



A review of deep learning-based three-dimensional medical image registration methods

Haonan Xiao¹, Xinzhi Teng¹, Chenyang Liu¹, Tian Li¹, Ge Ren¹, Ruijie Yang², Dinggang Shen^{3,4,5}, Jing Cai¹

¹Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China; ²Department of Radiation Oncology, Peking University Third Hospital, Beijing, China; ³School of Biomedical Engineering, ShanghaiTech University, Shanghai, China; ⁴Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China; ⁵Department of Artificial Intelligence, Korea University, Seoul, Republic of Korea

Contributions: (I) Conception and design: J Cai; (II) Administrative support: J Cai; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: H Xiao; (V) Data analysis and interpretation: H Xiao, X Teng, C Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jing Cai, PhD. Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China. Email: jing.cai@polyu.edu.hk.

Abstract: Medical image registration is a vital component of many medical procedures, such as image-guided radiotherapy (IGRT), as it allows for more accurate dose-delivery and better management of side effects. Recently, the successful implementation of deep learning (DL) in various fields has prompted many research groups to apply DL to three-dimensional (3D) medical image registration. Several of these efforts have led to promising results. This review summarized the progress made in DL-based 3D image registration over the past 5 years and identify existing challenges and potential avenues for further research. The collected studies were statistically analyzed based on the region of interest (ROI), image modality, supervision method, and registration evaluation metrics. The studies were classified into three categories: deep iterative registration, supervised registration, and unsupervised registration. The studies are thoroughly reviewed and their unique contributions are highlighted. A summary is presented following a review of each category of study, discussing its advantages, challenges, and trends. Finally, the common challenges for all categories are discussed, and potential future research topics are identified.

Keywords: Artificial intelligence; deep learning (DL); image registration; image-guided radiotherapy (IGRT)

Submitted Feb 09, 2021. Accepted for publication Jul 15, 2021.

doi: 10.21037/qims-21-175

View this article at: <https://dx.doi.org/10.21037/qims-21-175>

Introduction

Medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound (US) have been used to aid clinical procedures and diagnoses for decades. In the field of radiation therapy (RT), image guidance using cone-beam CT (CBCT) (1), MRI (2-4), or US (5) facilitates accurate localization of targets during treatment and reduces the irradiation of normal tissues (6,7). These features realize better target

dose delivery and better management of side effects, which has increased the popularity of image-guided radiotherapy (IGRT). Many clinical applications require medical images to be acquired at different time points, by different scanners, and from different patients. After obtaining images across these different modalities, image registration is used to create a fusion image or match medical images to the corresponding method and patient.

In a typical image registration process, a moving image (an image that needs to be moved, also called a source

image) and a fixed image (an image that is used as the template, also called a target image) are first received by an algorithm. The moving image is then moved to match the fixed image, based on the parameters determined by the algorithm. Medical image registration has been studied for decades and remains an actively developing field (8). Medical image registration can be categorized based on different components of the process, as mono-model and multi-model registration (based on the type of input), as intra-patient and inter-patient registration (based on the objects to be registered), as rigid, affine, and deformable image registration (DIR) (based on the type of deformation), as three-dimensional (3D)-3D, 3D-two dimensional (2D), and 2D-2D registration (based on spatial dimensions), or as brain registration, lung registration, prostate registration, etc. [based on the region of interest (ROI)].

As image registration has become more popular and easier to use, it has been applied to various scenarios in IGRT, including target motion tracking (9-11), organ segmentation (12), and adaptive radiotherapy (13,14). However, the demand for more accurate and efficient registration has not abated and remains a priority for clinical applications. In conventional registration algorithms, deformation fields are obtained by iteratively optimizing objective functions. This method usually takes a long time and limits the applications of image registration in clinical settings. Deep learning (DL)-based models can advance the iterative optimization process to the training stage and yield the desired results with the implementation of a single forward computation. Therefore, DL-based models hold the promise of improving clinical tasks and meeting their requirements of high efficiency.

The applications of DL to medical images have been extensively studied and the field has witnessed significant levels of research. There are many systematic reviews on DL-based medical image analysis (15-24), but few have focused on the role of DL in 3D medical image registration. The rapid development of DL models and hardware has made DL-based registration more accurate and allowed the input shift from 2D slices to 3D volumes. This review aimed to summarize the progress made in DL-based medical image registration over the past 5 years and identify existing challenges and future trends.

We present a brief background of DL in Section “DL”. Research studies on DL-based 3D medical image registration published in the last 5 years (2017–2021) are surveyed and the statistical analyses are described in Section “Statistical analysis”. The details of studies from the three

categories, which comprise studies grouped in terms of their supervision methods, i.e., deep iterative registration, supervised registration, and unsupervised registration, are discussed in Sections “Deep iterative registration models”, “Supervised registration models”, and “Unsupervised registration models”, respectively. Finally, existing challenges and future research opportunities are discussed in Section “Discussion”.

DL

DL is an important part of machine learning and DL models are characterized by their large number of layers and parameters. Currently, typical DL models contain more than 100 layers and millions of parameters (25), enabling them to learn complex textures and make accurate predictions. DL models generally have sequential architectures (26), the first several layers of which usually learn simple features, such as edges. Subsequent layers combine these extracted features for advanced object detection. This property makes DL a suitable tool for image-related tasks, such as computer-aided diagnosis (27,28), image enhancement (29), image synthesis (30,31), and functional information derivation (32,33).

Convolutional neural network (CNNs)

DL methods have been widely used in computer vision, with CNNs being one of the most successful types of models. Unlike ANNs, which process the whole image at once, CNNs reduce the number of model parameters and the computational cost by utilizing shift-invariant filters across multiple layers. The “encoder-decoder” architecture is a popular choice for CNNs used for medical imaging applications, as the outputs, such as a 2D slice, a 3D volume, or a dense displacement-vector field (DVF), are of higher dimensions. U-Net is a successful example of this architecture and has been used in various studies (34,35). For registration tasks, CNNs usually receive the moving and fixed images as inputs and produce transformation parameters and moved images as outputs.

Generative adversarial network (GAN)

GANs form an important category of DL models (36). They usually consist of two competing networks, namely a generator and a discriminator. During training, the generator is trained to generate artificial data to fool the discriminator,

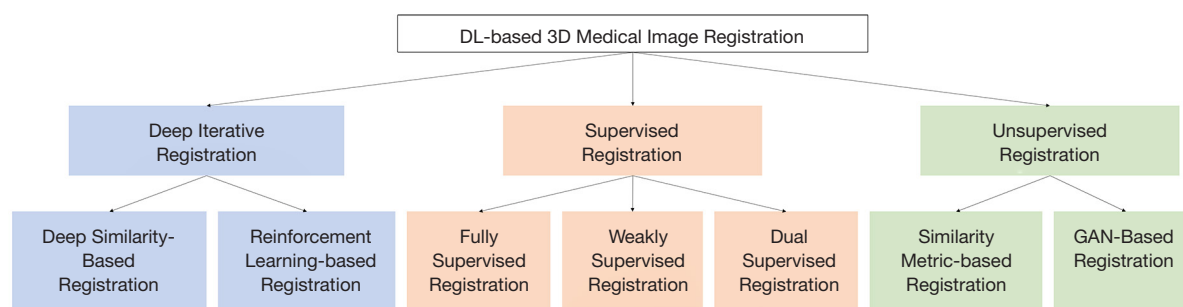


Figure 1 An overview of DL-based 3D medical image registration methods. DL, deep learning; 3D, three-dimensional; GAN, generative adversarial network.

whereas the discriminator is trained to distinguish real data from artificial data. GAN-based models have been successfully used in medical imaging applications, such as image synthesis and translation (30), super-resolution (37), and registration (38,39). For medical image registration, GAN-based models are usually used to provide additional regularization or to translate multi-modality registration to mono-modality registration via image synthesis.

