method, Hering et al. [40] built upon the progress made with respect to both dual and weak supervision by introducing a label and similarity metric-based loss function for cardiac motion tracking via the deformable registration of 2D cine-MR images. Both segmentation overlap and edge-based normalized gradient fields distance were used to construct the loss function in this approach. Their method outperformed a multilevel registration approach similar to the one proposed in [88].

3.2.2 Discussion and assessment

Direct transformation estimation marked a major breakthrough for deep learning-based image registration. With full

supervision, promising results have been obtained. However, at the same time, those techniques require a large amount of detailed annotated images for training. Partially/weakly supervised transformation estimation methods alleviated the limitations associated with the trustworthiness and expense of ground truth labels. However, they still require manually annotated data (e.g., ground truth and/or segmentations). On the other hand, weak supervision allows for similarity quantification in the multimodal case. Further, partial supervision allows for the aggregation of methods that can be used to assess the quality of a predicted registration. As a result, there is growing interest in these research areas.

4 Unsupervised transformation estimation

Despite the success of the methods described in the previous sections, the difficult nature of the acquisition of reliable ground truth remains a significant hindrance [103]. This has motivated a number of different groups to explore unsupervised approaches [24,61]. One key innovation that has been useful to these works is the spatial transformer network (STN) [48]. Several methods use an STN to perform the deformations associated with their registration applications [32,58]. This section discusses unsupervised methods that utilize image similarity metrics (Sect. 4.1) and feature representations of image data (Sect. 4.2) to train their networks. A description of notable works is given in Table 3.

4.1 Similarity metric-based unsupervised transformation estimation

4.1.1 Standard methods

This section begins by discussing approaches that use a common similarity metric with common regularization strategies to define their loss functions. Later in the section, approaches that use more complex similarity metric-based strategies

Table 3 Unsupervised transformation estimation methods

References	Loss function	Transform	Modality	ROI	Model
[50]	SSD	Deformable	CT	Chest	Multi-scale CNN
[33]	UB SSD	Deformable	MR	Brain	19-layer FCN
[118]	**MSD**	**Deformable**	**MR**	**Brain**	**ICNet**
[95]	MSE	Deformable	SEM	Neurons	11-layer CNN
[21]	**MSE**	**Deformable**	**MR**	**Brain**	**VoxelMorph**
[93]	MSE	Deformable	MR	Cardiac cine	8-layer FCNet
[58]	CC	Deformable	MR	Brain	FAIM
[61]	NCC	Deformable	MR	Brain	8-layer FCN
[12]	NCC	Deformable	CT, MR	Pelvis	U-Net
[24]	NCC	Deformable	MR	Cardiac cine	DIRNet
[23]	NCC	Deformable	MR	Cardiac cine	DLIR
[32]	NCC	Deformable	X-ray, MR	Bone cardiac cine	U-Net STN
[101]	L2 distance + image gradient	Deformable	MR, US	Brain	FCN
[81]	Predicted TRE	Deformable	CT	Head/neck	FCN
[30]	BCE	Deformable	MR	Brain	GAN
[73]	NMI + SSIM + VGG outputs	Deformable	MR, FA/color fundus	Cardiac retinal	GAN
[74]	NMI + SSIM + VGG outputs + BCE	Deformable	X-ray	Bone	GAN
[117]	MSE AE output	Deformable	ssEM	Neurons	CAE STN
[110]	MSE stacked AE outputs	Deformable	MR	Brain	Stacked AE
[109]	NCC of ISA outputs	Deformable	MR	Brain	Stacked ISA
[56]	Log likelihood	Deformable	MR	Brain	cVAE STN
[67]	SSD MIND + PCANet outputs	Deformable	CT, MR	Chest brain	FCN PCANet
[54]	SSD VGG outputs	Rigid	MR	Brain	CNN MLP

Bold values rows use diffeomorphisms

Fig. 9 A visualization of deep single-step registration where the network is trained using a metric that quantifies image similarity. Therefore, the approach is unsupervised
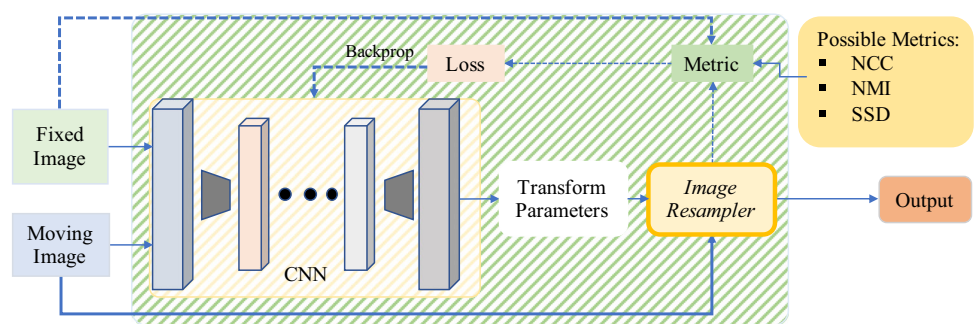
are discussed. A visualization of standard similarity metric-based transformation estimation is given in Fig. 9.

