Deep Learning in Medical Image Registration: A Review

Yabo Fu[1], Yang Lei[1], Tonghe Wang[1,2], Walter J. Curran[1,2], Tian Liu[1,2] and Xiaofeng Yang[1,2*]
[1]Department of Radiation Oncology, Emory University, Atlanta, GA
[2]Winship Cancer Institute, Emory University, Atlanta, GA

Corresponding author:
Xiaofeng Yang, PhD
Department of Radiation Oncology
Emory University School of Medicine
1365 Clifton Road NE
Atlanta, GA 30322
E-mail: xiaofeng.yang@emory.edu

Abstract

This paper presents a review of deep learning (DL) based medical image registration methods. We summarized the latest developments and applications of DL-based registration methods in the medical field. These methods were classified into seven categories according to their methods, functions and popularity. A detailed review of each category was presented, highlighting important contributions and identifying specific challenges. A short assessment was presented following the detailed review of each category to summarize its achievements and future potentials. We provided a comprehensive comparison among DL-based methods for lung and brain registration using benchmark datasets. Lastly, we analyzed the statistics of all the cited works from various aspects, revealing the popularity and future trend of DL-based medical image registration.

1 Introduction

Image registration, also known as image fusion or image matching, is the process of aligning two or more images based on image appearances. Medical image registration seeks to find an optimal spatial transformation that best aligns the underlying anatomical structures. Medical image registration is used in many clinical applications such as image guidance [22, 123, 148, 170], motion tracking [13, 46, 172], segmentation [44, 57, 174, 171, 173, 176], dose accumulation [1, 153], image reconstruction [91] and so on. Medical image registration is a broad topic which can be grouped from various perspectives. From input image point of view, registration methods can be divided into unimodal, multimodal, interpatient, intra-patient (*e.g.* same- or different-day) registration. From deformation model point of view, registration methods can be divided in to rigid, affine and deformable methods. From region of interest (ROI) perspective, registration methods can be grouped according to anatomical sites such as brain, lung registration and so on. From image pair dimension perspective, registration methods can be divided into 3D to 3D, 3D to 2D and 2D to 2D/3D.

Different applications and registration methods face different challenges. For multi-modal image registration, it is difficult to design an accurate image similarity measures due to the inherent appearance differences between different imaging modalities. Inter-patient registration can be tricky since the underlying anatomical structures are different across patients. Different-day intra-patient registration is challenging due to image appearance changes caused by metabolic processes, bowel movement, patient gaining/losing weight and so on. It is crucial for the registration to be computationally efficient in order to provide real-time image guidance. Examples of such application include 3D-MR to 2D/3D-US prostate registration to guide brachytherapy catheter placement and 3D-CT to 2D X-ray registration in intraoperative surgeries. For segmentation and dose accumulation, it is important to ensure the registration has high spatial accuracy. Motion tracking can be used for motion management in radiotherapy such as patient-setup and treatment planning. Motion tracking could also be used to assess respiratory function through 4D-CT lung registration and to access cardiac function through myocardial tissue tracking. In addition, motion tracking could be used to compensate for irregular motion in image reconstruction. In terms of deformation model, rigid transformation is often too simple to represent the actual tissue deformation while free-form transformation is ill-conditioned and hard to regularize. One limitation of 2D-2D registration is it ignores the out-of-plane deformation. Nevertheless, 3D-3D registration is usually computationally demanding, resulting in slow registration.

Many methods have been proposed to deal with the above-mentioned challenges. Popular registration methods include optical flow [169, 167], demons [154], ANTs [3], HAMMER [131], ELASTIX [75] and so on. Scale invariant feature transform (SIFT) and mutual information (MI) have been proposed for multi-modal image similarity calculation [55]. For 3D image registration, GPU has been adopted to accelerate the computational speed [128]. Multiple transformation regularization methods including spatial smoothing [168], diffeomorphic [154], spline-based [147], FE-based [9] and other deformable models have been proposed. Though medical image registration has been extensively studied, it remains a hot research topic. The field of medical image registration has been evolving rapidly with hundreds of papers published each year. Recently, DL-based methods have changed the landscape of medical image processing research and achieved the-state-of-art performances in many applications[25, 27, 45, 58, 84, 85, 86, 88, 89, 97, 98, 156, 157, 158, 160, 161]. However, deep learning in medical image registration has not been extensively studied until the past three to four years. Though several review papers on deep learning in medical image analysis have been published [73, 93, 96, 105, 106, 121, 132, 182], there are very few review papers that are specific to deep learning in medical image registration [60]. The goal of this paper is to summarize the latest developments, challenges and trends in DL-based medical image registration methods. With this survey, we aim to

1. Summarize the latest developments in DL-based medical image registration.

2. Highlight contributions, identify challenges and outline future trends.

3. Provide detailed statistics on recent publications from different perspectives.

2 Deep Learning

2.1 Convolutional Neural Network

Convolutional neural network (CNN) is a class of deep neural networks with regularized multilayer perceptron. CNN uses convolution operation in place of general matrix multiplication in simple neural networks. The convolutional filters and operations in CNN make it suitable for visual imagery signal processing. Because of its excellent feature extraction ability, CNN is one of the most successful models for image analysis. Since the breakthrough of AlexNet [79], many variants of CNN have been proposed and have achieved the-state-of-art performances in various image processing tasks. A typical CNN usually consists of multiple convolutional layers, max pooling layers, batch normalization layers, dropout layers, a sigmoid or softmax layer. In each convolutional layer, multiple channels of feature maps were extracted by sliding trainable convolutional kernels across the input feature maps. Hierarchical features with high-level abstraction are extracted using multiple convolutional layers. These feature maps usually go through multiple fully connected layer before reaching the final decision layer. Max pooling layers are often used to reduce the image sizes and to promote spatial invariance of the network. Batch normalization is used to reduce internal covariate shift among the training samples. Weight regularization and dropout layers are used to alleviate data overfitting. The loss function is defined as the difference between the predicted and the target output. CNN is usually trained by minimizing the loss via gradient back propagation using optimization methods. Many different types of network architectures have been proposed to improve the performance of CNN [93]. U-Net proposed by Ronneberger *et al.* is among one of the most used network architectures [120]. U-Net was originally used to perform neuronal structures segmentation. U-Net adopts symmetrical contractive and expansive paths with skip connections between them. U-Net allows effective feature learning from a small number of training datasets. Later, He *et al.* proposed a residual network (ResNet) to ease the difficulty of training deep neural networks [61]. The difficulty in training deep networks is caused by gradient degradation and vanishing. They reformulated the layers as learning residual functions instead of directly fitting a desired underlying mapping. Inspired by residual network, Huang *et al.* later proposed a densely connected convolutional network (DenseNet) by connecting each layer to every other layer [68]. Inception module was first used in GoogLeNet to alleviate the problem of gradient vanishing and allow for more efficient computation of deeper networks [146]. Instead of performing convolution using a kernel with fixed size, an inception module uses multiple kernels of different sizes. The resulting feature maps were concatenated and processed by the next layer. Recently, attention gate was used in CNN to improve performance in image classification and segmentation [124]. Attention gate could learn to suppress irrelevant features and highlight salient features useful for a specific task.

2.2 Autoencoder

An autoencoder (AE) is a type of neural network that learns to copy its input to its output without supervision [7]. Autoencoder usually consists of an encoder which encodes the input into a low-dimensional latent state space and a decoder which restore the original input from the low-dimensional latent space. To prevent an autoencoder from learning an identity function, regularized autoencoders were invented. Examples of regularized autoencoders include sparse autoencoder, denoising autoencoder and contractive autoencoder [151]. Recently, convolutional autoencoder (CAE) was proposed to combine CNN with traditional autoencoders [15]. CAE replaces the fully connected layer in traditional AE with convolutional layers and transpose-convolutional layers. CAE has been used in multiple medical image processing tasks such as lesion detection, segmentation, image restoration [93]. Different from above-mentioned AEs, variational AE (VAE) is generative model that learns latent representation using a variational approach [65]. VAE has been used for anomaly detection [185] and image generation [30].

2.3 Recurrent Neural Network

A recurrent neural network (RNN) is a type of neural network that was used to model dynamic temporal behavior [53]. RNN is widely used for natural language processing [20]. Unlike feedforward networks such as CNN, RNN is suitable for processing temporal signal. The internal state of RNN was used to model and memorize previously processed information. Therefore, the output of RNN was dependent on not only its immediate input but also its input history. Long short-term memory (LSTM) is one type of RNN which has been used in image processing tasks [4]. Recently, Cho *et al.* proposed a simplified version of LSTM, called gated recurrent unit [18].

2.4 Reinforcement Learning

Reinforcement learning (RL) is a type of machine learning that focused on predicting the best actions to take given its current state in an environment [150]. RL is usually modeled as a Markov decision process using a set of environment states and actions. An artificial agent is trained to maximize its cumulative expected rewards. The training process often involves an exploration-exploitation tradeoff. Exploration means to explore the whole space to gather more information while exploitation means to explore the promising areas given current information. Q-learning is a model-free RL algorithm, which aims to learn a Q function that models the action-reward relationship. Bellman equation is often used in Q-learning for reward calculation. The Bellman equation calculates the maximum future reward as the immediate reward the agent gets for entering the current state plus a weighted maximum future reward for the next state. For image processing, the Q function is often modeled as CNN, which could encode input images as states and learn the Q function via supervised training [51, 78, 92, 108].

2.5 Generative Adversarial Network

A typical generative adversarial network (GAN) consists of two competing networks, a generator and a discriminator [56]. The generator is trained to generate artificial data that approximate a target data distribution from a low-dimensional latent space. The discriminator is trained to distinguish the artificial data from actual data. The discriminator encourages the generator to predict realistic data by penalizing unrealistic predictions via learning. Therefore, the discriminative loss could be considered as a dynamic network-based loss term. The generator and discriminator both are getting better during training to reach Nash equilibrium. Multiple variants of GAN include conditional GAN (cGan) [109], InfoGan [16], , CycleGAN [184], StarGan [19] and so on. In medical imaging, GAN has been used to perform image synthesis for inter- or intra-modality, such as MR to synthetic CT [84, 89], CT to synthetic MR [27, 83], CBCT to synthetic CT [58], non-attenuation correction (non-AC) PET to CT [26], low-dose PET to synthetic full-dose PET [88], non-AC PET to AC PET [28], low-dose CT to full-dose CT [159] and so on. In medical image registration, GAN is usually used to either provide additional regularization or translate multi-modal registration to unimodal registration. Out of medical imaging, GAN has been widely used in many other fields including science, art, games and so on.

3 Deep learning in medical image registration

DL-based registration methods can be classified according to deep learning properties, such as network architectures (CNN, RL, GAN etc.), training process (supervised, unsupervised etc.), inference types (iterative, one-shot prediction), input image sizes (patch-based, whole image-based), output types (dense transformation, sparse transformation on control points, parametric regression of transformation model etc.) and so on. In this paper, we classified DL-based medical image registration methods according to its methods, functions and popularity in to seven categories, including 1) RL-based methods, 2) Deep similarity-based methods, 3) Supervised transformation predication, 4) Unsupervised transformation prediction, 5) GAN in medical image registration, 6) Registration validation using deep learning, and 7) Other learning-based methods. In each category, we provided a comprehensive table, listing all the surveyed works belonging to this category and summarizing their important features.

Before we delve into the details of each category, we provided a detailed overview of DL-based medical image registration methods with their corresponding components and features in Fig. 1. The purpose of Fig. 1 is to give the readers an overall understanding of each category by putting its important features side by side with each other. CNN was initially designed to process highly structured datasets such as images, which are usually expressed by regular grid-sampling data points. Therefore, almost all cited methods have utilized convolutional kernels in their deep learning design. This explains why the CNN module is in the middle of Fig. 1.

Works cited in this review were collected from various databases, including Google Scholar, PubMed, Web of Science, Semantic Scholar and so on. To collect as many works as possible, we used a variety of keywords including but not limited to machine learning, deep learning, learning-based, convolutional neural network, image registration, image fusion, image alignment, registration validation, registration error prediction, motion tracking, motion management and so on. We totally collected over 150 papers that are closely related to deep learning-based medical image registration. Most of these works were published between the year of 2016 and 2019. The number of publications is plotted against year by stacked bar charts in Fig. 2. Number of papers were counted by categories. The total number
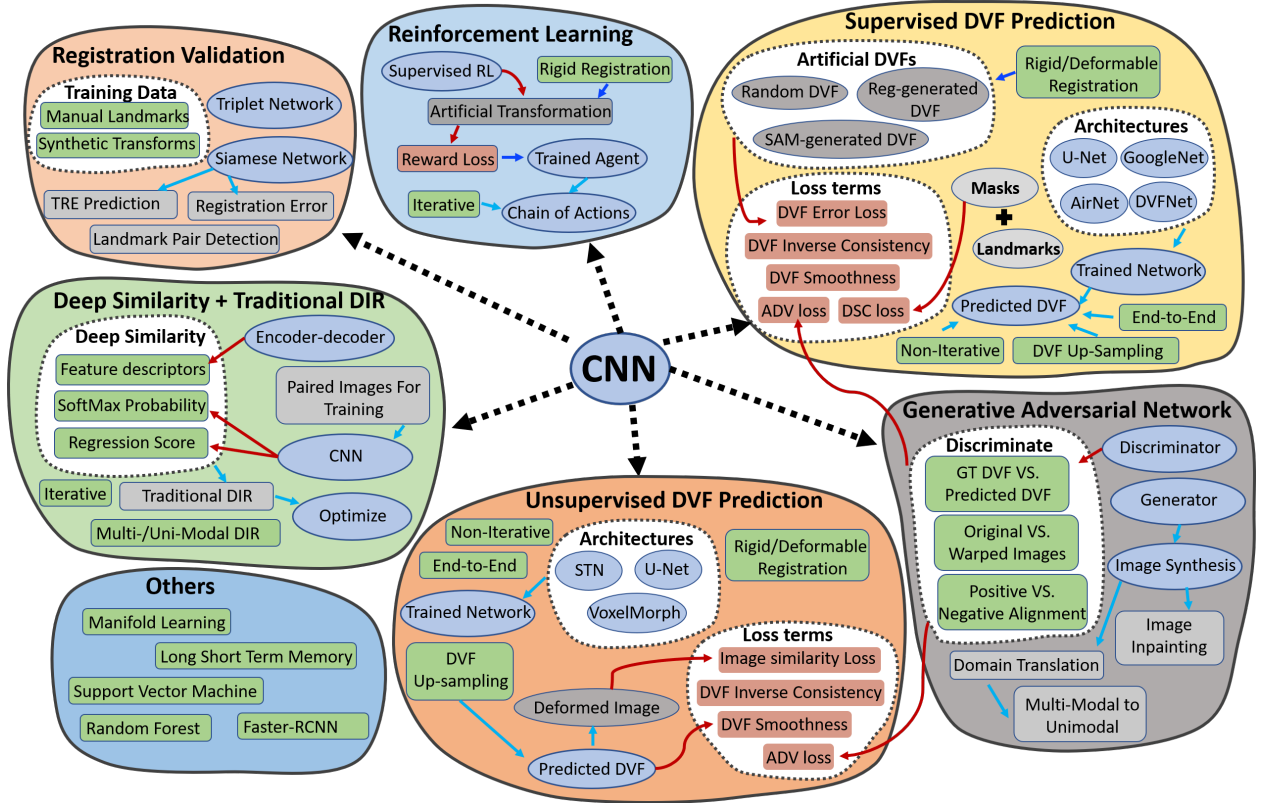
Fig. 1. Overview of seven categories of DL-based methods in medical image registration

of publications has grown dramatically over the last few years. Fig. 2 shows a clear trend of increasing interest in supervised transformation prediction (SupCNN) and unsupervised transform prediction (UnsupCNN). Meanwhile, GAN are gradually gaining popularity. On the other hand, the number of papers of RL-based medical image registration has decreased in 2019, which may indicate decreasing interest in RL for medical image registration. The 'DeepSimilarity' in Fig. 2 represents the category of deep similarity-based registration methods. The number of papers in this category has also increased, however, only slightly as compared to 'SupCNN' and 'UnsupCNN' categories. In addition, more and more studies were published on using DL for medical image registration validations.