DL methods in 3D medical image registration

Based on the type of DL methods and the training regime, DL-based 3D medical image registration can be classified into three categories: deep iterative registration, supervised registration, and unsupervised registration. An overview of these methods is shown in *Figure 1*. In deep iterative registration, DL models are integrated into conventional iterative registration methods. Here, the intensity-based similarity metrics are usually replaced by deep similarity metrics. Based on which DL methods are applied, deep iterative registration can be further classified as deep similarity-based registration or reinforcement learning (RL)-based registration. In supervised registration, DL models are trained with either labeled reference data or different types of reference data. Based on which type of reference data are used, these models can be further classified as fully supervised registration, weakly supervised registration, or dual supervised registration. In unsupervised registration, the DL models can be further classified as similarity metric-based registration, in which the models are optimized based on the similarity losses between warped and fixed images, or GAN-based registration, in which the generator and discriminator compete and are optimized in an adversarial manner.

Statistical analysis

To include as many relevant studies as possible, various keywords were used in our search. The keywords included but were not limited to DL, machine learning, image registration, image fusion, motion estimation, and deep similarity. The publications were obtained from Google Scholar, PubMed, Web of Science, and arXiv. Papers of poor methods and validations were excluded, including but not limited to insufficient subjects in training or test set ($n < 10$) and unclear implementation description. After our search, we included 68 studies that were closely related to DL-based 3D medical image registration. *Figure 2* describes the results of the statistical analysis of these studies.

The studies we selected investigated several types of ROIs. As shown in *Figure 2A*, the brain, lung, and prostate are the three most popular ROIs in DL-based registration studies and account for 40%, 24%, and 10% of all studies, respectively. With respect to image modality, both multi-modality and mono-modality registration have been studied using DL-based registration. As shown in *Figure 2B*, MR-MR and CT-CT registration are the most studied modalities. Of the three categories of DL-based methods, there are significantly higher numbers of supervised and unsupervised registration studies than of deep iterative registration studies. The corresponding percentages are shown in *Figure 2C*.

Although the various DL models can differ in terms of architecture and training strategies, they are similar in that their performance improves as the size of the training dataset increases. Compared to self-collected datasets, training and evaluating DL models on multiple public datasets can also demonstrate their generalizability and help compare their

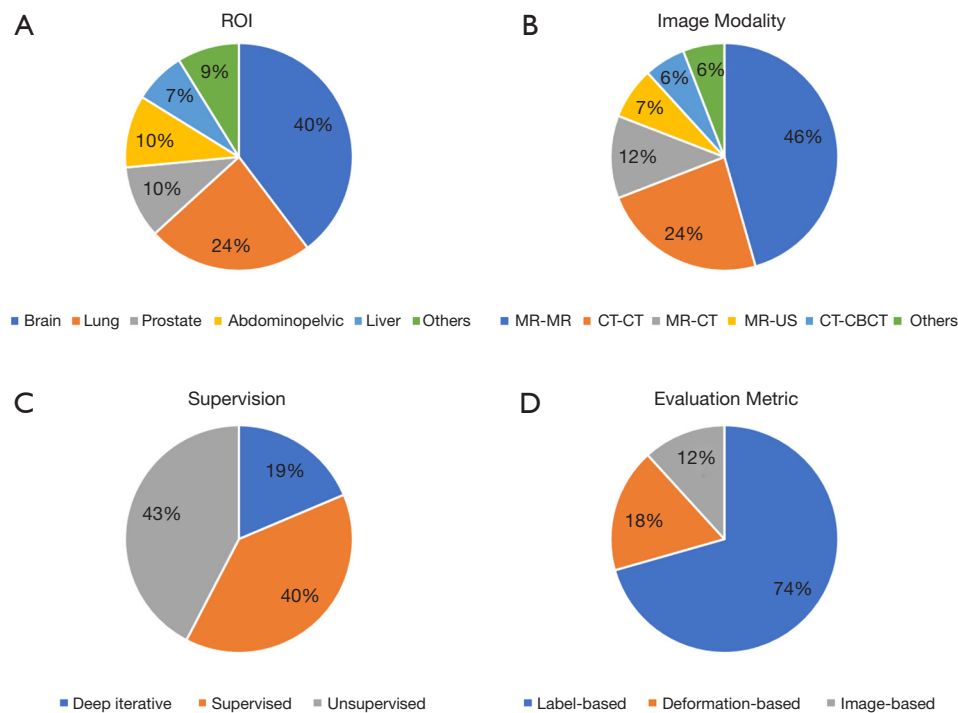


Figure 2 Statistical analysis of selected publications. CBCT, cone-beam computed tomography; CT, computed tomography; MR, magnetic resonance; US, ultrasound; ROI, region of interest.

accuracy on fair benchmarks. In addition, public datasets can also make the studies easier to be reproduced and adapted to new scenarios. The public datasets that have been used in DL-based registration studies are summarized in *Table 1*, and the datasets utilized by. Brain MR and lung CT images constitute the majority of the datasets and datasets with annotated images are more frequently used. For example, the most used datasets for brain (LONI LPBA40) and lung (DIR-Lab) images provide brain MR images with segmentation labels and four dimensional (4D)-CT images with 300 surrogates, respectively.

Registration evaluation is another challenge of using DL models. The commonly used evaluation metrics can be classified into three categories: image-based metrics, label-based metrics, and deformation-based metrics. Image-based metrics are often used when no labels are available and focus either on absolute values or the distribution of voxel intensities. The commonly used metrics are mean absolute error (MAE), mean square error (MSE), cross-correlation (CC), normalized cross-correlation (NCC), mutual information (MI), feature similarity index metric

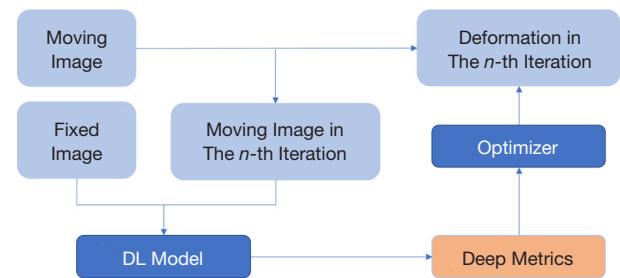
(FSIM), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). When labels are provided, label-based metrics that focus on the differences between warped and ground truth labels can be used. For organ segmentation masks, Dice coefficient (DSC), average symmetric surface distance (ASSD), surface registration error (SRE), Hausdorff distance (HD), and mean surface distance (MSD) are commonly used. For point-wise surrogates, target registration error (TRE) and fiducial registration error (FRE) are the most popular metrics. The metrics of the first two categories measure the accuracy of image matching, whereas deformation-based metrics are used to measure the accuracy and plausibility of the deformation. For example, the Euclidean distance measures the numerical difference between predicted and reference deformations, the Jacobian determinant (Jaco. Det.) of DVFs is used to quantify the singularity of the deformation field and bending energy (BE) is used to measure its smoothness. *Figure 2D* shows the corresponding percentage values for each category; it can be seen that label-based metrics are most used to evaluate image registration.

Table 1 Public datasets used in deep learning-based three-dimensional medical image registration

ROI	Dataset	Modality
Brain	OASIS (40)	MRI
	LONI LPBA40, IBSR18, CUMC12, MGH10 (41)	MRI
	MindBoggle101 (42)	MRI
	IXI [†]	MRI
	ABIDE (43)	MRI
	ADHD (44)	MRI
	MCIC (45)	MRI
	PPMI (46)	MRI
	HABS (47)	MRI
	Harvard GSP (48)	MRI
	FreeSurfer Buckner40 (49)	MRI
	BraTS18 (50)	MRI
	ANDI (51)	MRI
	ALBERTs (52)	MRI
	Simulated Brain(53)	MRI
	RESECT(54)	MRI, US
Knee	OAI [‡]	MRI
Lung	POPI(55)	CT
	DIR-Lab (56-58)	CT
	VISCERAL (59)	CT
	SPREAD (60)	CT
	SPARE (61)	CT
Abdomen	KITS19 (62)	CT
	Medical Segmentation Decathlon (63)	CT
	Pancreas-CT (64)	CT
Spine	SpineWeb library (65)	MRI, CT
Prostate	Prostate Fused-MRI-Pathology(66)	-
	Prostate-3T (67)	MRI
	PROMISE12 (68)	MRI

[†], available online at <http://brain-development.org/ixi-dataset/>;

[‡], available online at <https://nda.nih.gov/oai/>. ROI, region of interest; CT, computed tomography; MRI, magnetic resonance imaging.