Inspired to overcome the difficulty associated with obtaining ground truth data, Li et al. [60,61] trained an FCN to perform deformable intersubject registration of 3D brain MR volumes using "self-supervision." NCC between the warped and fixed images and several common regularization terms (e.g., smoothing constraints) constitute the loss function in this method. Although many manually defined similarity metrics fail in the multimodal case (with the occasional exception of MI), they are often suitable for the unimodal case. The method detailed in this work outperforms

Advanced Neuroimaging Tools (ANTs)-based registration [7] and the deep learning methods proposed by Sokooti et al. [99] (discussed previously) and Yoo et al. [117] (discussed in the next section).

Further, de Vos et al. [24] used NCC to train an FCN to perform the deformable registration of 4D cardiac cine-MR volumes. A DVF is used in this method to deform the moving volume. Their method outperforms registration that is performed using the Elastix toolbox [53].

In another work, de Vos et al. [23] use a multistage, multiscale approach to perform unimodal registration on several datasets. NCC and a bending-energy regularization term are

used to train the networks that predict an affine transformation and subsequent coarse-to-fine deformations using a B-Spline transformation model. In addition to validating their multistage approach, they show that their method outperforms registration that is performed using the Elastix toolbox [53] with and without bending energy.

The unsupervised deformable registration framework used by Ghosal et al. [33] minimizes the upper bound of the SSD (UB SSD) between the warped and fixed 3D brain MR images. The design of their network was inspired by the SKIP architecture [68]. This method outperforms log-demons-based registration.

Shu et al. [95] used a coarse-to-fine, unsupervised deformable registration approach to register images of neurons that are acquired using a scanning electron microscope (SEM). The mean squared error (MSE) between the warped and fixed volumes is used as the loss function here. Their approach is competitive with and faster than the sift flow framework [64].

Sheikhjafari et al. [93] used learned latent representations to perform the deformable registration of 2D cardiac cine-MR volumes. Deformation fields are thus obtained by embedding. This latent representation is used as the input to a network that is composed of eight fully connected layers to obtain the transformation. The sum of absolute errors (SAE) is used as the loss function. Here, the registration performance was seen to be influenced by the B-Spline grid spacings. This method outperforms a moving mesh correspondence-based method described in [83].

Stergios et al. [100] used a CNN to both linearly and locally register inhale–exhale pairs of lung MR volumes. Therefore, both the affine transformation and the deformation are jointly estimated. The loss function is composed of an MSE term and regularization terms. Their method outperforms several state-of-the-art methods that do not utilized ground truth data, including Demons [69], SyN [6], and a deep learning-based method that uses an MSE loss term. Further, the inclusion of the regularization terms is validated by an ablation study.

The successes of deep similarity metric-based unsupervised registration motivated Neylon et al. [81] to use a neural network to learn the relationship between image similarity metric values and TRE when registering CT image volumes. This is done in order to robustly assess registration performance. The network was able to achieve subvoxel accuracy in 95% of cases. Similarly inspired, Balakrishnan et al. [8,9] proposed a general framework for unsupervised image registration, which can be either unimodal or multimodal theoretically. The neural networks are trained using a selected, manually defined image similarity metric (e.g., NCC, NMI, etc.).

In a follow-up paper, Dalca et al. [21] casted deformation prediction as variational inference. Diffeomorphic integration is combined with a transformer layer to obtain

a velocity field. Squaring and rescaling layers are used to integrate the velocity field to obtain the predicted deformation. MSE is used as the similarity metric that, along with a regularization term, defines the loss function. Their method outperforms ANT-based registration [7] and the deep learning-based method described in [8].

Shortly after, Kuang et al. [58] used a CNN and STN inspired framework to perform the deformable registration of T1-weighted brain MR volumes. The loss function is composed of a NCC term and a regularization term. This method uses Inception modules, a low capacity model, and residual connections instead of skip connections. They compare their method with VoxelMorph (the method proposed by Balakrishnan et al., described above) [9] and uTIlzReg GeoShoot [105] using the LBPA40 and Mindboggle 101 datasets and demonstrate superior performance with respect to both.

Building upon the progress made by the previously described metric-based approaches, Ferrante et al. [32] used a transfer learning-based approach to perform unimodal registration of both X-ray and cardiac cine images. In this work, the network is trained on data from a source domain using NCC as the primary loss function term and tested in a target domain. They used a U-net like architecture [87] and an STN [48] to perform the feature extraction and transformation estimation, respectively. They demonstrated that transfer learning using either domain as the source or the target domain produces effective results. This method outperformed registration obtained using the Elastix toolbox [53] with parameters determined using grid search.

Although applying similarity metric-based approaches to the multimodal case is difficult, Sun et al. [101] proposed an unsupervised method for 3D MR/US brain registration that uses a 3D CNN that consists of a feature extractor and a deformation field generator. This network is trained using a similarity metric that incorporates both pixel intensity and gradient information. Further, both image intensity and gradient information are used as inputs into the CNN.

4.1.2 Extensions

Cao et al. [12] also applied similarity metric-based training to the multimodal case. Specifically, they used intra-modality image similarity to supervise the multimodal deformable registration of 3D pelvic CT/MR volumes. The NCC between the moving image that is warped using the ground truth transformation and the moving image that is warped using the predicted transformation is used as the loss function. This work utilizes "dual" supervision (i.e., the intra-modality supervision previously described is used for both the CT and the MR images). This is not to be confused with the dual supervision strategies described earlier.