3.1 Deep similarity-based methods

Conventional intensity-based similarity metrics include sum-of-square distance (SSD), mean square distance (MSD), (normalized) cross correlation (CC), and (normalized) mutual information (MI). Generally, conventional similarity measures work quite well for unimodal image registration where the image pair shares the same intensity distribution such as CT-CT, MR-MR image registration. However, noise and artifacts in images such as US and CBCT often cause conventional similarity measures to perform poorly even in unimodal image registration. Metrics such as SSD and MSD does not work for multi-modal image registration. To develop a similarity measure for multi-modality image registration, handcrafted descriptors such as MI were proposed. To improve its performance, a variety of MI variants such as correlation ratio-based MI [54], contextual conditioned MI [118] and modality independent neighborhood descriptor (MIND) [63] have been proposed. Recently, CNN has achieved huge success in tasks such as image classification and segmentation problems. However, CNN has not been widely used in image registration tasks until the last three to four years. To take the advantage of CNN, several groups tried to replace the traditional image similarity measures such as SSD, MAE and MI with DL-based similarity measures, achieving promising registration results. In the following section, we described several important works that attempted to use DL-based similarity measures in medical image registration.

3.1.1 Overview of works

Cheng *et al.* proposed a deep similarity learning network to train a binary classifier [17]. The
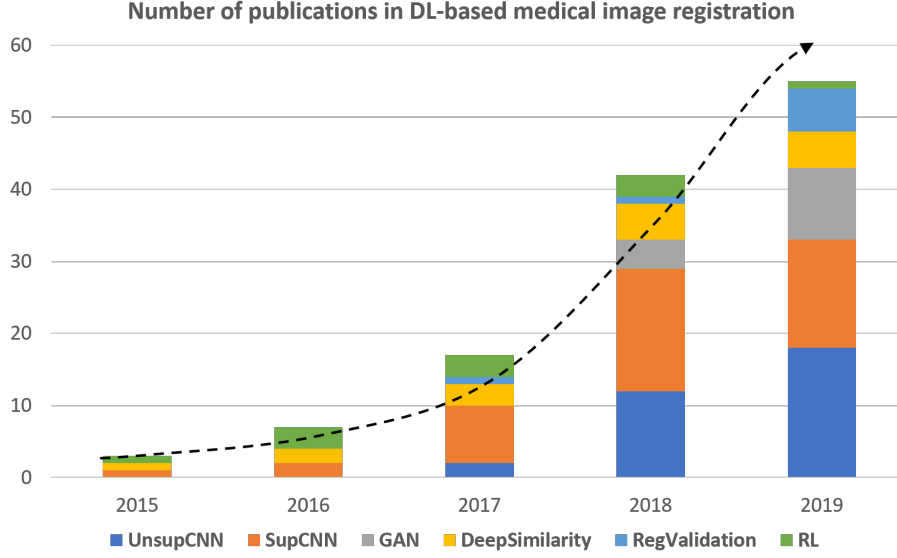
Fig. 2. Overview of number of publications in DL-based medical image registration. The dotted line indicates increase interest in DL-based registration methods over the years. 'DeepSimilarity' is the category of using DL-based similarity measures in traditional registration frameworks. 'RegValidation' represents the category of using DL for registration validation.

Table 1 Overview of deep similarity-based methods

References	ROI	Dimension	Modality	Transformation	Supervision
[135]	Brain	3D-3D	T1-T2	Deformable	Supervised
[164]	Brain	3D-3D	MR	Deformable	Unsupervised
[137]	Brain	3D-3D	MR	Rigid, Deformable	Supervised
[17]	Brain	2D-2D	MR-CT	Rigid	Supervised
[59]	Prostate	3D-3D	MR-US	Rigid	Supervised
[125]	Brain	2D-2D	MR	Rigid	Weakly Supervised
[42]	Brain, HN, Abdomen	3D-3D	MR, CT	Deformable	Weakly Supervised

network was trained to learn the correspondence of two image patches from CT-MR image pair. The continuous probabilistic value was used as the similarity score. Similarly, Simonovsky *et al.* proposed a 3D similarity network using a few aligned image pairs [135]. The network was trained to classify whether an image pair is aligned or not. They observed that hinge loss performed better than cross entropy. The learnt deep similarity metric was then used to replace MI in traditional deformable image registration (DIR) for brain T1-T2 registration. It is important to ensure the smoothness of first order derivative in order to fit the deep similarity metrics into traditional DIR frameworks. The gradient of the deep similarity metric with respect to transformation was calculated using chain rule. They found out that high overlap of neighboring patches led to smoother and more stable derivatives. They have trained the network using IXI brain datasets and tested it using a completely independent datasets called ALBERTs in order to show the good generality of the learnt metric. They showed that the learnt deep similarity metric outperformed MI by a significant margin.

Compared to CT-MR and T1-T2 image registration, MR-US image registration is more challenging due to the fundamental imaging acquisition differences between MR and US. A deep learning-based similarity measure is desired for MR-US image registration. Haskins *et al.* proposed to use CNN to predict the target registration error (TRE) between 3D MR and transrectal US (TRUS) images [59]. The predicted TRE was used as image similarity metric for MR-US rigid registration. TREs obtained from expert-aligned images were used as ground truth. The CNN was trained to regress to the TRE as similarity prediction. The learnt metric was non-smooth and non-convex, which hinders gradient-based optimization. To address this issue, they performed multiple TRE predictions throughout the optimization. The average TRE estimate was used as the similarity metric to mitigate the non-convex problem and to expand the capture range. They claimed that the learnt similarity metric outperformed MI and its variant MIND [63].

Table 2 Overview of RL in medical image registration

References	ROI	Dimension	Modality	Transformation
[50, 51]	Cardiac, HN	2D, 3D	MR, CT, US	NA
[78]	Prostate	3D-3D	MR	Deformable
[92]	Spine, Cardiac	3D-3D	CT-CBCT	Rigid
[101]	Chest, Abdomen	2D-2D	CT-Depth Image	Rigid
[108]	Spine	3D-2D	3DCT-Xray	Rigid
[183]	Spine	3D-2D	3DCT-Xray	Rigid
[144]	Nasopharyngeal	2D-2D	MR-CT	Rigid with scaling

In previous works, accurate image alignment is needed for deep similarity metrics learning. However, it is very difficult to obtain well aligned multi-modal image pairs for network training. The quality of image alignment could affect the accuracy of the learnt deep similarity metrics. To mitigate this problem, Sedghi *et al.* used special data augmentation techniques called dithering and symmetrizing to discharge the need for well-aligned images for deep metric learning [125]. The learnt deep metric outperformed MI on 2D brain image registration. Though they managed to relax the absolute accuracy of image alignment in network training, roughly-aligned image pairs were still necessary. To eliminate the need for aligned image pairs, Wu *et al.* proposed to use stacked autoencoders (SAE) to learn intrinsic feature representations by unsupervised learning [164]. The convolutional SAE could encode an image to obtain low-dimensional feature representations for image similarity calculation. The learnt feature representations were used in Demons and HAMMER to perform brain image DIR. They showed that the image registration performance has improved consistently using the learnt feature representations in terms of dice similarity coefficient (DSC). To test the generality of the learnt feature representation, they reused network trained using LONI dataset on ADNI datasets. The results were comparable to the case of learning feature representation from the same datasets.

It was shown that combining multi-metric measures could produce more robust registration results compared to using the metrics individually. Ferrante *et al.* used support vector machine (SVM) to learn weights of an aggregation of similarity measures including anatomical segmentation maps and displacement vector labels [42]. They have showed that the multi-metric outperformed conventional single-metric approaches. To deal with the non-convex of the aggregated similarity metric, they optimized a regularized upper bound of the loss using CCCP algorithm [180]. One limitation of this method was that segmentation masks of the source images were needed at testing stage.

3.1.2 Assessments

Deep similarity metric has shown its potential to outperform traditional similarity metrics in medical image registration. However, it is difficult to ensure that its derivative is smooth for optimization. The above-mentioned measures of using a large overlap [135] or performing multiple TRE predictions [59] are computationally demanding and only mitigate the problem of non-convex derivatives. Well-aligned image pairs are difficult to obtain for deep similarity network training. Though Wu *et al.* [164] has demonstrated that deep similarity network could be trained in an unsupervised manner, they only tested on unimodal image registration. Extra experiments on multi-modal images need to be performed to show its effectiveness. The biggest limitation of this category maybe that the registration process still inherits the iterative nature of traditional DIR frameworks, which slows the registration process. As more and more papers on direct transformation prediction emerge, it is expected that this category will be less attractive in the future.

3.2 Reinforcement learning in medical image registration

One disadvantage of the previous category is that the registration process is iterative and time-consuming. It is desired to develop a method to predict transformation in one shot. However, one shot transformation prediction is very difficult due to the high dimensionality of the output parameter space. RL has recently gained a lot of attention since the publications from Mnih *et al.* [110] and Silver *et al.* [134]. They combined RL with DNN to achieve human-level performances on Atari and Go. Inspired by the success of RL, and to circumvent the challenge of high dimensionality in one shot transformation prediction, several groups proposed to combine CNN with RL to decompose the registration task into a sequence of classification problems. The strategy is to find a series of actions, such as rotation and translation along certain axis by a certain value, to iteratively improve image alignment.

3.2.1 Overview of works

Table 2 shows a list of selected references that used RL in medical image registration. Liao *et al.* was one of the first to explore RL in medical image registration [92]. The task was to perform 3D-3D, rigid, cone beam CT (CBCT)-CT image registration. Specific challenges of the registration include large differences in field of views (FOVs) between the CT and CBCT in spine registration and the severe streaking artifacts in CBCT. An artificial agent was trained using a greedy supervised approach to perform rigid image registration. The artificial agent was modelled using CNN, which took raw images as input and output the next optimal action. The action space consists of 12 candidate transformations, which are 1mm of translation and 1 degree of rotation along the x, y, and z axis, respectively. Ground truth alignment were obtained using iterative closest point registration of expert-defined spine landmarks and epicardium segmentation, followed by visual inspection and manual editing. Data augmentation was used to artificially de-align the image pair with known transformations. Different from Mnih *et al.* who trained their network with repeated trial and error, Liao *et al.* trained the network with greedy supervision, where the reward can be calculated explicitly via a recursive function. They showed that the network training process with supervision was a magnitude more efficient than the training process of Mnih *et al.*s network. They also claimed their network could reliably overcome local maxima, which was challenging for generic optimization algorithms when the underlying problem was non-convex. Motivated by [92], Miao *et al.* proposed a multi-agent system with an auto attention mechanism to rigidly register 3D-CT with 2D X-ray spine image [108]. Reliable 2D-3D image registration could map the pre-operative 3D data to real-time 2D X-ray images by image fusion. To deal with various image artifacts, they proposed to use an auto-attention mechanism to detect regions with reliable visual cues to drive the registration. In addition, they used a dilated FCN-based training mechanism to reduce the degree of freedom of training data to improve the training efficiency. They have outperformed single agent-based and optimization-based methods in terms of TRE. Sun *et al.* proposed to use an asynchronous RL algorithm with customized reward function for 2D MR-CT image registration [144]. They used datasets from 99 patients diagnosed as nasopharyngeal carcinoma. Ground truth image alignments were obtained using toolbox Elastix [75]. Different from previous works, Sun *et al.* incorporated scaling factor into the action space. The action space consists of 8 candidate transformations including 1 pixel for translation, 1 degree for rotation and 0.05 for scaling. CNN was used to encode image states and LSTM was used to encode hidden states between neighboring frames. Their method was better than Elastix in terms of TRE when the initial image alignment was poor. The use of actor-critic scheme [99] allowed the agent to explore transformation parameter spaces freely and avoided local minima when the initial alignment was poor. On the contrary, when the initial image alignment was good, Elastix was slightly better than their method. In the inference phase, a Monte Carlo rollout strategy was proposed to terminate the searching path to reach a better action. All of the above-mentioned methods focused on rigid registration since rigid transformation could be represented by a low-dimensional parametric space, such as rotation, translation and scaling. However, non-rigid, free-form transformation model has high dimensionality and non-linearity which would result in a huge action space. To deal with this problem, Krebs *et al.* proposed to build a statistical deformation model (SDM) with a low-dimensional parametric space [78]. Principal component analysis (PCA) was used to construct SDM on B-spline deformation vector field (DVF). Modes of the PCA of the displacement were used as the unknow vectors for the agents to optimize. They evaluated the method on inter-subject MR prostate image registration in both 2D and 3D. The method achieved DSC scores of 0.87 and 0.80 for 2D and 3D, respectively. Ghesu *et al.* proposed to use RL to detect 3D-landmarks in medical images [51]. This method was mentioned since it belongs to the category of RL and the detected landmarks could be used for landmark-based image registration. They reformulated the landmark detection task as a behavioral problem for the network to learn. To deal with local minima problem, a multi-scale approach was used. Experiments on 3D-CT scans were conducted to compare with another five methods. The results showed that the detection accuracy was improved by 20-30 percent while being 2-3 orders of magnitude faster.