**Figure 3** General workflow for deep similarity-based DL models. DL, deep learning.

Deep iterative registration models

Deep similarity-based registration

The past few decades have witnessed the development of several conventional iterative registration methods and toolboxes, including optical flow (69), demons (70), Elastix (71), advanced normalization tools (ANTs) (72), and the hierarchical attribute matching mechanism for elastic registration (HAMMER) (73). In most cases, a cost function is designed to measure the similarity between the warped image I_n in the n -th iteration and the fixed image J . The general mathematical formula for the cost function \mathcal{L} is:

$$\mathcal{L} = \text{Sim}(I_n, J) + \text{Reg}(\mathcal{T}_n) \quad [1]$$

where $\text{Reg}(\mathcal{T}_n)$ regularizes the transform \mathcal{T}_n in the n -th iteration for plausibility. The commonly used similarity metrics, such as MAE, MSE, CC, NCC, and MI, are usually intensity-based (74,75). In general, they perform well when applied to mono-modality registration. However, these metrics primarily focus on voxel intensity values [except HAMMER (73)] and are sensitive to image artefacts. Moreover, they treat all components of the image equally and therefore may not capture the most effective features. The optimal metrics for registration vary across modalities and the choice of metric relies on experience, which may introduce human errors. CNNs have shown great promise in image recognition and segmentation and are therefore more likely to capture underlying features. To make full use of the feature extraction capability of CNNs, some research groups have replaced the intensity-based similarity metrics with CNN-based metrics and achieved high registration performance. The general workflow of deep similarity-based registration is shown in Figure 3. An overview of these studies is described in Table 2.

Table 2 Overview of deep iterative registration studies

Study	Learning	Transform	Modality	ROI	Public dataset for training/validation	Evaluation metrics
(76)	Metric	Deformable	MRI-MRI	Brain	LONI LPBA40	DSC
(77)	Metric	Deformable	MRI-MRI	Brain	IXI, ALBERTs	DSC
(78)	Metric	Deformable	MRI-MRI	Brain	IXI	FRE, DSC
(79)	Metric	Rigid	MRI-US	Prostate	PROMISE12, Prostate-3T	TRE
(80)	Metric	Deformable	MRI-MRI	Brain	OASIS, ABIDE	DSC
(81)	Metric	Rigid & Deformable	MRI-MRI	Brain	Simulated brain	DSC
(82)	Metric	Deformable	MRI-MRI	Brain	LPBA40, CUMC12, MGH10, IBSR18	DSC, Jaco. Det.
(83)	RL agent	Rigid	CT-CBCT	Spine and heart	N/A	TRE
(84)	RL agent	Deformable	CT-CT	Thorax, abdomen, and pelvis	N/A	Euclidean distance
(85)	RL agent	Deformable	MRI-MRI	Prostate	PROMISE12, Prostate-3T	DSC, HD
(86)	RL agent	Affine	MR-CT	Brain	N/A	TRE

CBCT, cone-beam computed tomography; CT, computed tomography; DSC, Dice coefficient; FRE, fiducial registration error; HD, Hausdorff distance; MRI, magnetic resonance imaging; RL, reinforcement learning; ROI, region of interest; TRE, target registration error.

Applications

Wu *et al.* developed stacked autoencoders (SAEs) to learn discriminative features from input images and quantify their similarities (76). Given that SAEs can learn intrinsic image features, conventional algorithms integrated with SAEs have consistently shown improved registration accuracy across several datasets. Simonovsky *et al.* modeled the registration into a classification task, in which a CNN was used to identify whether the input image pairs were well aligned (77). Then, they replaced the MI in a conventional registration algorithm with the CNN and observed that the resulting algorithm gave significantly better T1–T2 brain MR image registration. Sedghi *et al.* adopted a similar idea and further developed this method for groupwise registration (78). Their deep metric performed well on difficult registration cases that the traditional MI metric had previously failed to manage.

In addition to discrete classification, deep metrics can also assign continuous values of physical meanings. For example, Haskins *et al.* used a deep CNN to learn a similarity metric for registration between MR and transrectal US (TRUS) images (79). The network took the MR-TRUS image pairs as inputs and estimated TREs to evaluate the registration. The training was supervised using the difference between the ground truth and the estimated TREs. The CNN outperformed classical MI-based and state-of-the-art modality-independent neighborhood descriptor (MIND) feature-based registration methods

with smaller TREs. Czolbe *et al.* used a pre-trained segmentation network to extract image features, and used the differences in image features between the moving and warped images as the cost function (80). They reported higher registration accuracy and faster convergence speeds compared to conventional methods. So *et al.* utilized a learning-based metric called Bhattacharyya Distances for both rigid and DIR and showed superior performance to MI (81).

Aside from the similarity term in Eq. [1], the regularization term can also be optimized via training. For example, Niethammer *et al.* integrated a shallow CNN for spatial adaptive regularization (82). Conventional transform regularization penalizes the spatial gradient of the deformation field for smoothness, which is shift-invariant and may over-smoothen regions with sharp changes. The integrated CNN received a precalculated DVF with corresponding images and outputted a locally smoothed DVF that was supervised by positive semidefinite matrices. This method improved the structure overlap after registration and reduced the number of negative points in the Jaco. Det. of the DVFs compared to the precalculated methods.

Summary

Deep similarity-based methods have shown great potential in obtaining image- and purpose-specific metrics. These novel metrics are better suited to registration tasks and have outperformed conventional intensity-based similarity

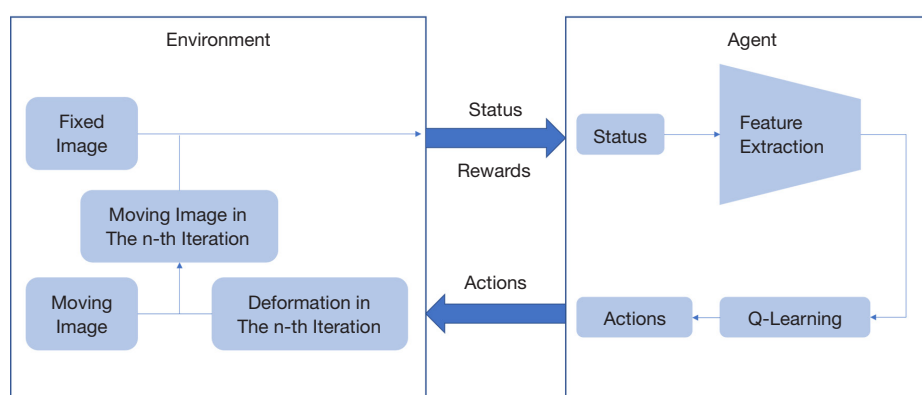


Figure 4 General workflow of reinforcement learning-based registration.

metrics. However, the CNNs used for obtaining deep similarity metrics need to be trained separately and the ground truth data used for training is difficult to obtain. Although the registration process developed by Czolbe *et al.* did not require ground truth labels, their segmentation network nevertheless required labeled images for training (80). Another limitation is that deep similarity metrics are difficult to interpret and validate, and errors can accumulate due to insufficient training. Finally, the implementation of deep similarity-based registration is most severely limited by the iterative process that is used by conventional registration methods. As more studies demonstrate the feasibility of one-shot registration using DL, deep similarity methods may be less attractive in the future. Compared to previous years (15,16), the number of studies in this category has decreased and this trend is expected to continue.

RL-based registration

RL is a subfield of machine learning. In an RL framework, an intelligent agent performs a sequence of actions in an environment designed to maximize rewards via successive trials and errors. RL has been widely studied in several decision-making tasks, such as robotic control, stock market trading, and recommendation systems. Recently, the combination of RL with DL, known as deep reinforcement learning (DRL), has been applied to image registration. As shown in Figure 4, RL-based registration is an iterative observer-action process that runs in a reward-driven system. The target and moving-image pair are constructed in a specified environment, and an artificial agent is trained by interacting with this environment to perform sequential

alignments (i.e., registration parameters).

Applications

Liao *et al.* were the first to use an RL framework to perform 3D rigid-body image registration (83). In this approach, a neural network-based agent is designed to predict sequential movements (i.e., ± 1 mm for translations or $\pm 1^\circ$ for rotations) for image alignment. The AI agent is trained in a greedy deep-supervised learning (DSL) fashion, which precludes the exploration history of the agent to improve training efficiency. In a similar design, Ma *et al.* utilized a deep Q-learning framework to extract image contextual features for rigid registration (84). In contrast to the above-mentioned rigid registration approach, Krebs *et al.* applied RL to deformable registration, and used a statistical deformation model to restrict the dimensionality of the action space (85). Recently, Hu *et al.* applied asynchronous RL to 2D affine registration (86). They incorporated a convolutional long-short-term-memory (conLSTM) module into the RL framework to extract spatiotemporal image features.