Inspired by the limiting nature of the asymmetric transformations that typical unsupervised methods estimate, Zhang et

al. [118] used their network inverse-consistent deep network (ICNet) to learn the symmetric diffeomorphic transformations for each of the brain MR volumes that are aligned into the same space. Different from other works that use standard regularization strategies, this work introduces an inverse-consistent regularization term and an anti-folding regularization term to ensure that a highly weighted smoothness constraint does not result in folding. Finally, the MSD between the two images allows this network to be trained in an unsupervised manner. This method outperformed SyN-based registration [6], Demons-based registration [69], and several deep learning-based approaches.

The next three approaches described in this section used a GAN for their applications. Unlike the GAN-based approaches described previously, these methods use neither ground truth data nor manually crafted segmentations. Mahapatra et al. [73] used a GAN to implicitly learn the density function that represents the range of plausible deformations of cardiac cine images and multimodal retinal images (retinal color fundus images and fluorescein angiography (FA) images). In addition to NMI, structural similarity index measure (SSIM), and a feature perceptual loss term (determined by the SSD between VGG outputs), the loss function is comprised of conditional and cyclic constraints, which are based on recent advances involving the implementation of adversarial frameworks. Their approach outperforms registration that is performed using the Elastix toolbox [53] and the method proposed by de Vos et al. [24].

Further, Fan et al. [30] used a GAN to perform unsupervised deformable image registration of 3D brain MR volumes. Unlike most other unsupervised works that use a manually crafted similarity metric to determine the loss function and unlike the previous approach that used a GAN to ensure that the predicted deformation is realistic, this approach uses a discriminator to assess the quality of the alignment. This approach outperforms diffeomorphic demons and SyN registration on every dataset except for MGH10. Further, the use of the discriminator for supervision of the registration network is superior to the use of ground truth data, SSD, and CC on all datasets.

Different from the hitherto previously described works (not just the GAN-based ones), Mahapatra et al. [74] proposed simultaneous segmentation and registration of chest X-rays using a GAN framework. The network takes three inputs: reference image, floating image, and the segmentation mask of the reference image and outputs the segmentation mask of the transformed image, and the deformation field. Three discriminators are used to assess the quality of the generated outputs (deformation field, warped image, and segmentation) using cycle consistency and a dice metric. The generator is additionally trained using NMI, SSIM, and a feature perceptual loss term.

Finally, instead of predicting a deformation field given a fixed parameterization as the other methods in this section do, Jiang et al. [50] used a CNN to learn an optimal parameterization of an image deformation using a multi-grid B-Spline method and L1-norm regularization. They use this approach to parameterize the deformable registration of 4D CT thoracic image volumes. Here, SSD is used as the similarity metric and L-BFGS-B is used as the optimizer. The convergence rate using the parameterized deformation model obtained using the proposed method is faster than the one obtained using a traditional L1-norm regularized multi-grid parameterization.

4.1.3 Discussion and assessment

Image similarity-based unsupervised image registration has received a lot of attention from the research community recently because it bypasses the need for expert labels of any kind. This means that the performance of the model will not depend on the expertise of the practitioner. Further, extensions of the original similarity metric-based method that introduce more sophisticated similarity metrics (e.g., the discriminator of a GAN) and/or regularization strategies have yielded promising results. However, it is still difficult to quantify image similarity for multimodal registration applications. As a result, the scope of unsupervised, image similarity-based works is largely confined to the unimodal case. Given that multimodal registration is often needed in many clinical applications, we expect to see more papers in the near future that will tackle this challenging problem.

## 4.2 Feature-based unsupervised transformation estimation

In this section, methods that use learned feature representations to train neural networks are surveyed. Like the methods surveyed in the previous section, the methods surveyed in this section do not require ground truth data. In this section, approaches that create unimodal registration pipelines are presented first. Then, an approach that tackles multimodal image registration is discussed. A visualization of featured-based transformation estimation is given in Fig. 10.

4.2.1 Unimodal registration

Yoo et al. [117] used an STN to register serial-section electron microscopy images (ssEMs). An autoencoder is trained to reconstruct fixed images and the L2 distance between reconstructed fixed images and corresponding warped moving images is used along with several regularization terms to construct the loss function. This approach outperforms the bUnwarpJ registration technique [5] and the Elastic registration technique [89].

Fig. 10 A visualization of feature-based unsupervised image registration. Here, a feature extractor is used to map inputted images to a feature space to facilitate the prediction of transformation parameters
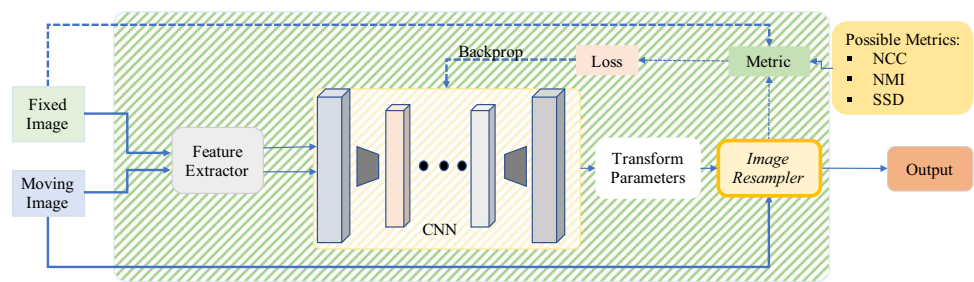
In the same year, Liu et al. [67] proposed a tensor-based MIND method using a principle component analysis-based network (PCANet) [14] for both unimodal and multimodal registration. Both inhale–exhale pairs of thoracic CT volumes and multimodal pairs of brain MR images are used for experimental validation of this approach. MI- and residual complexity (RC)-based [79], and the original MIND-based [38] registration techniques were outperformed by the proposed method.