3.2.2 Assessment

The biggest limitation of RL-based image registration is that the transformation model is highly constrained to low-dimensionality. As a result, most of the RL-based registration methods used rigid transformation models. Though Krebs *et al.* has applied RL to non-rigid image registration by predicting a low-dimensional parametric space of statistical deformation model, the accuracy and flexibility of the deformation model is highly constrained and may not be adequate to represent the actual deformation. RL-based image registration methods have shown its usefulness in enhancing the robustness of many algorithms in multi-modal image registration tasks. Despite the usefulness of RL, statistics

Table 3 Overview of supervised transformation prediction methods

References	ROI	Dimension	Patch-based	Modality	Transformation
[107]	Implant, TEE	2D-3D	Yes	Xray	Rigid
[114]	Cranial	2D-3D	No	CBCT-Xray	Deformable
[119]	Cardiac	3D-3D	No	MR	Deformable
[152]	Cardiac, Brain	2D-2D	No	MR	Deformable
[175]	Brain	3D-3D	Yes	MR	Deformable
[10]	Brain	3D-3D	Yes	MR	Deformable
[11]	Pelvic	3D-3D	Yes	MR-CT	Deformable
[64]	Cardiac	2D-2D	No	MR	Deformable
[66, 67]	Prostate	3D-3D	No	MR-US	Deformable
[100]	Abdomen	2D-2D	Yes	MR	Deformable
[122]	Brain	3D-3D/2D	No	MR	Rigid
[126]	Lung	3D-3D	Yes	CT	Deformable
[136]	Brain	2D-2D	No	T1-T2	Rigid
[145]	Liver	2D-2D	Yes	CT-US	Affine
[166]	Prostate	3D-3D	No	MR-US	Rigid + Affine
[35, 36]	Lung	3D-3D	No	CT	Deformable
[39]	Brain	3D-3D	Yes	MR	Deformable
[43]	Lung	3D-2D	No	CT	Deformable
[76]	Brain	3D-3D	No	T1, T2, Flair	Affine
[94]	Skull, Upper Body	2D-2D	No	DRR-Xray	Deformable
[113, 139, 140]	Lung	3D-3D	Yes	CT	Deformable

indicates loss of popularity of this category, evidenced by the decreasing number of papers in 2019. As the techniques advance, more and more direct transformation predication methods are proposed. The accuracy of the direct transformation prediction methods is constantly improving, achieving comparable accuracy to top traditional DIR methods. Therefore, the advantage of casting registration as a sequence of classification problems in RL-based registration methods is gradually vanishing.

3.3 Supervised transformation predication

Both deep similarity-based and RL-based registration methods are iterative methods in order to avoid the challenges of one-shot transformation prediction. Despite the difficulties, several groups have attempted to train networks to directly infer the final transformation in a single forward prediction. The challenges include 1) high dimensionality of the output parametric space, 2) lack of training datasets with ground truth transformations and 3) regularization of the predicted transformation. Methods including ground truth transformation generation, image re-sampling and transformation regularization methods have been proposed to overcome these challenges. Table 3 shows a list of selected references that used supervised transformation prediction for medical image registration.

3.3.1 Overview of works
3.3.1.1 Ground truth transformation generation
For supervised transformation prediction, it is important to generate many image pairs with known transformations for network training. Numerous data augmentation techniques were proposed for artificial transformations generation. Generally, these artificial transformation generation methods can be classified into three groups: 1) random transformation, 2) traditional registration-generated transformation and 3) model-based transformations.
A. Random transformation generation
Salehi *et al.* aimed to speed up and improve the capture range of 3D-3D and 2D-3D rigid image registration of fetal brain MR scans [122]. CNN was used to predict both rotation and translation parameters. The network was trained using datasets generated by randomly rotating and translating the original 3D images. Both MSE and geodesic distance were used for loss function calculation. Geodesic distance is the distance between two points on a unit sphere. They have showed significant improvement after combining the geodesic distance loss with the MSE loss. Sun *et al.* used expert aligned CT-US image pairs as ground truth [145]. Known artificial affine transformations were used to synthesize training datasets. The network was trained to predict the affine parameters. They have trained

network which worked for simulated CT-US registration. However, it does not work on real CT-US pairs due to the vast appearance differences between the simulated and the real US. They have tried multiple methods to counter-act overfitting, such as deleting dropout layers, less complex network, parameter regularization and weight decay. Unfortunately, none of them worked. Eppenhof *et al.* proposed to train a CNN using synthetic random transformations to perform 3D-CT lung DIR [36]. The output of the network was DVF on a thin plate spline transform grid. MSE between the predicted DVF and the ground truth DVF was used as loss function. They achieved 4.023.08 mm TRE on DIRLAB [12], which was much worse than 1.36+1.01 mm of traditional DIR method. They later improved their method to use a U-Net architecture [35]. The network was trained on whole image. Images were down-sampled to fit into GPU memory. Again, synthetic random transformation was used to train the network. Affine pre-registration was required prior to CNN transformation prediction. They managed to reduce the TRE from 4.023.08 mm to 2.171.89 mm on DIRLAB datasets. Despite the slightly worse TRE than traditional DIR methods, they have demonstrated the possibility of direct transformation prediction using CNN.

Eppenhof *et al.* proposed to train a CNN using synthetic random transformations to perform 3D-CT lung DIR [117]. The output of the network was DVF on a thin plate spline transform grid. MSE between the predicted DVF and the ground truth DVF was used as loss function. They achieved 4.02±3.08 mm TRE on DIRLAB [126], which was much worse than 1.36+1.01 mm of traditional DIR method. They later improved their method to use a U-Net architecture [118]. The network was trained on whole image. Images were down-sampled to fit into GPU memory. Again, synthetic random transformation was used to train the network. Affine pre-registration was required prior to CNN transformation prediction. They managed to reduce the TRE from 4.02±3.08 mm to 2.17±1.89 mm on DIRLAB datasets. Despite the slightly worse TRE than traditional DIR methods, they have demonstrated the possibility of direct transformation prediction using CNN.

B. Traditional registration-generated transformations

Later, several groups tried to use traditional registration methods to register an image pair to generate ground truth transformations for the network to learn. The rationale is that random transformation generation might be too different from the true transformation, which might deteriorate the performance of network. Sentker *et al.* used DVF generated from traditional DIRs including Plasti-Match [111], NiftyReg [127] and VarReg [162] as ground truth [126]. MSE between the predicted and the ground truth DVF was used as loss function to train a network for 3D-CT lung registration. On DIRLAB [12] datasets, they achieved better TRE using DVFs generated by VarReg as compared to Plasti-Match and NiftyReg. Results showed that their CNN-based registration method was comparable to the original traditional DIR in terms of TRE. The best TRE values they have achieved on DIRLAB is 2.501.16 mm. Fan *et al.* proposed a BIRNet to perform brain image registration using dual supervision [39]. Ground truth transformations were obtained using existing registration methods. MSE between the ground truth and the predicted transformations were used as loss function. They used not only the original image but also its difference and gradient images as input to the network.

C. Model-based transformation generation

Uzunova *et al.* aimed to generate a large and diverse set of training image pairs with known transformations from a few sample images [152]. They proposed to learn highly expressive statistical appearance models (SAM) from a few training samples. Assuming Gaussian distribution for the appearance parameters, they synthesized huge amounts of realistic ground truth training datasets. FlowNet [29] architecture was used to register 2D MR cardiac images. For comparison, they have generated ground truth transformations using three different methods, which are affine registration-generated, randomly-generated and the proposed SAM-generated transformations. They showed that CNN learnt from the SAM-generated transformation outperformed CNN learnt from randomly-generated and affine registration-generated transformation. Sokooti *et al.* generated artificial DVFs using model-based respiratory motion to simulate ground truth DVF for 3D-CT lung image registration [140]. For comparison, random transformations were also generated using single frequency and mixed frequencies. They tested different combinations of various network structures including U-Net whole image, multi-view based and U-Net advanced. The multi-view and U-Net advanced all used patch-based training. TRE and Jacobian determinant were used as evaluation metrics. After comparison, they claimed that the realistic model-based transformation performed better compared to random transformations in terms of TRE. On average, they achieved TRE of 2.32 mm and 1.86 mm for SPREAD and DIRLAB datasets, respectively.

3.3.1.2 Supervision methods

As neural network develops, many new supervision terms such as 'supervised', 'unsupervised', 'deeply supervised', 'weakly supervised', 'dual supervised', 'self-supervised' have emerged. Generally, neural network learns to perform a certain task by minimizing a predefined loss function via optimization. These terms refer to how the training datasets are prepared and how the networks are trained using the datasets. In the following paragraph, we briefly describe the definition of each supervision strategy in the context of DL-based image registration.

The learning process of a neural network is supervised if the desired output is already known in the training datasets. Supervised network means the network is trained with the ground truth transformation, which is a dense DVF for free deformation and a parametric vector of 6 for rigid transformation. On the other hand, unsupervised learning has no target output available in the training datasets, which means the desired DVFs or target transformation parameters are absent in the training datasets. Unsupervised network was also referred to self-supervised network since the warped image is generated from one of the input image pair and compared to another input image for supervision. Deep supervision usually means that the differences between outputs from multiple layers and the desired outputs are penalized during training whereas normal supervision only penalizes the difference between the final output and the desired output. In this manner, supervision was extended to deep layers of the network. Weak supervision represents scenario where ground truth other than the exact desired output is available in the training datasets and used to calculate the loss function. For example, a network is called weakly supervised if corresponding anatomical structural masks or landmark pairs, not the desired dense DVF, are used to train the network for direct dense DVF prediction. Dual supervision means that the network is trained using both supervised and unsupervised loss functions.

A. Weak supervision

Methods that use ground truth transformation generation were mainly supervised method for direct transformation prediction. Weakly supervised transformation prediction has also been explored. Instead of using artificially-generated transformations, Hu *et al.* proposed to use higher-level correspondence information such as labels of anatomical organs for network training [66]. They argued that such anatomical labels were more reliable and practical to obtain. They trained a CNN to perform deformable MR-US prostate image registration. The network was trained using weakly supervised method, meaning that only corresponding anatomical labels, not dense voxel-level spatial correspondence, were used for loss calculation. The anatomical labels were required only in the training stage for loss calculation. Labels were not required in inference stage to facilitate fast registration. Similarly, Hering *et al.* combined the complementary information from segmentation labels and image similarity to train a network [64]. They showed significant higher DSC scores than using only image similarity loss or segmentation label loss in 2D MR cardiac DIR.

B. Dual supervision

Technically, dual supervision is not strictly defined. It usually means the network was trained using two types of important loss functions. Cao *et al.* used dual supervision which includes a MR-MR loss and a CT-CT loss [11]. Prior to network training, they transformed the multi-modality to unimodality registration by using pre-aligned counterpart images, for MR-CT registration. The MR has a pre-aligned CT and CT has a pre-aligned MR. The loss function has a dual similarity loss including MR-MR and CT-CT loss. They showed that the dual-modality similarity performed better than SyN [2] and single modality similarity in terms of DSC and average surface distance (ASD) in pelvic image registration. Liu *et al.* used representation learning to learn feature-based descriptors with probability maps of confidence level [94]. Then, the learnt descriptor pairs across the image were used to build a geometric constraint using Hough voting or RANSAC. The network was trained using both supervised synthetic transformations and an unsupervised descriptor image similarity loss. Similarly, Fan *et al.* combined both supervised and unsupervised loss terms for dual supervision in MRI brains image registration [39].

3.3.2 Assessment

In recent two to three years, we have seen a huge interest in supervised CNN direct transformation prediction, evidenced by increasing number of publications. Though direct transformation prediction has yet to outperform the-state-of-art traditional DIR methods, the registration accuracy has improved greatly. Some methods have achieved comparable registration accuracy to the traditional DIR methods. Ground truth transformation generation will continue to play an important role in network training. Limitations of using artificially generated image pair with known ground truth transformations include 1) the generated transformation might not reflect the true physiological motion, 2) the generated transformation might not capture the large range of variations of actual image registration scenarios and 3)

Table 4 Overview of unsupervised transformation prediction methods

References	ROI	Dimension	Patch-based	Modality	Transformation
[52]	Brain	3D-3D	No	MR	Deformable
[129]	Brain, Liver	2D-2D	No	MR, CT	Deformable
[155]	Cardiac	2D-2D	No	MR	Deformable
[177]	Neural tissue	2D-2D	No	EM	Deformable
[14]	Brain	3D-3D	No	MR	Affine
[37, 38]	Brain	3D-3D	Yes	MR	Deformable
[41]	Lung, Cardiac	2D-2D	No	MR, Xray	Deformable
[72]	HN	3D-3D	Yes	CT	Deformable
[77]	Cardiac	3D-3D	No	MR	Deformable
[90]	Brain	3D-3D	No	MR	Deformable
[102, 116]	Cardiac	2D-2D	No	MR	Deformable
[130]	Cardiac	2D-2D	No	MR	Deformable
[133]	Neuron Tissue	2D-2D	Yes	EM	Affine
[142]	Lung	3D-3D	No	MR	Deformable
[143]	Brain	3D-3D	No	MR-US	Deformable
[23]	Cardiac, Lung	3D-3D	Yes	MR, CT	Affine and Deformable
[181]	Brain	3D-3D	No	MR	Deformable
[5, 6, 21]	Brain	3D-3D	No	MR	Deformable
[32, 33]	Prostate	3D-3D	Yes	CT	Deformable
[38]	Brain, Pelvic	3D-3D	Yes	MR, CT	Deformable
[71]	Lung	2D-2D	Yes	CT	Deformable
[74]	Liver	3D-3D	No	CT	Deformable
[80, 81]	Brain	3D-3D	No	MR	Deformable
[82]	Liver	3D-3D	No	CT	Deformable
[87]	Abdomen	3D-3D	Yes	CT	Deformable
[104]	Retina	2D-2D	No	FA	Deformable
[149]	Neural tissue	2D-2D	No	EM	Deformable
[179]	Abdominopelvic	3D-3D	Yes	CT-PET	Deformable
[40]	Lung, Cardiac	3D-3D	Yes	CT, MR	Deformable

the artificially generated image pairs in the training stage are different from the actual image pair in the inference stage. To deal with the first limitations, we can use various transformation generation models. Adequate data augmentation could be performed to mitigate the second limitation. Domain adaption [41, 183] could be used to account for the domain difference between the artificially-generated and the true images. Image registration is an ill-posed problem, the ground truth transformation could help to constrain the final transformation prediction. Combinations of different loss functions and DVF regularization methods have also been examined to improve the accuracy of registration. We expect DL-based registration of this category to keep growing in the future.

3.4 Unsupervised transformation prediction

It is desired to develop unsupervised image registration methods to overcome the lack of training datasets with known transformations. However, it is difficult to define proper loss function of the network without ground truth transformations. In 2015, Jaderberg *et al.* proposed a spatial transformer network (STN) which explicitly allows spatial manipulation of data within the network [69]. Importantly, the spatial transformer network was a differentiable module that can be inserted in to existing CNN architectures. The publication of STN has inspired many unsupervised image registration methods since STN enables image similarity loss calculation during the training process. A typical unsupervised transformation prediction network for DIR takes an image pair as input and directly output dense DVF, which was used by STN to warp the moving image to generate warped images. The warped images were then compared to fixed images to calculate image similarity loss. DVF smoothness constraint was normally used to regularize the predicted DVF.