Summary

Recent studies have demonstrated that RL has immense potential in image registration applications. For certain tasks, RL-based methods can achieve similar or higher registration accuracies compared to other registration methods. A major challenge faced by RL-based registration is the long training time that arises due to difficulties in convergence. For example, Ma *et al.* spent 4 days in training (84) and Hu *et al.* spent approximately 13 hours, even with accelerated training (86). Very few studies have used RL for image registration. Improvements in computational power are expected to accelerate the development of RL-based

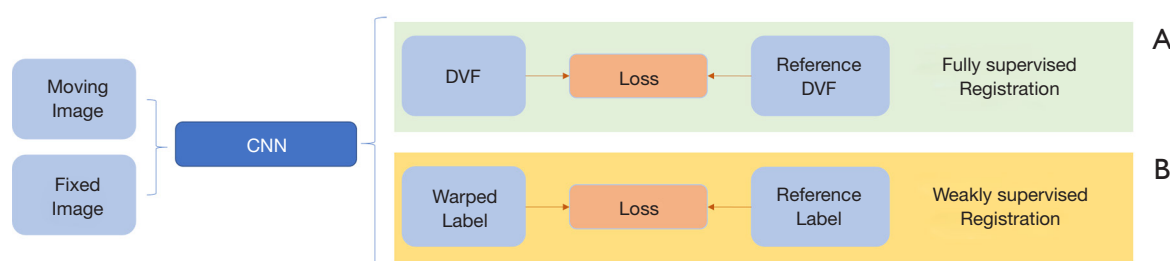


Figure 5 General workflow of (A) fully supervised registration and (B) weakly supervised registration. CNN, convolutional neural network; DVF, displacement vector field.

approaches for image registration.

Supervised registration models

Supervised training for DL models is a straightforward idea upon which many registration models are based. Based on the type of supervision used during model training, these models can be divided into three sub-categories: fully supervised, weakly supervised, and dual supervised registration. In fully supervised registration, ground truth DVFs from conventional registration algorithms are used to supervise the training. Here, the loss function is usually the difference between the ground truth and the predicted DVFs, as shown in *Figure 5A*. In the weakly supervised registration, certain indirect reference labels are used for training instead of reference DVFs, which are the most commonly anatomical contours, as shown in *Figure 5B*. In dual supervised registration, more than two kinds of reference data are used for training, typically consisting of a combination of image similarity, reference DVFs, and anatomical structure contours. An overview of supervised registration models was shown in *Table 3*.

Applications

Fully supervised registration

A few of the studies we selected predicted the motions at patch centers for convenience and obtained dense DVFs via interpolation. For example, Cao *et al.* used a CNN model with three neurons in the output layer, with each neuron representing the motion amplitude along the x, y and z directions, respectively, at the center of a small patch (87,109). They achieved accuracies that were comparable to conventional algorithms. Teng *et al.* built a patch-based CNN to perform inter-phase registration of lung 4D-CT and 4D-CBCT (88). The inputs were moving and target

patch pairs and the deformable vectors in the centers of the moving patches were predicted. Three evaluation metrics, CC, MSE, and SSIM, were used to evaluate registration performance in the diaphragm region. Other studies have utilized CNNs to directly predict dense DVFs. For example, Yang *et al.* used a U-Net-like CNN model to learn DVFs of the same resolution as that of input brain MR images, and achieved high registration accuracy across multiple datasets (89). Rohé *et al.* used a similar network architecture for cardiac MR image registration, and obtained comparable results with conventional methods in terms of contour overlap (90). Wang *et al.* introduced a framework to tune the regularization parameter for smoothness of diffeomorphic transformations in brain MR images (91). They built a CNN predictive model to learn the regularization parameters from pairwise image registration. Their network predicted appropriate regularization parameters in a time-efficient and memory-saving manner.

To overcome the shortage of training data, some studies on supervised registration used artificial DVFs to supervise DL model training. Artificially created DVFs would alleviate the cost of collecting densely labeled data and supervise the training in voxel level. Sokooti *et al.* used a fully supervised DIR method for registering lung CT images, in which the reference DVFs were artificially created by combining different spatial frequencies to mimic both large and small motions (92,93). This method achieved a reasonable registration accuracy across multiple datasets. There are also other methods by which artificial DVFs can be generated. For example, Eppenhof *et al.* sampled random numbers from a specified range in a coarse-to-fine grid to generate artificial motions (94,95), and Guo *et al.* used an error-scaling method to generate a training dataset with a target distribution (96). The registration accuracy of all of these methods was either comparable to or better than that of conventional algorithms, in terms of TRE and the Dice score.

Table 3 Overview of supervised registration studies

Study	Supervision	Transform	Modality	ROI	Public dataset for training/validation	Evaluation metrics
(39)	Real DVF	Deformable	MRI-CBCT	Head & Neck	N/A	DSC, MSD, HD
(87)	Real DVF	Deformable	MRI-MRI	Brain	LONI LPBA40, ADNI	DSC, ASSD
(88)	Real DVF	Deformable	4D-CT/CBCT	Lung	N/A	CC, MSE, SSIM
(89)	Real DVF	Deformable	MRI-MRI	Brain	LPBA40, IBSR18, MGH10, CUMC12, OASIS	DSC
(90)	Real DVF	Deformable	MRI-MRI	Heart	N/A	DSC, HD, Jaco. Det.
(91)	Real DVF	Deformable	MRI-MRI	Brain	ADNI, OASIS, LPBA40	Regularization parameter difference
(92,93)	Artificial DVF	Deformable	CT-CT	Lung	SPREAD	TRE, Jaco. Det.
(94,95)	Artificial DVF	Deformable	CT-CT	Lung	DIR-Lab, POPI	TRE
(96)	Artificial DVF	Deformable	MRI-TRUS	Prostate	N/A	SRE
(97)	Real DVF	Deformable	MRI-TRUS	Prostate	N/A	DSC, MSD, HD
(98)	Contours	Deformable	MRI-MRI	Brain	N/A	DSC, BE
(99)	Contours	Deformable	MRI-MRI	Brain	N/A	MI, DSC
(100)	Contours	Deformable	MRI-MRI	Brain, Knee	MindBoogle101, OAI	DSC
(101)	Contours	Affine & Deformable	MRI-MRI	Brain	Mindboggle101, LPBA40, IXI	DSC, HD, ASSD
(102)	Contours	Deformable	CT-CT	Abdomen	KITS19, Medical Segmentation Decathlon, Pancreas-CT	DSC
(103)	Contours	Deformable	CT-CT	Lung	DIR-Lab	DSC
(104)	Dual	Deformable	MRI-MRI	Brain	LONI LPBA40	DSC
(105)	Dual	Deformable	MRI-MRI	Brain	LONI LPBA40, IXI	DSC, HD
(106)	Dual	Deformable	CT-CT	Abdomen	VISCERAL	DSC
(107)	Real DVF	Deformable	CT-CT	Lung	N/A	Euclidean distance
(108)	Contours	Deformable	MRI-MRI	Prostate	N/A	DSC, TRE

ASSD, average symmetric surface distance; CBCT, cone-beam computed tomography; CC, cross-correlation; CT, computed tomography; DSC, Dice coefficient; DVF, displacement vector field; FRE, fiducial registration error; HD, Hausdorff distance; Jaco. Det, Jacobian determinant; MI, mutual information; MRI, magnetic resonance imaging; MSD, mean surface distance; MSE, mean square error; ROI, region of interest; RL, reinforcement learning; SRE, surface registration error; SSIM, structural similarity index measure; TRE, target registration error; TRUS, transrectal ultrasound.