Krebs et al. [56,57] performed the registration of 2D brain and cardiac MRs and bypassed the need for spatial regularization using a stochastic latent space learning approach. A conditional variational autoencoder [25] is used to ensure that the parameter space follows a prescribed probability distribution. The negative log likelihood of the fixed image given the latent representation and the warped volume and KL divergence of the latent distribution from a prior distribution is used to define the loss function. This method outperforms the Demons technique [69] and the deep learning method described in [8].

4.2.2 Multimodal registration

Unlike all of the other methods described in this section, Kori et al. perform feature extraction and affine transformation parameter regression for the multimodal registration of 2D T1 and T2 weighted brain MRs in an unsupervised capacity using pre-trained networks [54]. The images are binarized, and then, the Dice score between the moving and the fixed images is used as the cost function. As the appearance difference between these two modalities is not significant, the use of these pre-trained models can be reasonably effective.

4.2.3 Discussion and assessment

Performing multimodal image registration in an unsupervised capacity is significantly more difficult than performing unimodal image registration because of the difficulty associated with using manually crafted similarity metrics to quantify the similarity between the two images, and generally using the unsupervised techniques described above to establish/detect voxel-to-voxel correspondence. The use of unsupervised learning to learn feature representations

to determine an optimal transformation has generated significant interest from the research community recently. Along with the previously discussed unsupervised image registration method, we expect feature-based unsupervised registration to continue to generate significant interest from the research community. Further, extension to the multimodal case (especially for applications that use image with significant appearance differences) is likely to be a prominent research focus in the next few years.

5 Research trends and future directions

In this section, we summarize the current research trends and future directions of deep learning in medical image registration. As we can see from Fig. 2, some research trends have emerged. First, deep learning-based medical image registration seems to be following the observed trend for the general application of deep learning to medical image analysis. Second, unsupervised transformation estimation methods have been garnering more attention recently from the research community. Further, deep learning-based methods consistently outperform traditional optimization-based techniques [80]. Based on the observed research trends, we speculate that the following research directions will receive more attention in the research community.

5.1 Deep adversarial image registration

We further speculate that GANs will be used more frequently in deep learning-based image registration in the next few years. As described above, GANs can serve several different purposes in deep learning-based medical image registration: using a discriminator as a learned similarity metric, ensuring that predicted transformations are realistic, and using a GAN to perform image translation to transform a multimodal registration problem into a unimodal registration problem.

GAN-like frameworks have been used in several works to directly train transformation predicting neural networks. Several recent works [30,111] use a discriminator to discern between aligned and misaligned image pairs. Although the training paradigm borrows from an unsupervised training strategy, the discriminator requires pre-aligned image pairs.

Therefore, it will have limited success in multimodal or challenging unimodal applications where it is difficult to register images. Because discriminators are trained to assign all misaligned image pairs the same label, they will likely be unable to model a spectrum of misalignments. Despite this limitation, the application of GANs to medical image registration is still quite promising and will be described below.

Unconstrained deformation field prediction can result in warped moving images with unrealistic organ appearances. A common approach is to add the L2 norm of the predicted deformation field, its gradient, or its Laplacian to the loss function. However, the use of such regularization terms may limit the magnitude of the deformations that neural networks are able to predict. Therefore, Hu et al. [42] explored the use of a GAN-like framework to produce realistic deformations. Constraining the deformation prediction using a discriminator results in superior performance relative to the use of L2 norm regularization in that work.

Lastly, GANs can be used to map medical images in a source domain (e.g., MR) to a target domain (e.g., CT) [19,46,66,116], regardless of whether or not paired training data are available [121]. This image appearance reduction technique would be advantageous because many unimodal unsupervised registration methods use similarity metrics that often fail in the multimodal case. If image translation is performed as a preprocessing step, then commonly used similarity metrics could be used to define the loss function of transformation predicting networks.

5.2 Reinforcement learning-based registration

We also project that reinforcement learning will also be more commonly used for medical image registration in the next few years because it is very intuitive and can mimic the manner in which physicians perform registration. It should be noted that there are some unique challenges associated with deep learning-based medical image registration: including the dimensionality of the action space in the deformable registration case. However, we believe that such limitations are surmountable because there is already one proposed method that uses reinforcement learning-based registration with a deformable transformation model [55].

5.3 Raw imaging domain registration

This article has focused on surveying methods performing registration using reconstructed images. However, we speculate that it is possible to incorporate reconstruction into an end-to-end deep learning-based registration pipeline. In 2016, Wang [107] postulated that deep neural networks could be used to perform image reconstruction. Further, several works [85,98,115,120] recently demonstrated the ability of deep learning to map data points in the raw data domain to

the reconstructed image domain. Therefore, it is reasonable to expect that registration pipelines that take raw data as input and output registered, reconstructed images can be developed within the next few years.