3.4.1 Overview of works

Yoo *et al.* proposed to use a convolution autoencoder (CAE) to encode image to a vector to calculate similarity, called feature-based similarity which is different from handcrafted feature similarity such

as SIFT [177]. They showed this feature-based similarity measure was better than intensity-based similarity measure for DIR. They have combined the deep similarity metrics and STN for unsupervised transformation estimation in 2D electron microscopy (EM) neural tissue image registration. Balakrishnan *et al.* proposed an unsupervised CNN-based DIR method for MR brain atlas-based registration [5, 6]. They used a U-Net like architecture and named it VoxelMorph. In the training, the network penalized the differences in image appearances with the help of STN. Smoothness constraint was used to penalize local spatial variations in the predicted transformation. They have achieved comparable performance to ANT [3] registration method in terms of DSC score of multiple anatomical structures. Later, they extended their method to leverage auxiliary segmentations available in the training data. A DSC loss function was added to the original loss functions in the training stage. Segmentation labels were not required during testing. They investigated unsupervised brain registration, with and without segmentation label DSC loss. Their results showed that the segmentation loss could help yield improved DSC scores. The performance is comparable to ANT and NiftyReg, while being x150 faster than ANTs and x40 faster than NiftyReg. Like [6], Qin *et al.* also used segmentation as complementary information for cardiac MR image registration [116]. They found out that the feature learnt by registration CNN could be used in segmentation as well. The predicted DVF was used to deform the masks of moving image to generate masks of the fixed image. They trained a joint segmentation and registration model for cardiac cine image registration and proved that the joint mode could generate better results than the separate models alone in both segmentation and registration tasks. Similar idea has been explored in [103] as well. They claimed registration and segmentation are complementary functions and combining them can improve each others performance. Later, Zhang *et al.* proposed a network with trans-convolutional layers for end-to-end DVF prediction in MR brain DIR [181]. They focused on the diffeomorphic mapping of the transformation. To encourage smoothness and avoid folding of the predicted transformation, they proposed an inverse-consistent regularization term to penalize the difference between two transformations from the respective inverse mappings. The loss function consists of an image similarity loss, a transformation smoothness loss, an inverse consistent loss and an anti-folding loss. Their method has outperformed Demons and Syn, in terms of DSC score, sensitivity, positive predictive value, average surface distance and Hausdorff distance. A similar idea was proposed by Kim *et al.* who used cycle consistent loss to enforce DVF regularization [74]. They also used identity loss where the output DVF should be zero if the moving and fixed image are the same image. For 3D-CT image registration, Lei *et al.* used an unsupervised CNN to perform abdominal image registration [87]. They used a dilated inception module to extract multi-scale motion features for robust DVF prediction. Apart from the image similarity loss and DVF smoothness loss, they integrated a discriminator to provide additional adversarial loss for DVF regularization. Vos *et al.* proposed an unsupervised affine and DIR framework by stacking multiple CNN into a larger network [23]. The network was tested on cardiac cine MRI and 3D CT lung image registration. They showed their method was comparable to conventional DIR method while being several orders of magnitude faster. Like [23], Lau *et al.* cascaded affine and deformable networks for CT liver DIR [82]. Recently, Jiang *et al.* proposed a multi-scale framework with unsupervised CNN for 3D CT lung DIR [71]. They cascaded three CNN models with each model focusing on its own scale level. The network was trained using image patches to optimize an image similarity loss and a DVF smoothness loss. They showed that network trained on SPARE datasets could generalize to a different DIRLAB datasets. In addition, the same trained network also performed well on CT-CBCT and CBCT-CBCT registration without retraining or fine-tuning. They achieved an average TRE of 1.661.44 mm on DIRLAB datasets. Fu *et al.* proposed an unsupervised method for 3D-CT lung DIR [48]. They first performed whole-image registration on down-sampled image using a CoarseNet to warp the moving image globally. Then, image patches of the globally warped moving image were registered to the image patches of the fixed image using a patch-based FineNet. They also incorporated a discriminator to provide adversarial loss by penalizing unrealistic warped images. Vessel enhancement was performed prior to DIR to improve the registration accuracy. They have achieved an average TRE of 1.591.58 mm, which outperformed some traditional DIR methods. Interestingly, both Jiang *et al.* and Fu *et al.* have achieved better TRE values using unsupervised methods than the supervised methods in [35] and [126].

3.4.2 Assessment

Compared to supervised transformation prediction, unsupervised methods effectively alleviate the problem of lack of training datasets. Various regularization terms have been proposed to encourage plausible transformation prediction. Several groups have achieved comparable or even better results in terms of TRE on DIRLAB 3D-CT lung DIR. However, most of the methods in this category focused

Table 5 Overview of registration methods using GAN

References	ROI	Dimension	Patch-based	Modality	Transformation
[37]	Brain	3D-3D	Yes	MR	Deformable
[67]	Prostate	3D-3D	No	MR-US	Deformable
[166]	Prostate	3D-3D	No	MR-US	Deformable
[122]	Brain	3D-3D	No	MR	Rigid
[33]	Prostate	3D-3D	Yes	CT	Deformable
[38]	Brain, Pelvic	3D-3D	Both	MR, CT	Deformable
[87]	Abdomen	3D-3D	Yes	CT	Deformable
[48]	Lung	3D-3D	Yes	CT	Deformable
[102, 103, 104]	Retina, Cardiac	2D-2D	No	FA, Xray	Deformable
[115]	Lung, Brain	2D-2D	No	T1-T2, CT-MR	Deformable

on unimodality registration. There has been a lack of investigation in multi-modality image registration using unsupervised methods. To provide additional supervision, several groups have combined supervised with unsupervised methods for transformation prediction [39]. The combination seems beneficial; however, more investigation was needed to justify its effectiveness. Given the promising results of the unsupervised methods, we expect a continuous growth of interest in this category.

3.5 GAN in medical image registration

The use of GAN in medical image registration can be generally categorized in two groups: 1) to provide additional regularization of the predicted transformation; 2) to perform cross-domain image mapping.

3.5.1 Overview of works

3.5.1.1 GAN-based regularization

Since image registration is an ill-posed problem, it is crucial to have adequate regularization to encourage plausible transformations and to prevent unrealistic transformations such as tissue folding. Commonly used regularization terms include DVF smoothness constraint, anti-folding constraint and inverse consistency constraint. However, it remains ambiguous whether these constraints are adequate for proper regularization. Recently, GAN-based regularization terms have been introduced to the realm of image registration. The idea is to train an adversarial network to introduce a network-based loss for transformation regularization. In the literature, discriminators were trained to distinguish three types of inputs, including 1) whether a transformation is predicted or ground truth, 2) whether an image is realistic or warped by predicted transformation, 3) whether an image pair alignment is positive or negative. Yan *et al.* trained an adversarial network to tell whether an image was deformed using ground truth transformation or predicted transformation [166]. Randomly generated transformations from manually aligned image pairs were used as ground truth to train a network to perform MR-US prostate image registration. The trained discriminator could provide not only an adversarial loss for regularization but also a discriminator score for alignment evaluation. Fan *et al.* used a discriminator to distinguish whether an image pair were well aligned [38]. In unimodal image registration, they have defined a positive image alignment case as weighted linear combination of the fixed and the moving images. In multi-modal image registration case, positive image alignments were pre-defined using paired MR and CT images. They performed on MR brain images for unimodal registration and on pelvic CT-MR for multi-modal registration. They have showed that the performance increased with the adversarial loss. Lei *et al.* used a discriminator to judge whether the warped image is realistic enough to the original images [87]. Fu *et al.* used a similar idea and showed that the inclusion of adversarial loss could improve registration accuracy in 3D-CT lung DIR [48]. The above GAN-based methods have tried to introduce regularization from the image or transformation appearance perspective. Differently, Hu *et al.* tried to introduce biomechanical constraints to 3D MR-US prostate image registration by discriminating whether a transformation is predicted or generated by finite element analysis [67]. Instead of adding the adversarial loss to existing smoothness loss, they replaced the smoothness loss with the adversarial loss. They showed that their method could predict physically plausible deformation without any other smoothness penalty.

3.5.1.2 GAN-based cross-domain image mapping

Table 6 Overview of registration validation methods using deep learning

References	ROI	Dimension	Modality	End point
[112]	HN	3D	CT	TRE prediction
[34]	Lung	3D	CT	Registration error
[31]	Brain	3D	MRI	DSC score
[47]	Lung	3D	CT	Landmark Pairs
[49]	Lung	3D	CT	Registration error
[138]	Lung	3D	CT	Registration error

For multi-modal image registration, progresses have been made by using deep similarity metrics in traditional DIR frameworks. Using iterative methods, several works have outperformed the-state-of-art MI similarity measures. However, in terms of direct transformation prediction, multi-modal image registration has not benefited from DL as much as unimodal image registration has. This is mainly due to the vast appearance differences between different modalities. To overcome this challenge, GAN has been used to translate multi-modal to unimodal image registration by mapping images from one modality to another. Salehi *et al.* trained a CNN using T2-weighted images to perform fetal brain MR registration. They tested the network on T1-weighted images by first mapping the T1 to T2 image domain using a conditional GAN [122]. They showed the trained network generalized well on the synthesized T2 images. Qin *et al.* used an unsupervised image-to-image translation framework to cast multi-modal to unimodal image registration [115]. The image to image translation method assumes the images could be decomposed into content code and style code. They have showed comparable results to MIND and Elastix on BraTs datasets in terms of RMSE of DVF error. On COPDGene datasets, they outperformed MIND and Elastix in terms of DICE, mean contour distance (MCD) and Hausdorff distance. Mahapatra *et al.* combined cGan [109] and registration network together to directly predict both DVF and warped image [104]. They implicitly transformed image in one modality to another modality. They outperformed Elastix on 2D retinal image registration in terms of Hausdorff distance, MAD and MSE. Elmahdy *et al.* claimed that inpainting gas pockets in the rectum could enhance rectum and seminal vesicle registration [33]. They used GAN to detect and inpaint rectum gas pocket prior to image registration.

3.5.2 Assessment

GAN has been shown to be promising in medical image registration via either novel adversarial loss or image domain translation. For adversarial losses, GAN could provide learnt network-based regularizations that are complementary to traditional handcrafted regularization terms. For image domain translation, GAN effectively cast the more challenging multi-modal registration to unimodal image registration, which allows many existing unimodal registration algorithms to be applied to multi-modal image registration. However, the absolute intensity mapping accuracy of GAN is yet to be investigated. GAN has also been applied to deep similarity metric learning in registration and alignment validation. As evidenced by the trend in Fig. 2, we expect to see more papers using GAN in image registration tasks in the future.

3.6 Registration validation using deep learning

The performance of image registration could be evaluated using image similarity metrics such as SSD, NCC and MI. However, the image similarity metrics only evaluate the overall alignment on the whole image. To have a deeper insight into local registration accuracy, we usually rely on manual landmark pair selection. Nevertheless, manual landmark pair selection is time-consuming, subjective and error-prone especially when many landmarks were to be selected. Fu *et al.* used a Siamese network for large quantity landmark pair detection on 3D-CT lung images [47]. The network was trained using the manual landmark pairs from DIRLAB datasets. They performed experiments comparisons, showing that the network could outperform human in landmark pair detection. Neylon *et al.* proposed to use a deep neural network to predict TRE for given image similarity metrics [112]. The network was trained using patient-specific biomechanical models of head-neck anatomy. They demonstrated that the network could rapidly and accurately quantify registration performance.

3.6.1 Overview of works

Table 7 Overview of other deep learning-based image registration methods

References	ROI	Dimension	Modality	Transformation	Methods
[70]	Lung	3D-3D	CT	Deformable	Multi-grid Inference
[163]	Brain	3D-3D	MR-US	Rigid	LSTM
[8]	Brain	2D-2D	CT, T1, T2, PD	Rigid	Manifold Learning
[165]	Brain, Abdomen	2D-3D	CT-PET, CT-MRI	Deformable	CAE, DSCNN
[178]	Spine	3D-2D	3DCT-Xray	Rigid	FasterRCNN
[54]	Brain	2D-2D	T1-T2, T1-PD	Deformable	FCN
[183]	Spine	3D-2D	3DCT-Xray	Rigid	Domain adaptation

Eppenhof *et al.* proposed a TRE alternative to assess DIR registration accuracy. They used synthetic transformations as ground truth to avoid the need for manual annotations [34]. The ground truth error map was the L2 difference between ground truth transformations and the predicted transformations. They trained a network to robustly estimate registration errors with sub-voxel accuracy. Galib *et al.* predicted an overall registration error index, which is the ratio between good alignment sub-volumes and poor alignment sub-volumes [49]. They justified the choice of threshold TRE of 3.5mm as a cutoff value of good and bad alignment. Their network was trained using manually labeled landmarks from DIRLAB. Sokooti *et al.* proposed a random forest regression method for quantitative error prediction of DIR [138]. They used both intensity-based features such as MIND and registration-based features such as transformation Jacobian determinant. Dubost *et al.* used ventricle DSC score to evaluate brain registration [31]. The ventricle was segmented using deep learning-based method.

3.6.2 Assessment

The number of papers using deep learning for registration evaluation has increased significantly in 2019. Most works treated registration error prediction as a supervised regression problem. Network was trained using manually annotated datasets. It is important to make sure the ground truth datasets are of high quality. Most of existing methods focused on lung because benchmark datasets with manual landmark pairs exists for 3D CT lung such as DIRLAB. It would be interesting to see the method be applied on many other treatment sites. Unsupervised registration error prediction is another interesting research topic to eliminate the need for manual annotated datasets.

3.7   Other learning-based methods in medical image registration

Jiang *et al.* proposed to use CNN to lean and infer expressive sparse multi-grid configurations prior to B-spline coefficient optimization [70]. Liu *et al.* used a ten-layer FCN for image synthesis without GAN to transform multimodal to unimodal registration among T1-weighted, T2-weighed, and proton density images [95]. Then, they used Elastix software with SSD similarity metric for the registration of brain phantom and IXI datasets. They outperformed MI similarity index. Wright *et al.* proposed to use LSTM network to predict a rigid transformation and an isotropic scaling factor for MR-US fetal brain registration [163]. Bashiri *et al.* used Laplacian eigenmap as a manifold learning method to implement a multi-modal to unimodal image translation in 2D brain image registration [8].

Yu *et al.* proposed to use FasterRCNN [117] for vertebrae bounding box detection [178]. The detected bounding box was then matched to doctor-annotated bounding box on the X-ray image. Zheng *et al.* proposed a domain adaptation module to cope with the domain variance between synthetic data and real data [183]. The adaptation module can be trained using a few paired real and synthetic data. The trained module could be plugged into the network to transfer the real features to approach the synthetic features. Since network was trained on synthetic data, the network should perform well on synthetic data. Hence, it is reasonable to transfer the real data features to synthetic features.