In addition to accuracy, reference deformations can also provide biomechanical information and increase the feasibility of DL models. For example, Fu *et al.* designed a registration framework for MR-TRUS prostate image registration (97), in which the ROIs were first segmented and then volumetric point clouds were generated from the segmentation using tetrahedron meshing. Reference deformations were obtained from these point clouds via finite-element modeling with biomechanical constraints. The registration framework demonstrated a promising

registration performance when evaluated using DSC, MSD, HD, and TRE. To demonstrate its generalizability, Fu *et al.* further applied this registration framework to register multi-parametric MR images with CBCT (39). This method performed better in terms of TRE compared to traditional intensity-based rigid registration.

Weakly supervised registration

The idea of including segmentation in registration training has been adopted by several researchers for brain MR-MR

registration. Hu *et al.* used a deformable registration method that used global and local label-driven learning with CNN, and obtained high-accuracy intra-subject and inter-subject registration of brain MR images (98). Li *et al.* and Estienne *et al.* developed hybrid CNNs that achieved both image registration and segmentation within a single framework (99,110). The segmentation similarity aided the registration process and improved the efficiency and the accuracy of registration. Xu *et al.* adopted a different approach (100), in which they introduced the existing segmentation as an input to the network. Balakrishnan *et al.* mainly focused on unsupervised registration, but also provided an option to perform weak supervision using contours (111). All the above-described networks achieved promising registration accuracy across various datasets. Moreover, aside from purely deformable registration, Zhu *et al.* used a registration scheme that combined both affine and deformable MR-MR brain registration methods (101). The affine network used global similarity as the loss function, whereas the deformable network used local similarity. In addition, overall anatomical similarity was used to supervise the training of the registration network. This method outperformed state-of-art methods when evaluated using different metrics.

Weakly supervised registration has also been successfully implemented for CT images. Estienne *et al.* used a registration method that applied spatial gradients and noisy segmentation labels to abdominal CT-CT registration (102). They developed a symmetrical formulation that predicted transformations from source to target and from target to source. They also integrated various publicly available datasets into the training process. Hering *et al.* used a multilevel variational image-registration network to perform large-scale CT-CT lung registration (103). Their multi-level approach was able to achieve significantly better registration results than conventional methods.

Dual supervised registration

Although reference DVFs facilitate efficient supervision, they are not foolproof. DL models will never surpass conventional methods if they are trained only with reference DVFs. To compensate for imperfect DVFs, Fan *et al.* designed a fully convolutional network with dual guidance to register brain MR images (104). The network was supervised using both the Euclidean distance between predicted and reference DVFs and the MSE between the warped and fixed image. Unlike classical U-Net models, their model was equipped with gap filling and hierarchical loss capabilities

to improve performance. They also implemented multi-source strategies to augment the training data. Compared to state-of-the-art methods, their method showed promising registration accuracy and efficiency on a variety of datasets. Ahmad *et al.* demonstrated a two-step process to register brain MR images (105). In their method, before implementing DL training in a manner similar to that used by Fan *et al.* (104), the input images were represented using a graph and clustered through iterative graph coarsening. This deformation initialization enabled groupwise registration to converge significantly faster and with a competitive level of accuracy compared to conventional methods, thereby facilitating large-scale image studies. Ha *et al.* developed a concept for large-scale image deformation of abdominal CTs using supervised learning (106). The network architecture was designed to predict discrete heatmaps for the relative displacement between two scans using graph and deformable-field convolutions. The MSE of DVFs and the smoothness of transformation were used to estimate the sparse displacement between two scans. This method showed a clear improvement in accuracy compared to state-of-the-art DL approaches for abdominal CTs.

Summary

The fully supervised registration method has been successfully used on various ROIs. The reference deformations from well-developed conventional registration algorithms have allowed certain challenging registrations to be achieved, including multi-modality and large motion registration, with performance efficiencies comparable to those of conventional methods. Reference DVFs play an important role in fully supervised registration, and special constraints can be learned using specifically prepared training samples, such as biomechanical constraints. Despite the presence of these advantages, the shortage of training data severely limits the applications of fully supervised registration. This problem can be addressed using artificial DVFs and data augmentation strategies. These two methods can track the exact motions between moving and fixed images and eliminate the uncertainties introduced by imperfect reference deformations. This is particularly applicable to multi-modality registration, for which reference deformations are less likely to be accurate.

However, both these strategies may fail to reflect true physiological motions. Therefore, concerted efforts are required to generate more realistic training samples. Weakly supervised registration is another important sub-

category. Unlike reference DVFs, which are difficult to obtain and can be potentially imperfect, organ contours are more available and reliable. This makes weakly supervised training easier to conduct. The supervision provided by contours could also solve difficult registration problems, such as noisy labels and large movements (102,103). Dual supervision combines the advantages of the above-described supervision methods and, as expected, outperforms both methods. However, the sample preparation for dual supervision is more laborious.

In summary, supervised registration has been successfully implemented and is particularly suitable for scenarios in which non-differentiable biomechanical constraints are considered. In light of its promise, we expect that supervised registration will be developed further in the future.

Unsupervised registration models

Although many methods (including data augmentation and weak supervision) have used to address the data shortage problem of supervised registration, the preparation of training samples remains a time-consuming process. Thus, it is more convenient to implement unsupervised registration, wherein the input only consists of moving and fixed image pairs from which the DL model learns the deformation. An overview of this category is shown in *Table 4*. This category nevertheless requires a loss function in training that is very similar to the loss function used in conventional iterative registration. As shown in Eq. [1], the loss function usually comprises an image similarity term and a DVF regularization term. Specifically, due to the nature of intrinsic convolution, a few similarity metrics such as localized NCC (LNCC) are modified to focus on small patches. Additionally, special loss terms can be added to perform adjustments, such as cycle-consistency loss to reduce singularity and identity loss to avoid over-fitting. The general workflow of similarity-based unsupervised registration is shown in *Figure 6A*. GAN-based unsupervised registration is a special sub-category of this method. Here, instead of intensity-based metrics, a discriminator quantifies the similarity between warped and fixed images, as shown in *Figure 6B*. This method is similar to deep similarity-based registration, which focuses not only on superficial voxel intensities but also on underlying textures and information.

Similarity metric-based registration

Balakrishnan *et al.* used an encoder-decoder-like network

named VoxelMorph to perform DIR of brain MR images (111). The network predicted dense DVFs with paired brain MR images as inputs. The network training was supervised using an intensity-based metric, which was either MSE or NCC depending on image modalities in implementation. The plausibility of the DVFs was regularized using their spatial gradient. Similarly, Estienne *et al.* also developed an encoder-decoder-like network named U-ReSNet for brain MR image registration (110). Their network was novel in that segmentation was achieved by extracting image features using a shared encoder that was also used for registration and reconstructing anatomical labels using a separate decoder. These segmentation results guided training and improved registration accuracy.

To better regularize DVFs, some researchers modified the training strategy to a cycle-consistent way. To achieve this, they processed the warped image through the network and transformed it back to the moving image (112,113,134). This strategy reduced the number of negative values in the Jaco. Det. and allowed more plausible DVFs to be obtained. Kim *et al.* further added an identity loss term in which identical images were inputs and any deformation was penalized (113).

Registration across multiple stages and the integration of affine registration are popular methods by which researchers account for large motions. de Vos *et al.* implemented both affine and deformable registration in a multi-resolution and multi-level manner (114). They down-sampled source images in multiple stages such that both large and small motions could be captured. Other studies utilized a similar coarse-to-fine strategy for lung images and achieved high registration accuracies (38,115,116). Shen *et al.* and Shao *et al.* also integrated affine registration into their DL models and successfully addressed large deformations in prostate and knee images (118,119).

In addition to simple intensity-based metrics, pre-trained CNNs can be used to extract image features and quantify their similarities. Some researchers have used separate networks to learn deep metrics for better registration, which is a process termed perceptual loss. For example, Duan *et al.* used a spatial weighting-based metric network to learn the deep similarity between CT and CBCT, and demonstrated a high capability of tolerating CBCT artefacts (120).