6 Conclusion

In this article, the recent works that use deep learning to perform medical image registration have been examined. As each application has its own unique challenges, the creation of the deep learning-based frameworks must be carefully designed. Many deep learning-based medical image registration applications share similar challenges including the lack of a robust similarity metric for multimodal applications, in which there are significant image appearance differences and/or different fields of view (e.g., MR-TRUS registration) [36], the lack of availability of large datasets, the challenge associated with obtaining segmentations and ground truth registrations, and quantifying the uncertainty of a model's prediction. Application-specific similarity metrics, patch-wise frameworks, unsupervised approaches, and variational autoencoder inspired registration frameworks are examples of popular solutions to these challenges. Furthermore, despite the sophistication of many of the methods discussed in this survey, resampling and interpolation are often not among the components of registration that are learned by the neural network. While researchers started to pay attention to this aspect [2], we expect more works to incorporate these components into their deep learning-based methods as the field continues to mature. Recent successes have demonstrated the impact of the application of deep learning to medical image registration. This trend can be observed across medical imaging applications. Many future exciting works are sure to build on the recent progress that has been outlined in this paper.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: a system for large-scale machine learning. In: OSDI, vol. 16, pp. 265–283 (2016)
2. Ali, S., Rittscher, J.: Conv2Warp: an unsupervised deformable image registration with continuous convolution and warping. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) Machine Learning in Medical Imaging, pp. 489–497. Springer International Publishing, Cham (2019)
3. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B.C., Awwal, A.A.S., Asari, V.K.: The history began from alexnet: a comprehensive survey on deep learning approaches. Preprint (2018). arXiv:1803.01164
4. Ambinder, E.P.: A history of the shift toward full computerization of medicine. J. Oncol. Pract. **1**(2), 54–56 (2005)
5. Arganda-Carreras, I., Sorzano, C.O., Marabini, R., Carazo, J.M., Ortiz-de Solorzano, C., Kybic, J.: Consistent and elastic regis-

tration of histological sections using vector-spline regularization. In: International Workshop on Computer Vision Approaches to Medical Image Analysis, pp. 85–95. Springer, Berlin (2006)

6. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)

7. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)

8. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9252–9260 (2018)

9. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: A learning framework for deformable medical image registration. Preprint (2018). arXiv:1809.05231

10. Blendowski, M., Heinrich, M.P.: Combining MRF-based deformable registration and deep binary 3D-CNN descriptors for large lung motion estimation in COPD patients. Int. J. Comput. Assisted Radiol. Surg. **14**, 1–10 (2018)

11. Cao, T., Singh, N., Jojic, V., Niethammer, M.: Semi-coupled dictionary learning for deformation prediction. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 691–694. IEEE, New York (2015)

12. Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., Shen, D.: Deep learning based inter-modality image registration supervised by intra-modality similarity. Preprint (2018). arXiv:1804.10735

13. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D.: Deformable image registration based on similarity-steered CNN regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 300–308. Springer, New York (2017)

14. Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: Pcanet: A simple deep learning baseline for image classification? IEEE Trans. Image Process. **24**(12), 5017–5032 (2015)

15. Chee, E., Wu, J.: Airnet: self-supervised affine registration for 3D medical images using neural networks. Preprint (2018). arXiv:1810.02583

16. Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., Zhang, Z.: Mxnet: a flexible and efficient machine learning library for heterogeneous distributed systems. Preprint (2015). arXiv:1512.01274

17. Cheng, X., Zhang, L., Zheng, Y.: Deep similarity learning for multimodal medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2016)

18. Cheng, X., Zhang, L., Zheng, Y.: Deep similarity learning for multimodal medical images. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **6**(3), 248–252 (2018)

19. Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., Choo, J.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8789–8797 (2018)

20. Chollet, F., et al.: Keras (2015). https://keras.io

21. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. Preprint (2018). arXiv:1805.04605

22. De Silva, T., Uneri, A., Ketcha, M., Reaungamornrat, S., Kleinszig, G., Vogt, S., Aygun, N., Lo, S., Wolinsky, J., Siewerdsen, J.: 3D–2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. Phys. Med. Biol. **61**(8), 3009 (2016)

23. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised

24. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 204–212. Springer, Berlin (2017)

25. Doersch, C.: Tutorial on variational autoencoders. Preprint (2016). arXiv:1606.05908

26. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)

27. Ehrhardt, J., Schmidt-Richberg, A., Werner, R., Handels, H.: Variational registration. In: Bildverarbeitung für die Medizin 2015, pp. 209–214. Springer, Berlin (2015)

28. Eppenhof, K.A., Pluim, J.P.: Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans. Med. Imaging **38**, 1097–1105 (2018)

29. Eppenhof, K.A.J., Pluim, J.P.: Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks. J. Med. Imaging **5**(2), 024003 (2018b)

30. Fan, J., Cao, X., Xue, Z., Yap, P.-T., Shen, D.: Adversarial similarity network for evaluating image alignment in deep learning based registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 739–746. Springer, Berlin (2018)