4   Benchmark

Benchmarking is important for readers to understand through comparison the advantages and disadvantages of each method. For image registration, both registration accuracy and computational time could be benchmarked. However, researchers have been reporting registration accuracies more than the computational speed. Computational speed is largely dependent on the hardware, which is often

Table 8 Comparison of Target Registration Error (TRE) values among different methods on DIRLAB datasets, TRE unit: (mm), *: Traditional DIR methods

Set	Initial	Heinrich* et al. [62]	Delmon* et al. [24]	Staring* et al. [141]	Eppenhof et al. [35]	De Vos et al. [23]	Sentker et al. [126]	Fu et al. [48]	Sokooti et al. [140]	Jiang et al. [71]	Fechter et al. [40]
1	3.89±2.78	0.97±0.5	1.2±0.6	0.99±0.57	1.45±1.06	1.27±1.16	1.20±0.60	0.98±0.54	1.13±0.51	1.20±0.63	1.21±0.88
2	4.3±3.9	0.96±0.5	1.1±0.6	0.94±0.53	1.46±0.76	1.20±1.12	1.19±0.63	0.98±0.52	1.08±0.55	1.13±0.56	1.13±0.65
3	6.94±4.05	1.21±0.7	1.6±0.9	1.13±0.64	1.57±1.10	1.48±1.26	1.67±0.90	1.14±0.64	1.33±0.73	1.30±0.70	1.32±0.82
4	9.83±4.86	1.39±1.0	1.6±1.1	1.49±1.01	1.95±1.32	2.09±1.93	2.53±2.01	1.39±0.99	1.57±0.99	1.55±0.96	1.84±1.76
5	7.48±5.51	1.72±1.6	2.0±1.6	1.77±1.53	2.07±1.59	1.95±2.10	2.06±1.56	1.43±1.31	1.62±1.30	1.72±1.28	1.80±1.60
6	10.89±6.9	1.49±1.0	1.7±1.0	1.29±0.85	3.04±2.73	5.16±7.09	2.90±1.70	2.26±2.93	2.75±2.91	2.02±1.70	2.30±3.78
7	11.03±7.4	1.58±1.2	1.9±1.2	1.26±1.09	3.41±2.75	3.05±3.01	3.60±2.99	1.42±1.16	2.34±2.32	1.70±1.03	1.91±1.65
8	15.0±9.01	2.11±2.4	2.2±2.3	1.87±2.57	2.80±2.46	6.48±5.37	5.29±5.52	3.13±3.77	3.29±4.32	2.64±2.78	3.47±5.00
9	7.92±3.98	1.36±0.7	1.6±0.9	1.33±0.98	2.18±1.24	2.10±1.66	2.38±1.46	1.27±0.94	1.86±1.47	1.51±0.94	1.47±0.85
10	7.3±6.35	1.43±1.6	1.7±1.2	1.14±0.89	1.83±1.36	2.09±2.24	2.13±1.88	1.93±3.06	1.63±1.29	1.79±1.61	1.79±2.24
Mean	**8.46±5.48**	**1.43±1.3**	**1.66±1.14**	**1.32±1.24**	**2.17±1.89**	**2.64±4.32**	**2.50±1.16**	**1.59±1.58**	**1.86±2.12**	**1.66±1.44**	**1.83±2.35**

Table 9 Benchmark datasets and evaluation metrics used in brain registration

References	Datasets	Transformation	Evaluation Metrics
[54]	IXI	Deformable	TRE, MI
[81]	MindBoggle-101	Deformable	DSC
[76]	BRATS, ALBERT	Affine	DSC, MI, SSIM, MSE
[42]	IBSR	Deformable	DSC, MI, NCC, SAD, DWT
[39]	LONI, LPBA40, IBSR, CUMC, MGH, IXI	Deformable	DSC, ASD
[6]	OASIS, ABIDE, ADHD, MCIC, PPMI, HABS, Harvard GSP	Deformable	DSC
[181]	ADNI	Deformable	DSC, SEN, PPV, ASD, HD
[136]	OASIS, IXI, ISLES	Rigid	MSE
[125]	IXI	Rigid	MAE of degree and translation
[90]	ADNI	Deformable	DSC
[10]	LONI, ADNI, IXI	Deformable	DSC, ASD
[175]	OASIS, IBIS, LPBA, IBSR, MGH, CUMC	Deformable	Target Overlap
[129]	LPBA	Deformable	TRE, JACC
[52]	IXI	Deformable	SSD, PSNR, SSIM
[164]	LONI, ADNI	Deformable	DSC
[135]	IXI, ALBERT	Deformable	DSC, JACC

different from group to group. According to the statistics of the cited works, the top two ROIs of registration are brain and lung. Therefore, we summarized the registration datasets for brain, registration accuracies for lung.

4.1 Lung

DIRLAB is one of the most cited public datasets for 4D-CT chest image registration studies [12]. DIRLAB provides 300 manually selected landmark pairs for end-exhalation and end-inhalation phases. This dataset was frequently used for 4D-CT lung registration benchmarking. To provide the readers a better understanding of the latest DL-based registration, we have listed the TREs of three top performing traditional methods and seven DL-based lung registration methods. Table 8 shows that DL-based lung registration methods have yet outperformed the top traditional DIR methods. However, DL-based DIR methods have been making substantial improvement over the years, with Fu *et al.* and Jiang *et al.* almost achieving comparable and slightly better TRE than Delmon *et al.* TREs of traditional DIR on case 8 were consistently better than that of the DL-based DIR. Case 8 is one of the most challenging cases in the DIRLAB datasets with impaired image quality and significant lung motion. This phenomenon suggests that the robustness and competency of DL-based DIR need to be further improved.

4.2 Brain

Brain image registration has much wider options in databases than lung image registration. As a result, authors were not consistent on which database to use for training and testing and what metrics to use for validations. To facilitate benchmarking, we have listed a number of works on brain image registration in Table 9, which presents the datasets, the registration transformation model and the evaluation metrics. DSC of multiple ROI is the most commonly used evaluation metric. MI and surface distance measures are the next frequently used evaluation metrics.
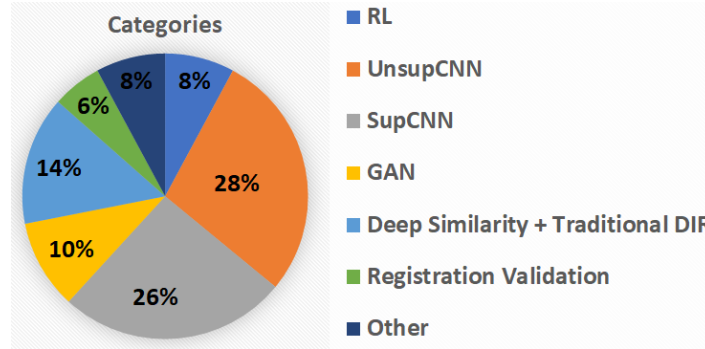
Fig. 3. Percentage pie chart of different categories.

5 Statistics

After careful study of each category, it is important to step back and look at the whole picture. Out of the 150+ papers cited, more than half of the papers were aimed at direct transformation prediction using either supervised or unsupervised transformation prediction. The category of deep similarity-based methods accounts for 14% of all methods while the category of GAN account for 10% of all methods. Publications from the same group (Conference papers which were extended into journal papers) were counted only once if there were no substantial differences in content. One paper may belong to multiple categories. For example, unsupervised CNN method could use GAN generated loss for additional transformation regularization. Details percentages are shown in Fig.3.

Besides the number of papers, we have also analyzed the percentage distributions of many other attributes including input image pair dimension, transformation model, image domain, patch-based training, DL frameworks and ROI of the cited works. The percentage distributions were shown in Fig. 4. 60% of the works were solving 3D-3D registration problems. The 2D-3D image registration works are mostly to register 3D-CT to 2D X-ray images for intraoperation image guidance. The percentages of the number of deformable, rigid and affine registration papers are 72%, 19% and 9%, respectively. Most of the rigid registration papers are for intra-patient brain and spine alignment. There are more publications on unimodal than multi-modal image registration. Due to the superior performance of DL-based similarity measures to traditional similarity measures, the number of DL-based multi-modality image registration papers is increasing and accounts for 41% of all the papers. Patch-based training was often adopted to save GPU memory. Fig.4 shows that 70% of all works used whole image-based training. The 70% includes not only 3D-3D but also 2D-3D and 2D-2D image registrations. Almost all 2D-2D registration used whole image-based training since 2D images are much less memory demanding than 3D images. Therefore, for 3D-3D image registration, there are roughly the same number of works that used whole image-based training and patch-based training. In terms of DL frameworks, Tensorflow is the leading framework which accounts for more than half of all papers. Pytorch is the second most popular DL framework which accounts for a quarter of all papers. Early works used Caffe and Theano, which was used less and less over the years as compared to Tensorflow and Pytorch. Theano has officially ceased development after version 1.0. The deep learning toolbox of Matlab is the least used framework perhaps due to licensing. In terms of the ROI, MR brain and CT lung are the most studied sites. Brain is the top registration target in all works. The reason for the wide adoption of brain include its clinical importance, its availability of public datasets and its relative simplicity of registration.

6 Discussion

Though image registration has been extensively studied, deep learning-based medical image registration is a relatively new research area. We have collected over 150 papers, most of which were published in the last 3 to 4 years. We generally classify these methods into seven non-exclusive categories. Many methods could be classified into multiple categories. For example, GAN was mostly used in combination with supervised or unsupervised transformation prediction methods as an auxiliary regularization or image pre-processing step. Supervised and unsupervised methods were combined for dual supervision in some works. Deep learning-based registration validation methods were included in this review because methods in this category often involve learning a deep similarity metric, therefore, could be used for image registration. RL and deep similarity-based methods are iterative whereas supervised
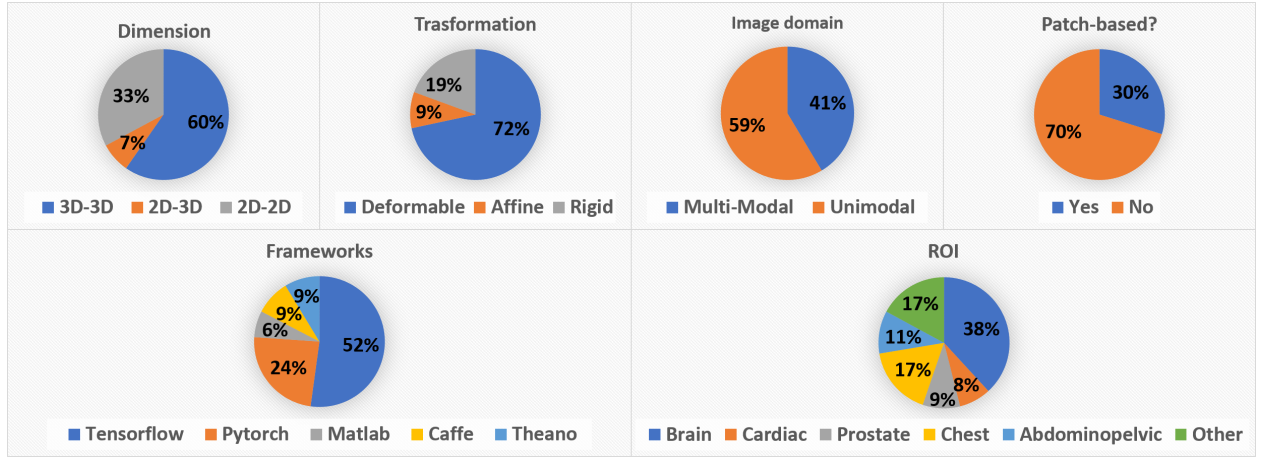
Fig. 4 Percentage pie chart of various attributes of DL-based image registration methods.

and unsupervised based methods are non-iterative. For iterative methods, multiple works have reported that deep similarity metrics have superior performance to handcrafted intensity-based image similarity metric. For non-iterative methods, DL-based methods have yet to outperform traditional DIR methods. Take lung registration for example, the best performing DL-based methods are only comparable to the-state-of-art traditional DIR methods in terms of TRE. However, DL-based direct transformation methods are generally order of magnitude faster than traditional DIR methods. This is mainly due to the non-iterative nature and the powerful GPU utilized. A common feature that is used in both traditional DIR and DL-based methods is multi-scale strategy. Multi-scale registration could help the optimization avoid local maxima and allow large deformation registration. Regarding network generality, Fu *et al.* and Jiang *et al.* both showed that network trained using one set of datasets could be readily applied to an independent set of datasets given that the two image domains are close to each other.

6.1 Whole image-based vs. patch-based transformation prediction

Whole image-based training and patch-based training have their own advantages and disadvantages. Due to limited GPU memory, the original images were often down-sampled to avoid memory overflow in whole image-based training. The down-sampling process could cause information loss and limit the registration accuracy. On the other hand, whole image training allows large inception field which enables registration of large deformations and mitigate the problem of local maxima in registration. Unless data augmentation is used, whole image-based training usually suffers from shortage of training datasets. On the contrary, patch-based training were not affected by the shortage of training datasets as much since many image patches could be sampled from the original images. In addition, patch-based training usually has better performance locally than whole-image based training. Recently, several groups combined whole-image training with patch-based training as a multi-scale approach for image registration [48, 87]. They have achieved promising results in terms of registration accuracy. One challenge with patch-based image registration is the patch fusion process, which stack many image patches to generate the final whole-image transformation. The patch fusion process could generate grid-like artifacts along the edges of the patches. One way to mitigate the problem is to use large patch overlap prior to patch fusion. However, it would make the inference process computationally inefficient. Another method is to use a non-parametric registration model for transformation prediction. One such example is LDDMM model used in QuickSilver [175]. Instead of directly predicting final spatial transformation, QuickSilver predict the momentum of the LDDMM model. The LDDMM model can generate diffeomorphic spatial transformation without the need of smooth momentum predictions.

6.2 Loss functions

Despite large variations in details, loss function definitions of the cited works share many common features. Almost all loss function definitions consist of one or more combinations of the following six types of losses, which are 1) intensity-based image appearance loss, 2) deep similarity-based image appearance loss, 3) transformation smoothness constraint, 4) transformation physical fidelity loss, 5)

transformation error loss with respect to ground truth transformation and 6) adversarial loss. Intensity-based image appearance loss includes SSD, MSE, MAE, MI, MIND, SSIM, CC and its variants. Deep similarity-based image appearance loss usually calculates the correlation between the learnt feature-based image descriptors. Transformation smoothness constraints usually involve the calculation of the first and second orders of spatial derivatives of predicted transformation. Transformation physical fidelity loss includes inverse consistency loss, negative Jacobian determinant loss, identity loss, anti-folding loss and so on. Transformation error loss was the error between predicted and ground truth transformations, which was only valid for supervised transformation prediction. Adversarial loss was the trainable network-based loss. Some auxiliary loss terms include the DSC loss of the anatomical labels or TRE loss of pre-selected landmark pairs.

6.3   Challenges and Opportunities

One of the most common challenges for supervised DL-based methods is the lack of training datasets with known transformations. This problem could be alleviated by various data augmentation methods. However, the data augmentation methods could introduce additional errors such as the bias of unrealistic artificial transformations and image domain shifts between training and testing stages. Several groups have demonstrated good generality of the trained network by applying them to datasets different from the training datasets. This inspired us to think that transfer learning maybe used to alleviate the problem of lack of training data. Surprisingly, transfer learning has not been used in medical image registration. For unsupervised methods, efforts were made to combine different kinds of regularization terms to constrain the predicted transformation. However, it is difficult to investigate the relative importance of each regularization term. Researchers are still trying to find an optimal set of transformation regularization terms that could help generate not only physically plausible but also physiologically realistic deformation field for a certain registration task. This is partially due to the lack of registration validation methods. Due to the unavailability of ground truth transformation between an image pair, it is hard to compare the performances of different registration methods. Therefore, registration validation methods are equally important as registration methods. We have observed an increased number of papers focusing on registration validation in 2019. More research on registration validation methods is desired in order to reliably evaluate the performances of different registration methods under different parametric configurations.

6.4   Trends

Judging from the statistics of the cited works, there is a clear trend of direct transformation prediction for fast image registration. So far, supervised and unsupervised transformation prediction methods are almost equally studied with close number of publications in either category. Either supervised or unsupervised methods have its own advantages and disadvantages. We speculate that more research will be focused on combining supervised and unsupervised methods in the future. GAN-based methods have gradually gaining popularity since GAN could be used to not only introduce additional regularizations but also perform image domain translation to cast multi-modal to unimodal image registration. We should see a steady growth of GAN-based medical image registration. New transformation regularization techniques have always been a hot topic due to the ill-posedness of the registration problem.