GAN-based registration

Yan *et al.* applied a GAN-based DL model to MR-US image registration (121), which for a long time had

Table 4 Overview of unsupervised registration studies

Study	Similarity loss	Transform	GAN-based	Modality	ROI	Public dataset for training/validation	Evaluation metrics
(38)	NCC	Deformable	Yes	CT-CT	Lung	DIR-Lab	TRE
(109)	NCC	Deformable	No	CT-MRI	Prostate	N/A	DSC, ASSD
(97)	CC	Deformable	No	PET-PET	Chest	N/A	MSE
(110)	MSE/NCC	Affine & Deformable	No	MRI-MRI	Brain	OASIS	DSC
(111)	MSE/LNCC	Deformable	No	MRI-MRI	Brain	OASIS, ABIDE, ADHD200, MCIC, PPMI, HABS, Harvard GSP, FreeSurfer Buckner40	DSC
(112)	CC	Deformable	No	MRI-MRI	Brain	MindBoggle101	DSC, Jaco. Det.
(113)	CC	Deformable	No	CT-CT	Liver	N/A	TRE, Jaco. Det.
(114)	NCC	Affine & Deformable	No	CT-CT	Lung	DIR-Lab	DSC, TRE, HD, ASSD, Jaco. Det.
(115)	NCC	Deformable	No	CT-CT	Lung	SPARE	TRE
(116,117)	N/A	Deformable	Yes	CT-CT	Lung	N/A	NCC, MAE, PSNR, TRE
(118)	MSE	Affine & Deformable	No	MR-Histology	Prostate	Prostate Fused-MRI-Pathology	DSC, HD, TRE
(119)	NCC	Affine & Deformable	No	MRI-MRI	Knee	N/A	DSC
(120)	Perceptual	Deformable	No	CT-CBCT	Lung	N/A	SSIM, TRE, DSC, ASSD
(121)	N/A	Affine	Yes	MR-US	Prostate	N/A	DSC, TRE
(122)	NCC	Deformable	Yes	CT-CT	Prostate	N/A	ASSD, HD
(123,124)	N/A	Deformable	Yes	MRI-MRI	Brain	LPBA40, IBSR18, CUMC12, MGH10	DSC, ASSD
				MRI-CT	Pelvic	N/A	
(125)	LNCC	Deformable	No	MRI-MRI	Brain	LONI LPBA40, IBSR18, CUMC12, MGH10	DSC, HD, Jaco. Det.
(126)	RMSE	Deformable	No	CT-CBCT	Head & Neck	N/A	MSE, MI, FSIM
(127)	NCC	Deformable	No	MRI-MRI	Brain	ADNI, LPBA40	DSC
(128)	MSE	Affine & Deformable	No	MRI-MRI	Lung	N/A	DSC, TRE
(129)	MSE	Deformable	No	MRI-US	Brain	RESECT	TRE
(130)	CC	Affine & Deformable	No	CT-CT	Liver	N/A	DSC, Jaco. Det.
				MRI-MRI	Brain	N/A	
(131)	NCC	Deformable	No	PET-CT	Body	N/A	NCC
(132)	SSD	Deformable	No	MRI-MRI	Brain	IXI, ANDI	SSIM, PSNR, SSD
(133)	MSE	Deformable	No	CT-CT	Lung	DIR-Lab	TRE

ASSD, average symmetric surface distance; CBCT, cone-beam computed tomography; CC, cross-correlation; CT, computed tomography; DSC, Dice coefficient; FRE, fiducial registration error; FSIM, feature similarity index metric; GAN, generative adversarial network; HD, Hausdorff distance; Jaco. Det., Jacobian determinant; LNCC, localized normalized cross-correlation; MAE, mean absolute error; MI, mutual information; MRI, magnetic resonance imaging; MSD, mean surface distance; MSE, mean square error; NCC, normalized cross-correlation; PET, positron emission tomography; PSNR, peak signal to noise ratio; RMSE, root mean square error; ROI, region of interest; RL, reinforcement learning; SRE, surface registration error; SSD, symmetric surface distance; SSIM, structural similarity index measure; TRE, target registration error; TRUS, transrectal ultrasound; US, ultrasound.

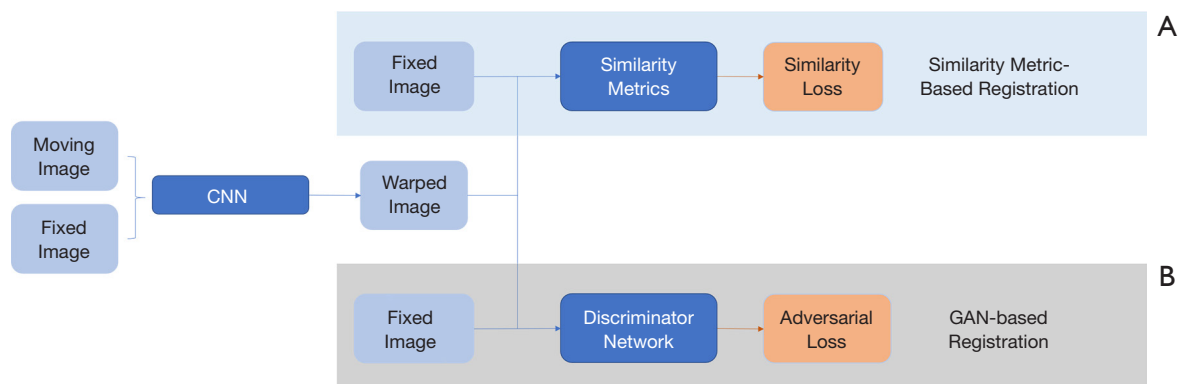


Figure 6 General workflow of (A) similarity metric-based registration and (B) GAN-based registration. CNN, convolutional neural network; GAN, generative adversarial network.

been regarded as a challenge because of the significant differences in image appearance and the large variations in image correspondence. They used a generator to predict the affine transformation parameters and a discriminator to distinguish warped and fixed images. Using this method, they obtained significantly higher DSCs and lower TREs after registration compared to conventional methods. Elmahdy *et al.* demonstrated the feasibility of using a shallow discriminator to perform joint prostate CT DIR and segmentation (122). Other researchers cropped the images to obtain small patches and integrated dilated convolutional layers into generators and discriminators to capture motions at multiple scales (38,116). Fan *et al.* further developed the discriminator such that it received a pair of images rather than a single image (123). In addition, they defined a positive case, i.e., a well-aligned image pair, as a fixed image and a combination of moving and fixed images to loosen the impractical requirements of perfect matching.

Deformations are rarely evenly distributed along the body, and therefore focusing on areas prone to larger deformations can improve the registration accuracy. Fu *et al.* and Lei *et al.* included attention modules in their DL models to assign higher weights to regions with large motions (38,117). Furthermore, Huang *et al.* cropped images into small patches and classified them into easy-to-register and hard-to-register patches based on the attention amplitude of each patch (125). The hard-to-register patches were further refined after classification.

Summary

Compared to supervised registration, unsupervised

registration is preferred, due to its convenient training process. Consequently, the number of publications on unsupervised registration has grown rapidly in recent years (15,16). Several groups have achieved accuracies that are either comparable to or higher than those of conventional algorithms (38,115). However, most studies on unsupervised registration focus on mono-modality registration, and thus unsupervised multi-modality registration warrants further investigation in the future. Unsupervised registration is also generally more challenging than supervised registration as no reference DVFs or anatomical contours are provided. Therefore, in addition to common network training, complex pre-processing procedures are required to achieve accurate registration. For example, Guo *et al.* (134) applied rigid registration before using the DL model to reduce motion amplitudes, whereas other groups applied binary masks to focus on ROIs (38,115). Fu *et al.* segmented and increased the intensity of pulmonary vessels by a factor of 1,000 to enrich the image details (38). Such pre-processing improves registration accuracy but complicates the training process and can affect model generalizability. In summary, unsupervised registration methods have easy training protocols and offer promising results. Accordingly, we expect research interest in this sub-category to continue growing.

Discussion

Although image registration has been studied for decades, the emergence of DL and its application have rejuvenated the field. The rapid development of DL models and hardware has allowed the applications of DL-based 3D medical image registration to range from mono-modality to

multi-modality registration tasks, from patch-based to whole image-based registration tasks, and from label-supervised to label-free registration tasks. DL models are capable of rapidly completing registration in a single forward calculation, within seconds or even milliseconds (111,114). In addition to the high computational efficiency, DL models also show comparable accuracy with conventional methods. For example, on the public benchmark DIR-Lab, Fu *et al.* (38) achieved a TRE of 1.59 ± 1.58 mm, which is very close to the lowest TRE of 0.92 ± 0.15 mm in conventional methods (135).