31. Fan, J., Cao, X., Yap, P.-T., Shen, D.: Birnet: brain image registration using dual-supervised fully convolutional networks. Preprint (2018). arXiv:1802.04692

32. Ferrante, E., Oktay, O., Glocker, B., Milone, D.H.: On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains. In: International Workshop on Machine Learning in Medical Imaging, pp. 294–302. Springer, Berlin (2018)

33. Ghosal, S., Ray, N.: Deep deformable registration: enhancing accuracy by fully convolutional neural net. Pattern Recogn. Lett. **94**, 81–86 (2017)

34. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT Press, Cambridge (2016)

35. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 2672–2680. Curran Associates, Inc. (2014). http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

36. Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P.A., Wood, B.J., Yan, P.: Learning deep similarity metric for 3D MR–TRUS image registration. Int. J. Comput. Assist. Radiol. Surg. **14**, 417–425 (2019)

37. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

38. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A.: Mind: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**(7), 1423–1435 (2012)

39. Heinrich, M.P., Jenkinson, M., Papież, B. W., Brady, M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 187–194. Springer, Berlin (2013)

40. Hering, A., Kuckertz, S., Heldmann, S., Heinrich, M.: Enhancing label-driven deep deformable image registration with local dis-

tance metrics for state-of-the-art cardiac motion tracking. Preprint (2018). arXiv:1812.01859

41. Hill, D.L., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Phys. Med. Biol. **46**(3), R1–R45 (2001)

42. Hu, Y., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Vercauteren, T., Noble, J.A., Barratt, D.C.: Adversarial deformation regularization for training image registration neural networks. Preprint (2018). arXiv:1805.10665

43. Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T.: Label-driven weakly-supervised learning for multimodal deformarle image registration. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1070–1074. IEEE, New York (2018)

44. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Med. Image Anal. **49**, 1–13 (2018c)

45. Ikeda, K., Ino, F., Hagihara, K.: Efficient acceleration of mutual information computation for nonrigid registration using CUDA. IEEE J. Biomed. Health Inf. **18**(3), 956–968 (2014)

46. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv Preprint (2017)

47. Ito, M., Ino, F.: An automated method for generating training sets for deep learning based image registration. In: The 11th International Joint Conference on Biomedical Engineering Systems and Technologies—Volume 2: BIOIMAGING, pp. 140–147. INSTICC SciTePress (2018)

48. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K: Spatial transformer networks. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems, pp. 2017–2025. Curran Associates, Inc. (2015). http://papers.nips.cc/paper/5854-spatial-transformer-networks.pdf

49. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM, New York (2014)

50. Jiang, P., Shackleford, J.A.: CNN driven sparse multi-level b-spline image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9281–9289 (2018)

51. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: a survey. J. Artif. Intell. Res. **4**, 237–285 (1996)

52. Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., Mukhopadhyay, A.: Gans for medical image analysis. Preprint (2018). arXiv:1809.06222

53. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)

54. Kori, A., Krishnamurthi, G.: Zero shot learning for multi-modal real time image registration (2019). arXiv:1908.06213

55. Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A.: Robust non-rigid registration through agent-based action learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 344–352. Springer, Berlin (2017)

56. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H.: Learning structured deformations using diffeomorphic registration. Preprint (2018) arXiv:1804.07172

57. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H.: Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Deep Learning in Medical Image

Analysis and Multimodal Learning for Clinical Decision Support, pp. 101–109. Springer, Berlin (2018)

58. Kuang, D., Schmah, T.: Faim—a convnet method for unsupervised 3D medical image registration. Preprint (2018). arXiv:1811.09243

59. Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G.B., Seo, J.B., Kim, N.: Deep learning in medical imaging: general overview. Korean J. Radiol. **18**(4), 570–584 (2017)

60. Li, H., Fan, Y.: Non-rigid image registration using fully convolutional networks with deep self-supervision. Preprint (2017). arXiv:1709.00799

61. Li, H., Fan, Y.: Non-rigid image registration using self-supervised fully convolutional networks without training data. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1075–1078. IEEE, New York (2018)

62. Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., Comaniciu, D.: An artificial agent for robust image registration. In: AAAI, pp. 4168–4175 (2017)

63. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)

64. Liu, C., Yuen, J., Torralba, A.: Sift flow: dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 978–994 (2011)

65. Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., Wang, J.: Applications of deep learning to mri images: a survey. Big Data Min. Anal. **1**(1), 1–18 (2018)

66. Liu, M.-Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems, pp. 700–708. Curran Associates, Inc. (2017) http://papers.nips.cc/paper/6672-unsupervised-image-to-image-translation-networks.pdf

67. Liu, Q., Leung, H.: Tensor-based descriptor for image registration via unsupervised network. In: 2017 20th International Conference on Information Fusion (Fusion), pp. 1–7. IEEE, New York (2017)

68. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)

69. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., (ADNI, A. D. N. I., et al.): LCC-demons: a robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage **81**, 470–483 (2013)

70. Lv, J., Yang, M., Zhang, J., Wang, X.: Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study. Br. J. Radiol. **91**, 20170788 (2018)