Disclosures
The authors declare no conflicts of interest.

References

[1] Else Stougrd Andersen, Karsten stergaard Noe, Thomas Sangild Srensen, Sren Kynde Nielsen, Lars Fokdal, Merete Paludan, Jacob Christian Lindegaard, and Kari Tanderup. Simple dvh parameter addition as compared to deformable registration for bladder dose accumulation in cervix cancer brachytherapy. *Radiotherapy and Oncology*, 107(1):52–57, 2013. ISSN 0167-8140.

[2] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*, 12(1):26–41, 2008. ISSN 1361-8415. doi: 10.1016/j.media.2007.06.004.

[3] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–44, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.09.025.

[4] Bram Bakker. Reinforcement learning with long short-term memory. In *NIPS*.

[5] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. *2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr)*, pages 9252–9260, 2018. ISSN 1063-6919. doi: 10.1109/Cvpr.2018.00964.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *Ieee Transactions on Medical Imaging*, 38(8):1788–1800, 2019. ISSN 0278-0062. doi: 10.1109/Tmi.2019.2897538.

[7] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In Isabelle Guyon, Gideon Dror, Vincent Lemaire, Graham Taylor, and Daniel Silver, editors, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27 of *Proceedings of Machine Learning Research*, pages 37–49, Bellevue, Washington, USA, 02 Jul 2012. PMLR.

[8] F. S. Bashiri, A. Baghaie, R. Rostami, Z. Y. Yu, and R. M. D'Souza. Multi-modal medical image registration with full or partial data: A manifold learning approach. *Journal of Imaging*, 5(1), 2019. ISSN 2313-433x. doi: ARTN510.3390/jimaging5010005.

[9] K. K. Brock, M. B. Sharpe, L. A. Dawson, S. M. Kim, and D. A. Jaffray. Accuracy of finite element model-based multi-organ deformable image registration. *Med Phys*, 32(6):1647–59, 2005. ISSN 0094-2405 (Print)0094-2405. doi: 10.1118/1.1915012.

[10] X. H. Cao, J. H. Yang, J. Zhang, Q. Wang, P. T. Yap, and D. G. Shen. Deformable image registration using a cue-aware deep regression network. *Ieee Transactions on Biomedical Engineering*, 65(9):1900–1911, 2018. ISSN 0018-9294. doi: 10.1109/Tbme.2018.2822826.

[11] Xiaohuan Cao, Jianhuan Yang, Li Wang, Zhong Xue, Qian Wang, and Dinggang Shen. Deep learning based inter-modality image registration supervised by intra-modality similarity. *Machine learning in medical imaging. MLMI*, 11046:55–63, 2018.

[12] Richard Castillo, Edward Castillo, Rudy Guerra, Valen E. Johnson, Travis McPhail, Amit K. Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*, 54(7):1849–1870, 2009. ISSN 0031-91551361-6560. doi: 10.1088/0031-9155/54/7/001.

[13] Raghavendra Chandrashekara, Anil Rao, Gerardo Ivar Sanchez-Ortiz, Raad H. Mohiaddin, and Daniel Rueckert. Construction of a statistical model for cardiac motion analysis using nonrigid image registration. Information Processing in Medical Imaging, pages 599–610. Springer Berlin Heidelberg. ISBN 978-3-540-45087-0.

[14] Evelyn Chee and Zhenzhou Wu. Airnet: Self-supervised affine registration for 3d medical images using neural networks. *ArXiv*, abs/1810.02583, 2018.

[15] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani. Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*, pages 1–1, 2017. ISSN 2372-2096. doi: 10.1109/TBDATA.2017.2717439.

[16] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *NIPS*.

[17] Xi Cheng, Li Zhang, and Yefeng Zheng. Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 6(3):248–252, 2018. ISSN 2168-1163. doi: 10.1080/21681163.2015.1135299.

[18] Kyunghyun Cho, Bart van Merrienboer, aglar Glehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *ArXiv*, abs/1406.1078, 2014.

[19] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8789–8797. ISBN 1063-6919. doi: 10.1109/CVPR.2018.00916.

[20] Junyoung Chung, aglar Glehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv*, abs/1412.3555, 2014.

[21] Adrian V. Dalca, Guha Balakrishnan, John V. Guttag, and Mert R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *MICCAI*.

[22] T. De Silva, A. Uneri, M. D. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. F. Lo, J. P. Wolinsky, and J. H. Siewerdsen. 3d-2d image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. *Phys Med Biol*, 61(8):3009–25, 2016. ISSN 0031-9155. doi: 10.1088/0031-9155/61/8/3009.

[23] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Isgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143, 2019. ISSN 1361-8415. doi: 10.1016/j.media.2018.11.010.

[24] V. Delmon, S. Rit, R. Pinho, and D. Sarrut. Registration of sliding objects using direction dependent b-splines decomposition. *Phys Med Biol*, 58(5):1303–14, 2013. ISSN 0031-9155. doi: 10.1088/0031-9155/58/5/1303.

[25] X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W. J. Curran, T. Liu, and X. Yang. Automatic multiorgan segmentation in thorax ct images using u-net-gan. *Med Phys*, 46(5):2157–2168, 2019. ISSN 0094-2405. doi: 10.1002/mp.13458.

[26] X. Dong, T. Wang, Y. Lei, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang. Synthetic ct generation from non-attenuation corrected pet images for whole-body pet imaging. *Phys Med Biol*, 64(21):215016, 2019. ISSN 0031-9155. doi: 10.1088/1361-6560/ab4eb7.

[27] Xue Dong, Yang Lei, Sibo Tian, Tonghe Wang, Pretesh Patel, Walter J. Curran, Ashesh B. Jani, Tian Liu, and Xiaofeng Yang. Synthetic mri-aided multi-organ segmentation on male pelvic ct using cycle consistent deep attention network. *Radiotherapy and Oncology*, 141:192–199, 2019. ISSN 0167-8140.

[28] Xue Dong, Yang Lei, Tonghe Wang, Kristin Higgins, Tian Liu, Walter J. Curran, Hui Mao, Jonathon A. Nye, and Xiaofeng Yang. Deep learning-based attenuation correction in the absence of structural information for whole-body pet imaging. *Physics in medicine and biology(accepted)*, 2019.

[29] A. Dosovitskiy, P. Fischer, E. Ilg, P. Husser, C. Hazirbas, V. Golkov, P. v. d. Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766. ISBN 2380-7504. doi: 10.1109/ICCV.2015.316.

[30] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. In *NIPS*.

[31] Florian Dubost, Marleen de Bruijne, Marco J. Nardin, Adrian V. Dalca, Kathleen L. Donahue, Anne-Katrin Giese, Mark R Etherton, Ona Wu, Marius de Groot, Wiro J. Niessen, Meike W. Vernooij, Natalia S. Rost, and Markus D. Schirmer. Automated image registration quality assessment utilizing deep-learning based ventricle extraction in clinical data. *ArXiv*, abs/1907.00695, 2019.

[32] M. S. Elmahdy, T. Jagt, R. T. Zinkstok, Y. C. Qiao, R. Shahzad, H. Sokooti, S. Yousefi, L. Incrocci, C. A. M. Marijnen, M. Hoogeman, and M. Staring. Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer. *Medical Physics*, 46(8):3329–3343, 2019. ISSN 0094-2405. doi: 10.1002/mp.13620.

[33] Mohamed S. Elmahdy, Jelmer M. Wolterink, Hessam Sokooti, Ivana Igum, and Marius Staring. Adversarial optimization for joint registration and segmentation in prostate ct radiotherapy. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*, pages 366–374. Springer International Publishing. ISBN 978-3-030-32226-7.

[34] K. A. J. Eppenhof and J. P. W. Pluim. Error estimation of deformable image registration of pulmonary ct scans using convolutional neural networks. *Journal of Medical Imaging*, 5(2), 2018. ISSN 2329-4302. doi: Artn02400310.1117/1.Jmi.5.2.024003.

[35] K. A. J. Eppenhof and J. P. W. Pluim. Pulmonary ct registration through supervised learning with convolutional neural networks. *Ieee Transactions on Medical Imaging*, 38(5):1097–1105, 2019. ISSN 0278-0062. doi: 10.1109/Tmi.2018.2878316.

[36] Koen A. J. Eppenhof, Maxime W. Lafarge, Pim Moeskops, Mitko Veta, and Josien P. W. Pluim. Deformable image registration using convolutional neural networks. In *Medical Imaging: Image Processing*.

[37] J. Fan, X. Cao, Z. Xue, P. T. Yap, and D. Shen. Adversarial similarity network for evaluating image alignment in deep learning based registration. *Med Image Comput Comput Assist Interv*, 11070: 739–746, 2018. doi: 10.1007/978-3-030-00928-1-83.

[38] J. F. Fan, X. H. Cao, Q. Wang, P. T. Yap, and D. G. Shen. Adversarial learning for mono- or multimodal registration. *Medical Image Analysis*, 58, 2019. ISSN 1361-8415. doi: UNSP10154510. 1016/j.media.2019.101545.

[39] J. F. Fan, X. H. Cao, E. A. Yap, and D. G. Shen. Birnet: Brain image registration using dual-supervised fully convolutional networks. *Medical Image Analysis*, 54:193–206, 2019. ISSN 1361-8415. doi: 10.1016/j.media.2019.03.006.

[40] Tobias Fechter and D. Baltas. One shot learning for deformable medical image registration and periodic motion tracking. *ArXiv*, abs/1907.04641, 2019.

[41] E. Ferrante, O. Oktay, B. Glocker, and D. H. Milone. On the adaptability of unsupervised cnn-based deformable image registration to unseen image domains. *Machine Learning in Medical Imaging: 9th International Workshop*, *Mlmi 2018*, 11046:294–302, 2018. ISSN 0302-9743. doi: 10.1007/978-3-030-00919-9-34.

[42] E. Ferrante, P. K. Dokania, R. M. Silva, and N. Paragios. Weakly supervised learning of metric aggregations for deformable image registration. *Ieee Journal of Biomedical and Health Informatics*, 23(4):1374–1384, 2019. ISSN 2168-2194. doi: 10.1109/Jbhi.2018.2869700.

[43] Markus D. Foote, Blake E. Zimmerman, Amit Sawant, and Sarang C. Joshi. Real-time 2d-3d deformable registration with deep learning and application to lung radiotherapy targeting. *Information Processing in Medical Imaging*, pages 265–276. Springer International Publishing. ISBN 978-3-030-20351-1.

[44] Y. Fu, S. Liu, H. Li, and D. Yang. Automatic and hierarchical segmentation of the human skeleton in ct images. *Phys Med Biol*, 62(7):2812–2833, 2017. ISSN 0031-9155. doi: 10.1088/1361-6560/ aa6055.

[45] Y. Fu, T. R. Mazur, X. Wu, S. Liu, X. Chang, Y. Lu, H. H. Li, H. Kim, M. C. Roach, L. Henke, and D. Yang. A novel mri segmentation method using cnn-based correction network for mri-guided adaptive radiotherapy. *Med Phys*, 45(11):5129–5137, 2018. ISSN 0094-2405. doi: 10.1002/mp. 13221.

[46] Y. B. Fu, C. K. Chui, C. L. Teo, and E. Kobayashi. Motion tracking and strain map computation for quasi-static magnetic resonance elastography. *Med Image Comput Comput Assist Interv*, 14(Pt 1):428–35, 2011. doi: 10.1007/978-3-642-23623-5-54.

[47] Y. B. Fu, X. Wu, A. M. Thomas, H. H. Li, and D. S. Yang. Automatic large quantity landmark pairs detection in 4dct lung images. *Medical Physics*, 46(10):4490–4501, 2019. ISSN 0094-2405. doi: 10.1002/mp.13726.

[48] Yabo Fu, Yang Lei, Tonghe Wang, Yingzi Liu, Pretesh Patel, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Lungregnet: an unsupervised deformable image registration method for 4d-ct lung. *Medical physics(submitted)*, 2019.

[49] S. M. Galib, H. K. Lee, C. L. Guy, M. J. Riblett, and G. D. Hugo. A fast and scalable method for quality assurance of deformable image registration on lung ct scans using convolutional neural networks. *Medical Physics*, 2019. ISSN 0094-2405. doi: 10.1002/mp.13890.

[50] F. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu. Multi-scale deep reinforcement learning for real-time 3d-landmark detection in ct scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):176–189, 2019. ISSN 1939-3539. doi: 10.1109/TPAMI.2017.2782687.

[51] Florin C. Ghesu, Bogdan Georgescu, Tommaso Mansi, Dominik Neumann, Joachim Hornegger, and Dorin Comaniciu. An artificial agent for anatomical landmark detection in medical images. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, pages 229–237. Springer International Publishing. ISBN 978-3-319-46726-9.

[52] S. Ghosal and N. Rayl. Deep deformable registration: Enhancing accuracy by fully convolutional neural net. *Pattern Recognition Letters*, 94:81–86, 2017. ISSN 0167-8655. doi: 10.1016/j.patrec.2017.05.022.

[53] C. L. Giles, G. M. Kuhn, and R. J. Williams. Dynamic recurrent neural networks: Theory and applications. *IEEE Transactions on Neural Networks*, 5(2):153–156, 1994. ISSN 1941-0093. doi: 10.1109/TNN.1994.8753425.

[54] L. Gong, H. Wang, C. Peng, Y. Dai, M. Ding, Y. Sun, X. Yang, and J. Zheng. Non-rigid mr-trus image registration for image-guided prostate biopsy using correlation ratio-based mutual information. *Biomed Eng Online*, 16(1):8, 2017. ISSN 1475-925x. doi: 10.1186/s12938-016-0308-5.

[55] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang. A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7):4328–4338, 2014. ISSN 1558-0644. doi: 10.1109/TGRS.2013.2281391.

[56] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*.

[57] Xiao Han, Mischa S. Hoogeman, Peter C. Levendag, Lyndon S. Hibbard, David N. Teguh, Peter Voet, Andrew C. Cowen, and Theresa K. Wolf. Atlas-based auto-segmentation of head and neck ct images. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2008, pages 434–441. Springer Berlin Heidelberg. ISBN 978-3-540-85990-1.

[58] J. Harms, Y. Lei, T. Wang, R. Zhang, J. Zhou, X. Tang, W. J. Curran, T. Liu, and X. Yang. Paired cycle-gan-based image correction for quantitative cone-beam computed tomography. *Med Phys*, 46(9):3998–4009, 2019. ISSN 0094-2405. doi: 10.1002/mp.13656.

[59] G. Haskins, J. Kruecker, U. Kruger, S. Xu, P. A. Pinto, B. J. Wood, and P. K. Yan. Learning deep similarity metric for 3d mr-trus image registration. *International Journal of Computer Assisted Radiology and Surgery*, 14(3):417–425, 2019. ISSN 1861-6410. doi: 10.1007/s11548-018-1875-7.

[60] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: A survey. *ArXiv*, abs/1903.02026, 2019.

[61] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. ISBN 1063-6919. doi: 10.1109/CVPR.2016.90.

[62] M. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel. Mrf-based deformable registration and ventilation estimation of lung ct. *IEEE Trans Med Imaging*, 32(7):1239–48, 2013. ISSN 0278-0062. doi: 10.1109/tmi.2013.2246577.