In this review, we selected 68 studies on DL-based 3D medical image registration that were published over the past 5 years. The studies were classified into three categories; for convenience, studies that belonged to multiple categories were listed in only one category. For example, the networks used for deep iterative registration require training to learn specific similarity metrics, and this training can be either supervised or unsupervised. Although both supervised and unsupervised methods could technically be combined into a third sub-category of dual supervision, for the purposes of this review we classified dual supervision as a sub-category of supervised registration. With respect to the registration accuracy of the three categories, several studies on deep iterative registration reported better performance than conventional methods (76,77), whereas supervised or unsupervised registration was either only comparable to or slightly better than state-of-the-art conventional registration methods. However, both supervised and unsupervised registrations omit the iterative process and are thus preferred due to their high efficiencies. Duan *et al.* integrated the deep similarity metric into an unsupervised DL model and obtained promising results (120). In future, the advantages of all of these categories may be combined, to better balance requirements for accuracy and efficiency.

Given the recent increases in GPU memory, several groups have adopted whole image-based DL model training, whereas others have opted to retain patch-based training. Both strategies have unique advantages and disadvantages. Whole image-based training has large receptive fields and is more capable of capturing large motions. It does not require patch processing, thereby reducing computation time (123). However, whole image-based DL models usually comprise down-sampling layers to save GPU memory, which may cause information loss and compromise registration accuracy. In addition, whole image-based training usually suffers from data shortage. In contrast, patch-based training does not face this problem, as sufficient patches can be

sampled from a single image volume. Moreover, DL models that undergo patch-based training can be deeper, and due to their smaller input size have fewer down-sampling layers, thereby improving local performance.

A key challenge of patch-based training is patch fusion before obtaining the final deformation. An overly large stride between patches can cause discontinuous deformation, whereas large overlaps between patches can significantly increase the computational cost. Some research groups have adopted multi-scale registration to combine the advantages of whole image-based and patch-based training and achieved promising results (115,117). Despite this, more rigorous studies are needed to validate this multi-scale strategy across different ROIs and image modalities.

Irrespective of the training strategy, the loss function remains the core of DL model training. This is especially true for unsupervised registration. Almost all loss functions are combinations of intensity-based similarity metrics, deep similarity metrics, a deformation smoothness constraint, a deformation physical fidelity constraint, error loss between predicted and reference deformation, adversarial loss, and other auxiliary losses. Intensity-based and deep similarity metrics are used to quantify the dissimilarities between warped and fixed images. A deformation smoothness constraint is used to make predicted DVFs smooth and is usually the first- or second-order derivative of the DVF. The deformation physical fidelity loss encourages the deformation to be physically realistic and comprises cycle-consistency loss, identity loss, and negative Jaco. Det. loss. The error loss between predicted and reference deformation is only used for fully supervised registration. The adversarial loss is specifically used for GAN-based DL models, and is a Boolean value generated by the discriminator that describes how similar the input image is to the fixed image. Auxiliary losses are used only if organ contours or surrogates are provided together with the training data. These losses are often represented as the DSC of contours or the TREs of pointwise surrogates.

Conventional registration algorithms have more types of losses that can improve registration accuracy. For example, on the public benchmark DIR-Lab, Vishnevskiy *et al.* achieved the lowest TRE by including isotropic total variation constrain (135). However, those losses are non-differentiable with simple chain-rule and special optimizers are needed, which are impractical for current DL libraries. DL-based registration can be further improved if differentiable approximations of these advanced losses can be developed.

Conclusions

DL-based 3D medical image registration has been successfully implemented in several studies. In this review, recent progress in DL-based 3D medical image registration is summarized. The collected studies are classified into three categories based on the supervision methods used. Our statistical analysis indicates that direct deformation prediction has increased in popularity, whereas the use of deep iterative registration is gradually decreasing. However, deep similarity metrics can be integrated into other categories to obtain higher registration accuracies. The number of studies on supervised and unsupervised registration are approximately equal, and both strategies have unique advantages and disadvantages. We expect the number of studies in both categories to increase, and new methods that combine the advantages of both strategies will become an emerging area of research.

Acknowledgments

Funding: This manuscript is supported by the following grants from Hong Kong: (I) GRF 151021/18M and GRF 151022/19M from the University Grants Committee (UGC); (II) HMRF 06173276 from the Food and Health Bureau (FHB).

Footnote

Provenance and Peer Review: With the arrangement by the Guest Editors and the editorial office, this article has been reviewed by external peers.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/qims-21-175>). The special issue “Artificial Intelligence for Image-guided Radiation Therapy” was commissioned by the editorial office without any funding or sponsorship. Jing Cai served as the unpaid Guest Editor of the special issue and reports this manuscript is supported by the following grants from Hong Kong: (I) GRF 151021/18M and GRF 151022/19M from the University Grants Committee (UGC); (II) HMRF 06173276 from the Food and Health Bureau (FHB). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related

to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the institutional Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. Informed consent was waived in this retrospective study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Mackie TR, Kapatoes J, Ruchala K, Lu W, Wu C, Olivera G, Forrest L, Tome W, Welsh J, Jeraj R, Harari P, Reckwerdt P, Paliwal B, Ritter M, Keller H, Fowler J, Mehta M. Image guidance for precise conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 2003;56:89-105.
2. Fallone BG, Murray B, Rathee S, Stancescu T, Steciw S, Vidakovic S, Blosser E, Tymofichuk D. First MR images obtained during megavoltage photon irradiation from a prototype integrated linac-MR system. *Med Phys* 2009;36:2084-8.
3. Zhou Y, Wong OL, Cheung KY, Yu SK, Yuan J. A pilot study of highly accelerated 3D MRI in the head and neck position verification for MR-guided radiotherapy. *Quant Imaging Med Surg* 2019;9:1255-69.
4. Yuan J, Wong OL, Zhou Y, Chueng KY, Yu SK. A fast volumetric 4D-MRI with sub-second frame rate for abdominal motion monitoring and characterization in MRI-guided radiotherapy. *Quant Imaging Med Surg* 2019;9:1303-14.
5. Troccaz J, Menguy Y, Bolla M, Cinquin P, Vassal P, Laieb N, Desbat L, Dusserre A, Dal Soglio S. Conformal external radiotherapy of prostatic carcinoma: requirements and experimental results. *Radiother Oncol* 1993;29:176-83.
6. Verellen D, De Ridder M, Linthout N, Tournel K, Soete G, Storme G. Innovations in image-guided radiotherapy. *Nat Rev Cancer* 2007;7:949-60.