71. Ma, K., Wang, J., Singh, V., Tamersoy, B., Chang, Y.-J., Wimmer, A., Chen, T.: Multimodal image registration with deep context reinforcement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 240–248. Springer, Berlin (2017)

72. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging **16**(2), 187–198 (1997)

73. Mahapatra, D.: Elastic registration of medical images with gans. Preprint (2018). arXiv:1805.02369

74. Mahapatra, D., Ge, Z., Sedai, S., Chakravorty, R.: Joint registration and segmentation of X-ray images using generative adversarial networks. In: International Workshop on Machine Learning in Medical Imaging, pp. 73–80. Springer, Berlin (2018)

75. Matthew, J., Hajnal, J.V., Rueckert, D., Schnabel, J.A.: LSTM spatial co-transformer networks for registration of 3D fetal US and MR brain images. In: Data Driven Treatment Response Assess-

ment and Preterm, Perinatal, and Paediatric Image Analysis, pp. 149–159. Springer, Berlin (2018)

76. Miao, S., Piat, S., Fischer, P., Tuysuzoglu, A., Mewes, P., Mansi, T., Liao, R.: Dilated FCN for multi-agent 2D/3D medical image registration. Preprint (2017). arXiv:1712.01651

77. Miao, S., Wang, Z.J., Liao, R.: A cnn regression approach for real-time 2D/3D registration. IEEE Trans. Med. Imaging **35**(5), 1352–1363 (2016a)

78. Miao, S., Wang, Z.J., Zheng, Y., Liao, R.: Real-time 2D/3D registration via CNN regression. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1430–1434. IEEE, New York (2016)

79. Myronenko, A., Song, X.: Intensity-based image registration by minimizing residual complexity. IEEE Trans. Med. Imaging **29**(11), 1882–1891 (2010)

80. Nazib, A., Fookes, C., Perrin, D.: A comparative analysis of registration tools: traditional vs. deep learning approach on high resolution tissue cleared data. Preprint (2018). arXiv:1810.08315

81. Neylon, J., Min, Y., Low, D.A., Santhanam, A.: A neural network approach for fast, automated quantification of dir performance. Med. Phys. **44**(8), 4126–4138 (2017)

82. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NIPS-W (2017)

83. Punithakumar, K., Boulanger, P., Noga, M.: A gpu-accelerated deformable image registration algorithm with applications to right ventricular segmentation. IEEE Access **5**, 20374–20382 (2017)

84. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems, pp. 91–99. Curran Associates, Inc. (2015) http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf

85. Rivenson, Y., Zhang, Y., Günaydın, H., Teng, D., Ozcan, A.: Phase recovery and holographic image reconstruction using deep learning in neural networks. Light Sci. Appl. **7**(2), 17141 (2018)

86. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 266–274. Springer, Berlin (2017)

87. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, Berlin (2015)

88. Rühaak, J., Heldmann, S., Kipshagen, T., Fischer, B.: Highly accurate fast lung CT registration. In: Medical Imaging 2013: Image Processing, Volume 8669, pp. 86690Y. International Society for Optics and Photonics (2013)

89. Saalfeld, S., Fetter, R., Cardona, A., Tomancak, P.: Elastic volume reconstruction from series of ultra-thin microscopy sections. Nat. Methods **9**(7), 717 (2012)

90. Salehi, S.S.M., Khan, S., Erdogmus, D., Gholipour, A.: Real-time deep registration with geodesic loss. Preprint (2018). arXiv:1803.05982

91. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)

92. Sedghi, A., Luo, J., Mehrtash, A., Pieper, S., Tempany, C.M., Kapur, T., Mousavi, P., Wells III, W.M.: Semi-supervised deep metrics for image registration. Preprint (2018). arXiv:1804.01565

93. Sheikhjafari, A., Noga, M., Punithakumar, K., Ray, N.: Unsupervised deformable image registration with fully connected generative neural network. In: International Conference on Medical Imaging with Deep Learning (2018)

94. Shen, D.: Image registration by local histogram matching. Pattern Recogn. **40**(4), 1161–1172 (2007)

95. Shu, C., Chen, X., Xie, Q., Han, H.: An unsupervised network for fast microscopic image registration. In: Medical Imaging 2018: Digital Pathology, vol. 10581, p. 105811D. International Society for Optics and Photonics (2018)

96. Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., Komodakis, N.: A deep metric for multimodal registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 10–18. Springer, Berlin (2016)

97. Sloan, J.M., Goatman, K.A., Siebert, J.P.: Learning rigid image registration-utilizing convolutional neural networks for medical image registration. In: 11th International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 89–99. SCITEPRESS-Science and Technology Publications (2018)

98. Smith, J.T., Yao, R., Sinsuebphon, N., Rudkouskaya, A., Un, N., Mazurkiewicz, J., Barroso, M., Yan, P., Intes, X.: Fast fit-free analysis of fluorescence lifetime imaging via deep learning. Proc. Natl. Acad. Sci. **116**(48), 24019–24030 (2019)

99. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P., Išgum, I., Staring, M.: Nonrigid image registration using multi-scale 3D convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 232–239. Springer, Berlin (2017)

100. Stergios, C., Mihir, S., Maria, V., Guillaume, C., Marie-Pierre, R., Stavroula, M., Nikos, P.: Linear and deformable image registration with 3D convolutional neural networks. In: Image Analysis for Moving Organ, Breast, and Thoracic Images, pp. 13–22. Springer, Berlin (2018)