[63] Mattias P. Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V. Gleeson, Sir Michael Brady, and Julia A. Schnabel. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012. ISSN 1361-8415.

[64] Alessa Hering, Sven Kuckertz, Stefan Heldmann, and Mattias P. Heinrich. Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking. In *Bildverarbeitung fr die Medizin*.

[65] R. Devon Hjelm, Sergey M. Plis, and Vince D. Calhoun. Variational autoencoders for feature detection of magnetic resonance imaging data. *ArXiv*, abs/1603.06624, 2016.

[66] Y. P. Hu, M. Modat, E. Gibson, W. Q. Li, N. Ghavamia, E. Bonmati, G. T. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis*, 49:1–13, 2018. ISSN 1361-8415. doi: 10.1016/j.media.2018.07.002.

[67] Yipeng Hu, Eli Gibson, Nooshin Ghavami, Ester Bonmati, Caroline M. Moore, Mark Emberton, Tom Vercauteren, J. Alison Noble, and Dean C. Barratt. Adversarial deformation regularization for training image registration neural networks. In *MICCAI*.

[68] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269. ISBN 1063-6919. doi: 10.1109/CVPR.2017.243.

[69] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. *ArXiv*, abs/1506.02025, 2015.

[70] Pingge Jiang and James A. Shackleford. Cnn driven sparse multi-level b-spline image registration. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9281–9289, 2018.

[71] Z. Jiang, F. F. Yin, Y. Ge, and L. Ren. A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration. *Phys Med Biol*, 2019. ISSN 0031-9155. doi: 10.1088/1361-6560/ab5da0.

[72] V. Kearney, S. Haaf, A. Sudhyadhom, G. Valdes, and T. D. Solberg. An unsupervised convolutional neural network-based algorithm for deformable image registration. *Physics in Medicine and Biology*, 63(18), 2018. ISSN 0031-9155. doi: ARTN18501710.1088/1361-6560/aada66.

[73] J. Ker, L. P. Wang, J. Rao, and T. Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2018. ISSN 2169-3536. doi: 10.1109/Access.2017.2788044.

[74] Boah Kim, Jieun Kim, June-Goo Lee, Dong Hwan Kim, Seong Ho Park, and Jong Chul Ye. Unsupervised deformable image registration using cycle-consistent cnn. In *MICCAI*.

[75] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*, 29(1):196–205, 2010. ISSN 0278-0062. doi: 10.1109/tmi.2009.2035616.

[76] Avinash Kori and Ganapathi Krishnamurthi. Zero shot learning for multi-modal real time image registration. *ArXiv*, abs/1908.06213, 2019.

[77] J. Krebs, T. Mansi, B. Mailhe, N. Ayache, and H. Delingette. Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Dlmia 2018*, 11045:101–109, 2018. ISSN 0302-9743. doi: 10.1007/978-3-030-00889-5-12.

[78] Julian Krebs, Tommaso Mansi, Herv Delingette, Li Zhang, Florin C. Ghesu, Shun Miao, Andreas K. Maier, Nicholas Ayache, Rui Liao, and Ali Kamen. Robust non-rigid registration through agent-based action learning. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pages 344–352. Springer International Publishing. ISBN 978-3-319-66182-7.

[79] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*.

[80] Dongyang Kuang. On reducing negative jacobian determinant of the deformation predicted by deep registration networks. *ArXiv*, abs/1907.00068, 2019.

[81] Dongyang Kuang and Tanya Schmah. Faim a convnet method for unsupervised 3d medical image registration. Machine Learning in Medical Imaging, pages 646–654. Springer International Publishing. ISBN 978-3-030-32692-0.

[82] Ting Fung Lau, Ji Luo, Shengyu Zhao, Eric I-Chao Chang, and Yan Xu. Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE journal of biomedical and health informatics*, 2019.

[83] Y. Lei, X. Dong, Z. Tian, Y. Liu, S. Tian, T. Wang, X. Jiang, P. Patel, A. B. Jani, H. Mao, W. J. Curran, T. Liu, and X. Yang. Ct prostate segmentation based on synthetic mri-aided deep attention fully convolution network. *Med Phys*, 2019. ISSN 0094-2405. doi: 10.1002/mp.13933.

[84] Y. Lei, J. Harms, T. Wang, Y. Liu, H. K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang. Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. *Med Phys*, 46(8):3565–3581, 2019. ISSN 0094-2405. doi: 10.1002/mp.13617.

[85] Y. Lei, X. Tang, K. Higgins, J. Lin, J. Jeong, T. Liu, A. Dhabaan, T. Wang, X. Dong, R. Press, W. J. Curran, and X. Yang. Learning-based cbct correction using alternating random forest based on auto-context model. *Med Phys*, 46(2):601–618, 2019. ISSN 0094-2405. doi: 10.1002/mp.13295.

[86] Y. Lei, S. Tian, X. He, T. Wang, B. Wang, P. Patel, A. B. Jani, H. Mao, W. J. Curran, T. Liu, and X. Yang. Ultrasound prostate segmentation based on multidirectional deeply supervised v-net. *Med Phys*, 46(7):3194–3206, 2019. ISSN 0094-2405. doi: 10.1002/mp.13577.

[87] Yang Lei, Yabo Fu, Joseph Harms, Tonghe Wang, Walter J. Curran, Tian Liu, Kristin Higgins, and Xiaofeng Yang. 4d-ct deformable image registration using an unsupervised deep convolutional neural network. Artificial Intelligence in Radiation Therapy, pages 26–33. Springer International Publishing. ISBN 978-3-030-32486-5.

[88] Yang Lei, Xue Dong, Tonghe Wang, Kristin Higgins, Tian Liu, Walter J. Curran, Hui Mao, Jonathon A. Nye, and Xiaofeng Yang. Whole-body pet estimation from low count statistics using cycle-consistent generative adversarial networks. *Physics in Medicine and Biology*, 64(21):215017, 2019. ISSN 1361-6560. doi: 10.1088/1361-6560/ab4891.

[89] Yang Lei, Joseph Harms, Tonghe Wang, Sibo Tian, Jun Zhou, Hui-Kuo Shu, Jim Zhong, Hui Mao, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Mri-based synthetic ct generation using semantic random forest with iterative refinement. *Physics in Medicine and Biology*, 64(8):085001, 2019. ISSN 1361-6560. doi: 10.1088/1361-6560/ab0b66.

[90] H. Li and Y. Fan. Non-rigid image registration using self-supervised fully convolutional networks without training data. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1075–1078. ISBN 1945-8452. doi: 10.1109/ISBI.2018.8363757.

[91] R. Li, X. Jia, J. H. Lewis, X. Gu, M. Folkerts, C. Men, and S. B. Jiang. Real-time volumetric image reconstruction and 3d tumor localization based on a single x-ray projection image for lung cancer radiotherapy. *Med Phys*, 37(6):2822–6, 2010. ISSN 0094-2405 (Print)0094-2405. doi: 10.1118/1.3426002.

[92] Rui Liao, Shun Miao, Pierre de Tournemire, Sasa Grbic, Ali Kamen, Tommaso Mansi, and Dorin Comaniciu. An artificial agent for robust image registration. *ArXiv*, abs/1611.10336, 2016.

[93] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sanchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. ISSN 1361-8415. doi: 10.1016/j.media.2017.07.005.

[94] C. Liu, L. H. Ma, Z. M. Lu, X. C. Jin, and J. Y. Xu. Multimodal medical image registration via common representations learning and differentiable geometric constraints. *Electronics Letters*, 55 (6):316–318, 2019. ISSN 0013-5194. doi: 10.1049/el.2018.6713.

[95] X. L. Liu, D. S. Jiang, M. N. Wang, and Z. J. Song. Image synthesis-based multi-modal image registration framework by using deep fully convolutional networks. *Medical and Biological Engineering and Computing*, 57(5):1037–1048, 2019. ISSN 0140-0118. doi: 10.1007/s11517-018-1924-y.

[96] Y. Liu, X. Chen, Z. F. Wang, Z. J. Wang, R. K. Ward, and X. S. Wang. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42:158–173, 2018. ISSN 1566-2535. doi: 10.1016/j.inffus.2017.10.007.

[97] Y. Liu, Y. Lei, T. Wang, O. Kayode, S. Tian, T. Liu, P. Patel, W. J. Curran, L. Ren, and X. Yang. Mri-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic ct generation method. *Br J Radiol*, 92(1100):20190067, 2019. ISSN 0007-1285. doi: 10.1259/bjr.20190067.

[98] Yingzi Liu, Yang Lei, Yinan Wang, Tonghe Wang, Lei Ren, Liyong Lin, Mark McDonald, Walter J. Curran, Tian Liu, Jun Zhou, and Xiaofeng Yang. Mri-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic ct generation method. *Physics in Medicine and Biology*, 64(14):145015, 2019. ISSN 1361-6560. doi: 10.1088/1361-6560/ab25bc.

[99] Wenhan Luo, Peng Sun, Fangwei Zhong, Wei Liu, Tong Zhang, and Yizhou Wang. End-to-end active object tracking via reinforcement learning. In *ICML*.

[100] J. Lv, M. Yang, J. Zhang, and X. Y. Wang. Respiratory motion correction for free-breathing 3d abdominal mri using cnn-based image registration: a feasibility study. *British Journal of Radiology*, 91(1083), 2018. ISSN 0007-1285. doi: ARTN2017078810.1259/bjr.20170788.

[101] Kai Ma, Jiangping Wang, Vivek Singh, Birgi Tamersoy, Yao-Jen Chang, Andreas Wimmer, and Terrence Chen. Multimodal image registration with deep context reinforcement learning. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pages 240–248. Springer International Publishing. ISBN 978-3-319-66182-7.

[102] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi. Deformable medical image registration using generative adversarial networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1449–1453. ISBN 1945-8452. doi: 10.1109/ISBI.2018.8363845.

[103] D. Mahapatra, Z. Y. Ge, S. Sedai, and R. Chakravorty. Joint registration and segmentation of xray images using generative adversarial networks. *Machine Learning in Medical Imaging: 9th International Workshop*, *Mlmi 2018*, 11046:73–80, 2018. ISSN 0302-9743. doi: 10.1007/978-3-030-00919-9-9.

[104] Dwarikanath Mahapatra, Suman Sedai, and Rahil Garnavi. Elastic registration of medical images with gans. *ArXiv*, abs/1805.02369, 2018.

[105] A. Maier, C. Syben, T. Lasser, and C. Riess. A gentle introduction to deep learning in medical image processing. *Zeitschrift Fur Medizinische Physik*, 29(2):86–101, 2019. ISSN 0939-3889. doi: 10.1016/j.zemedi.2018.12.003.

[106] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement. Survey on deep learning for radiotherapy. *Computers in Biology and Medicine*, 98:126–146, 2018. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2018.05.018.

[107] S. Miao, Z. J. Wang, and R. Liao. A cnn regression approach for real-time 2d/3d registration. *Ieee Transactions on Medical Imaging*, 35(5):1352–1363, 2016. ISSN 0278-0062. doi: 10.1109/Tmi.2016.2521800.

[108] Shun Miao, Sebastien Piat, Peter Walter Fischer, Ahmet Tuysuzoglu, Philip Walter Mewes, Tommaso Mansi, and Rui Liao. Dilated fcn for multi-agent 2d/3d medical image registration. *ArXiv*, abs/1712.01651, 2017.

[109] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *ArXiv*, abs/1411.1784, 2014.

[110] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. ISSN 1476-4687. doi: 10.1038/nature14236.

[111] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed*, 98(3):278–84, 2010. ISSN 0169-2607. doi: 10.1016/j.cmpb.2009.09.002.

[112] J. Neylon, Y. G. Min, D. A. Low, and A. Santhanam. A neural network approach for fast, automated quantification of dir performance. *Medical Physics*, 44(8):4126–4138, 2017. ISSN 0094-2405. doi: 10.1002/mp.12321.

[113] Jorge Onieva Onieva, Berta Marti-Fuster, Mara Pedrero de la Puente, and Ral San Jos Estpar. Diffeomorphic lung registration using deep cnns and reinforced learning. Image Analysis for Moving Organ, Breast, and Thoracic Images, pages 284–294. Springer International Publishing. ISBN 978-3-030-00946-5.

[114] Y. R. Pei, Y. G. Zhang, H. F. Qin, G. Y. Ma, Y. K. Guo, T. M. Xu, and H. B. Zha. Non-rigid craniofacial 2d-3d registration using cnn-based regression. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 10553:117–125, 2017. ISSN 0302-9743. doi: 10.1007/978-3-319-67558-9-14.

[115] C. Qin, B. B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen. Unsupervised deformable registration for multi-modal images via disentangled representations. *Information Processing in Medical Imaging, Ipmi 2019*, 11492:249–261, 2019. ISSN 0302-9743. doi: 10.1007/978-3-030-20351-1-19.

[116] Chen Qin, Wenjia Bai, Jo Schlemper, Steffen E. Petersen, Stefan K. Piechnik, Stefan Neubauer, and Daniel Rueckert. Joint learning of motion estimation and segmentation for cardiac mr image sequences. *ArXiv*, abs/1806.04066, 2018.

[117] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017. ISSN 1939-3539. doi: 10.1109/TPAMI.2016.2577031.

[118] H. Rivaz, Z. Karimaghaloo, V. S. Fonov, and D. L. Collins. Nonrigid registration of ultrasound and mri using contextual conditioned mutual information. *IEEE Trans Med Imaging*, 33(3):708–25, 2014. ISSN 0278-0062. doi: 10.1109/tmi.2013.2294630.

[119] Marc-Michel Roh, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: Learning deformable image registration using shape matching. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pages 266–274. Springer International Publishing. ISBN 978-3-319-66182-7.

[120] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015.

[121] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. S. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger. Deep learning in medical imaging and radiation therapy. *Medical Physics*, 46(1):e1–e36, 2019. ISSN 0094-2405. doi: 10.1002/mp.13264.

[122] Seyed Sadegh Mohseni Salehi, Shadab Khan, Deniz Erdomu, and Ali Gholipour. Real-time deep registration with geodesic loss. *ArXiv*, abs/1803.05982, 2018.

[123] David Sarrut. Deformable registration for image-guided radiation therapy. *Zeitschrift fr Medizinische Physik*, 16(4):285–297, 2006. ISSN 0939-3889.

[124] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias P. Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197207, 2018.

[125] Alireza Sedghi, Jie Luo, Alireza Mehrtash, Steven D. Pieper, Clare M. Tempany, Tina Kapur, Parvin Mousavi, and William M. Wells. Semi-supervised deep metrics for image registration. *ArXiv*, abs/1804.01565, 2018.

[126] Thilo Sentker, Frederic Madesta, and Ren Werner. Gdl-fire text 4d : Deep learning-based fast 4d ct image registration. In *MICCAI*.

[127] J. A. Shackleford, N. Kandasamy, and G. C. Sharp. On developing b-spline registration algorithms for multi-core processors. *Phys Med Biol*, 55(21):6329–51, 2010. ISSN 0031-9155. doi: 10.1088/0031-9155/55/21/001.

[128] R. Shams, P. Sadeghi, R. A. Kennedy, and R. I. Hartley. A survey of medical image registration on multicore and the gpu. *IEEE Signal Processing Magazine*, 27(2):50–60, 2010. ISSN 1558-0792. doi: 10.1109/MSP.2009.935387.

[129] Siyuan Shan, Xiaoqing Guo, Wen Yan, Eric I-Chao Chang, Yubo Fan, and Yan Xu. Unsupervised end-to-end learning for deformable medical image registration. *ArXiv*, abs/1711.08608, 2017.

[130] A. Sheikhjafari, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. Unsupervised deformable image registration with fully connected generative neural network.

[131] Dinggang Shen. Image registration by local histogram matching. *Pattern Recognition*, 40(4): 1161–1172, 2007. ISSN 0031-3203.

[132] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017. ISSN 1545-42741523-9829. doi: 10.1146/annurev-bioeng-071516-044442.

[133] Chang Shu, Xi Chen, Qiwei Xie, and Hua Han. *An unsupervised network for fast microscopic image registration*, volume 10581 of *SPIE Medical Imaging*. SPIE, 2018.

[134] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016. ISSN 1476-4687. doi: 10.1038/nature16961.

[135] Martin Simonovsky, Benjamn Gutirrez-Becker, Diana Mateus, Nassir Navab, and Nikos Komodakis. A deep metric for multimodal registration. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, pages 10–18. Springer International Publishing. ISBN 978-3-319-46726-9.

[136] James M. Sloan, Keith A. Goatman, and J. Paul Siebert. Learning rigid image registration - utilizing convolutional neural networks for medical image registration. In *BIOIMAGING*.

[137] R. W. K. So and A. C. S. Chung. A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using bhattacharyya distances. *Pattern Recognition*, 62:161–174, 2017. ISSN 0031-3203. doi: 10.1016/j.patcog.2016.09.004.

[138] H. Sokooti, G. Saygili, B. Glocker, B. P. F. Lelieveldt, and M. Staring. Quantitative error prediction of medical image registration using regression forests. *Medical Image Analysis*, 56:110–121, 2019. ISSN 1361-8415. doi: 10.1016/j.media.2019.05.005.

[139] Hessam Sokooti, Bob de Vos, Floris Berendsen, Boudewijn P. F. Lelieveldt, Ivana Igum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pages 232–239. Springer International Publishing. ISBN 978-3-319-66182-7.

[140] Hessam Sokooti, Bob D. de Vos, Floris F. Berendsen, Mohsen Ghafoorian, Sahar Yousefi, Boudewijn P. F. Lelieveldt, Ivana Igum, and Marius Staring. 3d convolutional neural networks image registration based on efficient supervised learning from artificial deformations. *ArXiv*, abs/1908.10235, 2019.

[141] Marius Staring, Stefan Klein, Wiro J. Niessen, Berend C. Stoel, and Erasmus Mc. Pulmonary image registration with elastix using a standard intensity-based algorithm.

[142] Christodoulidis Stergios, Sahasrabudhe Mihir, Vakalopoulou Maria, Chassagnon Guillaume, Revel Marie-Pierre, Mougiakakou Stavroula, and Paragios Nikos. Linear and deformable image registration with 3d convolutional neural networks. Image Analysis for Moving Organ, Breast, and Thoracic Images, pages 13–22. Springer International Publishing. ISBN 978-3-030-00946-5.

[143] Li Sun and Songtao Zhang. Deformable mri-ultrasound registration using 3d convolutional neural network. Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation, pages 152–158. Springer International Publishing. ISBN 978-3-030-01045-4.

[144] Shanhui Sun, Jing Hu, Mingqing Yao, Jinrong Hu, Xiaodong Yang, Qi Song, and Xi Wu. Robust multimodal image registration using deep recurrent reinforcement learning. Computer Vision ACCV 2018, pages 511–526. Springer International Publishing, . ISBN 978-3-030-20890-5.

[145] Yuanyuan Sun, Adriaan Moelker, Wiro J. Niessen, and Theo van Walsum. Towards robust ct-ultrasound registration using deep learning methods. Understanding and Interpreting Machine Learning in Medical Image Computing Applications, pages 43–51. Springer International Publishing, . ISBN 978-3-030-02628-8.

[146] C. Szegedy, Liu Wei, Jia Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. ISBN 1063-6919. doi: 10.1109/CVPR.2015.7298594.

[147] Richard Szeliski and James Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3):199–218, 1997. ISSN 1573-1405. doi: 10.1023/a:1007996332012.

[148] R. H. Taylor and D. Stoianovici. Medical robotics in computer-integrated surgery. *IEEE Transactions on Robotics and Automation*, 19(5):765–781, 2003. ISSN 2374-958X. doi: 10.1109/TRA.2003.817058.

[149] N. D. Thanh, I. Yoo, L. Thomas, A. Kuan, W. C. Lee, and W. K. Jeong. Weakly supervised learning in deformable em image registration using slice interpolation. *2019 Ieee 16th International Symposium on Biomedical Imaging (Isbi 2019)*, pages 670–673, 2019. ISSN 1945-7928.

[150] Sebastian Thrun. Efficient exploration in reinforcement learning.

[151] Michael Tschannen, Olivier Bachem, and Mario Lucic. Recent advances in autoencoder-based representation learning. *ArXiv*, abs/1812.05069, 2018.

[152] Hristina Uzunova, Matthias Wilms, Heinz Handels, and Jan Ehrhardt. Training cnns for image registration from few samples with model-based data augmentation. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pages 223–231. Springer International Publishing. ISBN 978-3-319-66182-7.

[153] Michael Velec, Joanne L. Moseley, Cynthia L. Eccles, Tim Craig, Michael B. Sharpe, Laura A. Dawson, and Kristy K. Brock. Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *International Journal of Radiation Oncology*Biology*Physics*, 80(1):265–272, 2011. ISSN 0360-3016.

[154] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supplement 1):S61–S72, 2009. ISSN 1053-8119.

[155] Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Igum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *DLMIA/ML-CDS@MICCAI*.

[156] B. Wang, Y. Lei, S. Tian, T. Wang, Y. Liu, P. Patel, A. B. Jani, H. Mao, W. J. Curran, T. Liu, and X. Yang. Deeply supervised 3d fully convolutional networks with group dilated convolution for automatic mri prostate segmentation. *Med Phys*, 46(4):1707–1718, 2019. ISSN 0094-2405. doi: 10.1002/mp.13416.

[157] T. Wang, B. B. Ghavidel, J. J. Beitler, X. Tang, Y. Lei, W. J. Curran, T. Liu, and X. Yang. Optimal virtual monoenergetic image in "twinbeam" dual-energy ct for organs-at-risk delineation based on contrast-noise-ratio in head-and-neck radiotherapy. *J Appl Clin Med Phys*, 20(2):121–128, 2019. ISSN 1526-9914. doi: 10.1002/acm2.12539.

[158] T. Wang, Y. Lei, N. Manohar, S. Tian, A. B. Jani, H. K. Shu, K. Higgins, A. Dhabaan, P. Patel, X. Tang, T. Liu, W. J. Curran, and X. Yang. Dosimetric study on learning-based cone-beam ct correction in adaptive radiation therapy. *Med Dosim*, 44(4):e71–e79, 2019. ISSN 1873-4022. doi: 10.1016/j.meddos.2019.03.001.

[159] T. Wang, Y. Lei, Z. Tian, X. Dong, Y. Liu, X. Jiang, W. J. Curran, T. Liu, H. K. Shu, and X. Yang. Deep learning-based image quality improvement for low-dose computed tomography simulation in radiation therapy. *J Med Imaging (Bellingham)*, 6(4):043504, 2019. ISSN 2329-4302 (Print)2329-4302. doi: 10.1117/1.Jmi.6.4.043504.

[160] T. Wang, N. Manohar, Y. Lei, A. Dhabaan, H. K. Shu, T. Liu, W. J. Curran, and X. Yang. Mri-based treatment planning for brain stereotactic radiosurgery: Dosimetric validation of a learning-based pseudo-ct generation method. *Med Dosim*, 44(3):199–204, 2019. ISSN 1873-4022. doi: 10.1016/j.meddos.2018.06.008.

[161] Tonghe Wang, Yang Lei, Haipeng Tang, Zhuo He, Richard Castillo, Cheng Wang, Dianfu Li, Kristin Higgins, Tian Liu, Walter J. Curran, Weihua Zhou, and Xiaofeng Yang. A learning-based automatic segmentation and quantification method on left ventricle in gated myocardial perfusion spect imaging: A feasibility study. *Journal of Nuclear Cardiology*, 2019. ISSN 1532-6551. doi: 10.1007/s12350-019-01594-2.

[162] R. Werner, A. Schmidt-Richberg, H. Handels, and J. Ehrhardt. Estimation of lung motion fields in 4d ct data by variational non-linear intensity-based registration: A comparison and evaluation study. *Phys Med Biol*, 59(15):4247–60, 2014. ISSN 0031-9155. doi: 10.1088/0031-9155/59/15/4247.

[163] Robert Wright, Bishesh Khanal, Alberto Gomez, Emily Skelton, Jacqueline Matthew, Jo V. Hajnal, Daniel Rueckert, and Julia A. Schnabel. Lstm spatial co-transformer networks for registration of 3d fetal us and mr brain images. Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis, pages 149–159. Springer International Publishing. ISBN 978-3-030-00807-9.

[164] G. R. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. G. Shen. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *Ieee Transactions on Biomedical Engineering*, 63(7):1505–1516, 2016. ISSN 0018-9294. doi: 10.1109/Tbme.2015.2496253.

[165] K. J. Xia, H. S. Yin, and J. Q. Wang. A novel improved deep convolutional neural network model for medical image fusion. *Cluster Computing-the Journal of Networks Software Tools and Applications*, 22:1515–1527, 2019. ISSN 1386-7857. doi: 10.1007/s10586-018-2026-1.

[166] Pingkun Yan, Sheng Xu, Ardeshir R. Rastinehad, and Bradford J. Wood. Adversarial image registration with application for mr and trus image fusion. In *MLMI@MICCAI*.

[167] D. Yang, H. Li, D. A. Low, J. O. Deasy, and I. El Naqa. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Phys Med Biol*, 53(21):6143–65, 2008. ISSN 0031-9155 (Print) 0031-9155. doi: 10.1088/0031-9155/53/21/017.

[168] D. Yang, S. M. Goddu, W. Lu, O. L. Pechenaya, Y. Wu, J. O. Deasy, I. El Naqa, and D. A. Low. Technical note: deformable image registration on partially matched images for radiotherapy applications. *Med Phys*, 37(1):141–5, 2010. ISSN 0094-2405 (Print) 0094-2405. doi: 10.1118/1.3267547.

[169] D. Yang, S. Brame, I. El Naqa, A. Aditya, Y. Wu, S. M. Goddu, S. Mutic, J. O. Deasy, and D. A. Low. Technical note: Dirart–a software suite for deformable image registration and adaptive radiotherapy research. *Med Phys*, 38(1):67–77, 2011. ISSN 0094-2405 (Print) 0094-2405. doi: 10.1118/1.3521468.

[170] X. Yang, H. Akbari, L. Halig, and B. Fei. 3d non-rigid registration using surface and local salient features for transrectal ultrasound image-guided prostate biopsy. *Proc SPIE Int Soc Opt Eng*, 7964:79642v, 2011. ISSN 0277-786X (Print) 0277-786x. doi: 10.1117/12.878153.

[171] X. Yang, D. Schuster, V. Master, P. Nieh, A. Fenster, and B. Fei. Automatic 3d segmentation of ultrasound images using atlas registration and statistical texture prior. *Proc SPIE Int Soc Opt Eng*, 7964, 2011. ISSN 0277-786X (Print) 0277-786x. doi: 10.1117/12.877888.

[172] X. Yang, P. Ghafourian, P. Sharma, K. Salman, D. Martin, and B. Fei. Nonrigid registration and classification of the kidneys in 3d dynamic contrast enhanced (dce) mr images. *Proc SPIE Int Soc Opt Eng*, 8314:83140b, 2012. ISSN 0277-786X (Print) 0277-786x. doi: 10.1117/12.912190.

[173] X. Yang, N. Wu, G. Cheng, Z. Zhou, D. S. Yu, J. J. Beitler, W. J. Curran, and T. Liu. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal mri study in head-and-neck radiation therapy. *Int J Radiat Oncol Biol Phys*, 90(5):1225–33, 2014. ISSN 0360-3016. doi: 10.1016/j.ijrobp.2014.08.350.

[174] X. Yang, P. J. Rossi, A. B. Jani, H. Mao, W. J. Curran, and T. Liu. 3d transrectal ultrasound (trus) prostate segmentation based on optimal feature learning framework. *Proc SPIE Int Soc Opt Eng*, 9784, 2016. ISSN 0277-786X (Print) 0277-786x. doi: 10.1117/12.2216396.

[175] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration a deep learning approach. *NeuroImage*, 158:378–396, 2017. ISSN 1053-8119.

[176] Xiaofeng Yang and Baowei Fei. 3d prostate segmentation of ultrasound images combining longitudinal image registration and machine learning. *Proceedings of SPIE–the International Society for Optical Engineering*, 8316:83162O–83162O, 2012. ISSN 0277-786X. doi: 10.1117/12.912188.

[177] Inwan Yoo, David G. C. Hildebrand, Willie F. Tobin, Wei-Chung Allen Lee, and Won-Ki Jeong. ssemnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features. In *DLMIA/ML-CDS@MICCAI*.

[178] H. C. Yu, Y. Fu, H. C. Yu, J. B. Jiao, Y. C. Wei, X. C. Wang, B. H. Wen, Z. Y. Wang, M. Bramlet, T. Kesavadas, H. H. Shi, and T. Huang. A novel framework for 3d-2d vertebra matching. *2019 2nd Ieee Conference on Multimedia Information Processing and Retrieval (Mipr 2019)*, pages 121–126, 2019. doi: 10.1109/Mipr.2019.00029.

[179] H. J. Yu, X. R. Zhou, H. Y. Jiang, H. J. Kang, Z. G. Wang, T. Hara, and H. Fujita. Learning 3d non-rigid deformation based on an unsupervised deep learning for pet/ct image registration. *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 10953, 2019. ISSN 0277-786x. doi: Artn109531x10.1117/12.2512698.

[180] A. L. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Comput*, 15(4):915–36, 2003. ISSN 0899-7667 (Print) 0899-7667. doi: 10.1162/08997660360581958.

[181] Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration. *ArXiv*, abs/1809.03443, 2018.

[182] Z. W. Zhang and E. Sejdic. Radiological images and machine learning: Trends, perspectives, and prospects. *Computers in Biology and Medicine*, 108:354–370, 2019. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2019.02.017.

[183] J. N. Zheng, S. Miao, Z. J. Wang, and R. Liao. Pairwise domain adaptation module for cnn-based 2-d/3-d registration. *Journal of Medical Imaging*, 5(2), 2018. ISSN 2329-4302. doi: Artn02120410. 1117/1.Jmi.5.2.021204.

[184] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251. ISBN 2380-7504. doi: 10.1109/ICCV.2017.244.

[185] David Zimmerer, Simon A. A. Kohl, Jens Petersen, Fabian Isensee, and Klaus H. Maier-Hein. Context-encoding variational autoencoder for unsupervised anomaly detection. *ArXiv*, abs/1812.05941, 2018.