101. Sun, L., Zhang, S.: Deformable MRI-ultrasound registration using 3D convolutional neural network. In: Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation, pp. 152–158. Springer, Berlin (2018)

102. Sun, Y., Moelker, A., Niessen, W.J., van Walsum, T.: Towards robust CT-ultrasound registration using deep learning methods. In: Understanding and Interpreting Machine Learning in Medical Image Computing Applications, pp. 43–51. Springer, Berlin (2018)

103. Uzunova, H., Wilms, M., Handels, H., Ehrhardt, J.: Training CNNS for image registration from few samples with model-based data augmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 223–231. Springer, Berlin (2017)

104. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)

105. Vialard, F.-X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. Int. J. Comput. Vis. **97**(2), 229–241 (2012)

106. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. Int. J. Comput. Vis. **24**(2), 137–154 (1997)

107. Wang, G.: A perspective on deep imaging. Preprint (2016). arXiv:1609.04375

108. Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., De Freitas, N.: Dueling network architectures for deep reinforcement learning. Preprint (2015). arXiv:1511.06581

109. Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., Shen, D.: Unsupervised deep feature learning for deformable registration of MR brain images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 649–656. Springer, Berlin (2013)

110. Wu, G., Kim, M., Wang, Q., Munsell, B.C., Shen, D.: Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans. Biomed. Eng. **63**(7), 1505–1516 (2016)

111. Yan, P., Xu, S., Rastinehad, A.R., Wood, B.J.: Adversarial image registration with application for MR and TRUS image fusion. Preprint (2018). arXiv:1804.11024

112. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M.K., Zhang, Y., Sun, L., Wang, G.: Low dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans. Med. Imaging **37**, 1348–1357 (2018)

113. Yang, X.: Uncertainty quantification, image synthesis and deformation prediction for image registration. Ph.D. Thesis, The University of North Carolina at Chapel Hill (2017)

114. Yang, X., Kwitt, R., Niethammer, M.: Fast predictive image registration. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (eds.) Deep Learning and Data Labeling for Medical Applications, pp. 48–57. Springer International Publishing, Cham (2016)

115. Yao, R., Ochoa, M., Intes, X., Yan, P.: Deep compressive macroscopic fluorescence lifetime imaging. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 908–911. IEEE, New York (2018)

116. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: unsupervised dual learning for image-to-image translation. arXiv Preprint (2017)

117. Yoo, I., Hildebrand, D.G., Tobin, W.F., Lee, W.-C.A., Jeong, W.-K.: ssEMnet: serial-section electron microscopy image registration using a spatial transformer network with learned features. In: Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., Nascimento, J.C., Cardoso, J.S., Belagiannis, V., Lu, Z. (eds.) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 249–257. Springer International Publishing, Cham (2017)

118. Zhang, J.: Inverse-consistent deep networks for unsupervised deformable image registration. arXiv Preprint (2018). arXiv:1809.03443

119. Zheng, J., Miao, S., Wang, Z.J., Liao, R.: Pairwise domain adaptation module for CNN-based 2-D/3-D registration. J. Med. Imaging **5**(2), 021204 (2018)

120. Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S.: Image reconstruction by domain-transform manifold learning. Nature **555**(7697), 487 (2018)

121. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv Preprint (2017)

122. Zitova, B., Flusser, J.: Image registration methods: a survey. Image Vis. Comput. **21**(11), 977–1000 (2003)

Grant Haskins is a graduate student at the Department of Biomedical Engineering at Rensselaer Polytechnic Institute. He graduated from UCSD with a Bachelor of Science degree in Physics with a specialization in Astrophysics and a minor in psychology. He also earned a Master of Science degree in Biomedical Engineering with a concentration in medical devices and diagnostics.

Uwe Kruger received the M.S. degree from the University of Essen, Duisburg, Germany, in 1996, and the Ph.D. degree from the University of Manchester, Manchester, UK, in 2002. He is currently a Professor of Practice in Biomedical Engineering with Rensselaer Polytechnic Institute, Troy, NY, USA. He has published over 160 research articles in leading international journals and conferences and wrote a monograph on the statistical monitoring of complex multivariate processes. He has four industrial patents and has been involved in numerous conference activities. His research includes multivariate statistics, data science, process monitoring, fault detection and diagnosis, system modeling, and developing innovative algorithmic as well as their application and industrial exploitation. Prof. Kruger is an Associate Editor of Control Engineering Practice.

Pingkun Yan is an Assistant Professor at the Department of Biomedical Engineering at Rensselaer Polytechnic Institute. He received his Ph.D. from National University of Singapore. His research interests include medical imaging informatics and image-guided intervention using machine learning and computer vision techniques. Dr. Yan has published over 90 peer-reviewed articles in well-recognized journals and top international conferences. His research work has been recognized by a number of awards including MICCAI best paper award and IJCARS-MICCAI best paper award. He is currently serving as an associate editor of multiple international journals, including Machine Vision and Applications and Neurocomputing. He co-organized six international workshops and four international journal special issues, and has served as a program committee member for more than 40 international conferences. He is a senior member of IEEE.