7. Dawson LA, Sharpe MB. Image-guided radiotherapy: rationale, benefits, and limitations. *Lancet Oncol* 2006;7:848-58.
8. Liang X, Yin FF, Wang C, Cai J. A robust deformable image registration enhancement method based on radial basis function. *Quant Imaging Med Surg* 2019;9:1315-25.
9. Christensen GE, Song JH, Lu W, El Naqa I, Low DA. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Med Phys* 2007;34:2155-63.
10. Schreibmann E, Thorndyke B, Li T, Wang J, Xing L. Four-dimensional image registration for image-guided radiotherapy. *Int J Radiat Oncol Biol Phys* 2008;71:578-86.
11. Zhang Y, Folkert MR, Huang X, Ren L, Meyer J, Tehrani JN, Reynolds R, Wang J. Enhancing liver tumor localization accuracy by prior-knowledge-guided motion modeling and a biomechanical model. *Quant Imaging Med Surg* 2019;9:1337-49.
12. Reed VK, Woodward WA, Zhang L, Strom EA, Perkins GH, Tereffe W, Oh JL, Yu TK, Bedrosian I, Whitman GJ, Buchholz TA, Dong L. Automatic segmentation of whole breast using atlas approach and deformable image registration. *Int J Radiat Oncol Biol Phys* 2009;73:1493-500.
13. Schwartz DL, Garden AS, Thomas J, Chen Y, Zhang Y, Lewin J, Chambers MS, Dong L. Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. *Int J Radiat Oncol Biol Phys* 2012;83:986-93.
14. Yang D, Brame S, El Naqa I, Aditya A, Wu Y, Goddu SM, Mutic S, Deasy JO, Low DA. Technical note: DIRART-A software suite for deformable image registration and adaptive radiotherapy research. *Med Phys* 2011;38:67-77.
15. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol* 2020;65:20TR01.
16. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *Mach Vis Appl* 2020;31:8.
17. Xiao H, Ren G, Cai J. A review on 3D deformable image registration and its application in dose warping. *Radiat Med Prot* 2020;1:171-8.
18. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
19. Ker J, Wang L, Rao J, Lim T. Deep Learning Applications in Medical Image Analysis. *IEEE Access* 2018;6:9375-89.
20. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017;19:221-48.
21. Liu Y, Chen X, Wang Z, Wang Z, Ward R, Wang X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf Fusion* 2018;42:158-73.
22. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. *Med Phys* 2019;46:e1-e36.
23. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys* 2019;29:86-101.
24. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. *Comput Biol Med* 2018;98:126-46.
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
26. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. *Neurocomputing* 2016;187:27-48.
27. Jeong J, Wang L, Ji B, Lei Y, Ali A, Liu T, Curran WJ, Mao H, Yang X. Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images: Introduction. *Quant Imaging Med Surg* 2019;9:1201-13.
28. Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, Xia Y, Wang G, Yang X. Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images. *Quant Imaging Med Surg* 2019;9:1288-302.
29. Liang X, Li N, Zhang Z, Yu S, Qin W, Li Y, Chen S, Zhang H, Xie Y. Shading correction for volumetric CT using deep convolutional neural network and adaptive filter. *Quant Imaging Med Surg* 2019;9:1242-54.
30. Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, Xie Y. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. *Quant Imaging Med Surg* 2020;10:1223-36.
31. Li W, Kazemifar S, Bai T, Nguyen D, Weng Y, Li Y, Xia J, Xiong J, Xie Y, Owrangi AM, Jiang SB. Synthesizing CT images from MR images with deep learning: model generalization for different datasets through transfer learning. *Biomed Phys Eng Express* 2021. [Epub ahead of print]. doi: 10.1088/2057-1976/abe3a7.
32. Ren G, Lam SK, Zhang J, Xiao H, Cheung AL, Ho WY, Qin J, Cai J. Investigation of a Novel Deep Learning-

- Based Computed Tomography Perfusion Mapping Framework for Functional Lung Avoidance Radiotherapy. *Front Oncol* 2021;11:644703.
33. Ren G, Zhang J, Li T, Xiao H, Cheung LY, Ho WY, Qin J, Cai J. Deep Learning-Based Computed Tomography Perfusion Mapping (DL-CTPM) for Pulmonary CT-to-Perfusion Translation. *Int J Radiat Oncol Biol Phys* 2021. [Epub ahead of print]. doi: 10.1016/j.ijrobp.2021.02.032.
 34. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*; Cham: Springer International Publishing; 2015:234-41.
 35. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*; Cham: Springer International Publishing; 2016:424-32.
 36. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. *Adv Neural Inf Process Syst* 2014;27:2672-80.
 37. You C, Li G, Zhang Y, Zhang X, Shan H, Li M, Ju S, Zhao Z, Zhang Z, Cong W, Vannier MW, Saha PK, Hoffman EA, Wang G. CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). *IEEE Trans Med Imaging* 2020;39:188-203.
 38. Fu Y, Lei Y, Wang T, Higgins K, Bradley JD, Curran WJ, Liu T, Yang X. LungRegNet: An unsupervised deformable image registration method for 4D-CT lung. *Med Phys* 2020;47:1763-74.
 39. Fu Y, Lei Y, Zhou J, Wang T, Yu D, Beitler J, Curran W, Liu T, Yang X. Synthetic CT-aided MRI-CT image registration for head and neck radiotherapy. *SPIE Medical Imaging* 2020. doi: 10.1117/12.2549092.
 40. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci* 2007;19:1498-507.
 41. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 2009;46:786-802.
 42. Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci* 2012;6:171.
 43. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry* 2014;19:659-67.
 44. HD-200 Consortium. The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. *Front Syst Neurosci* 2012;6:62.
 45. Gollub RL, Shoemaker JM, King MD, White T, Ehrlich S, Sponheim SR, et al. The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics* 2013;11:367-88.
 46. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol* 2011;95:629-35.
 47. Dagley A, LaPoint M, Huijbers W, Hedden T, McLaren DG, Chatwal JP, Papp KV, Amariglio RE, Blacker D, Rentz DM, Johnson KA, Sperling RA, Schultz AP. Harvard Aging Brain Study: Dataset and accessibility. *Neuroimage* 2017;144:255-8.
 48. Holmes AJ, Hollinshead MO, O'Keefe TM, Petrov VI, Fariello GR, Wald LL, Fischl B, Rosen BR, Mair RW, Roffman JL, Smoller JW, Buckner RL. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Sci Data* 2015;2:150031.
 49. Fischl B. FreeSurfer. *Neuroimage* 2012;62:774-81.
 50. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv* 2018: 1811.02629v3.
 51. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008;27:685-91.
 52. Gousias IS, Edwards AD, Rutherford MA, Counsell SJ, Hajnal JV, Rueckert D, Hammers A. Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants. *Neuroimage* 2012;62:1499-509.

53. Aubert-Broche B, Griffin M, Pike GB, Evans AC, Collins DL. Twenty new digital brain phantoms for creation of validation image data bases. *IEEE Trans Med Imaging* 2006;25:1410-6.
54. Xiao Y, Fortin M, Unsgård G, Rivaz H, Reinertsen I. REtroSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. *Med Phys* 2017;44:3875-82.
55. Vandemeulebroucke J, Rit S, Kybic J, Clarysse P, Sarrut D. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Med Phys* 2011;38:166-78.
56. Castillo R, Castillo E, Guerra R, Johnson VE, McPhail T, Garg AK, Guerrero T. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol* 2009;54:1849-70.
57. Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. Four-dimensional deformable image registration using trajectory modeling. *Phys Med Biol* 2010;55:305-27.
58. Castillo R, Castillo E, Fuentes D, Ahmad M, Wood AM, Ludwig MS, Guerrero T. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Phys Med Biol* 2013;58:2861-77.
59. Jimenez-Del-Toro O, Muller H, Krenn M, Gruenberg K, Taha AA, Winterstein M, et al. Cloud-Based Evaluation of Anatomical Structure Segmentation and Landmark Detection Algorithms: VISCERAL Anatomy Benchmarks. *IEEE Trans Med Imaging* 2016;35:2459-75.
60. Stolk J, Putter H, Bakker EM, Shaker SB, Parr DG, Piitulainen E, Russi EW, Grebski E, Dirksen A, Stockley RA, Reiber JH, Stoel BC. Progression parameters for emphysema: a clinical investigation. *Respir Med* 2007;101:1924-30.
61. Shieh CC, Gonzalez Y, Li B, Jia X, Rit S, Mory C, Riblett M, Hugo G, Zhang Y, Jiang Z, Liu X, Ren L, Keall P. SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan. *Med Phys* 2019;46:3799-811.
62. Heller N, Sathianathan N, Kalapara A, Walczak E, Moore K, Kaluzniak H, Rosenberg J, Blake P, Rengel Z, Oestreich M, Dean J, Tradewell M, Shah A, Tejpal R, Edgerton Z, Peterson M, Raza S, Regmi S, Papanikolopoulos N, Weight C. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes. *arXiv* 2019: 1904.00445.
63. Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, van Ginneken B, Kopp-Schneider A, Landman BA, Litjens G, Menze B, Ronneberger O, Summers RM, Bilic P, Christ PF, Do RKG, Gollub M, Golia-Pernicka J, Heckers SH, Jarnagin WR, McHugo MK, Napel S, Vorontsov E, Maier-Hein L, Cardoso MJ. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv* 2019: 1902.09063.
64. Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, Summers RM. DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*; Cham: Springer International Publishing; 2015:556-64.
65. Yao J, Burns JE, Munoz H, Summers RM. Detection of vertebral body fractures based on cortical shell unwrapping. *Med Image Comput Comput Assist Interv* 2012;15:509-16.
66. Madabhushi A, Feldman M. Fused Radiology-Pathology Prostate Dataset. The Cancer Imaging Archive 2016. doi: 10.7937/K9/TCIA.2016.TLPMR1AM.
67. Litjens G, Futterer J, Huisman H. Data From Prostate-3T. The cancer imaging archive. 2015. doi: 10.7937/K9/TCIA.2015.QJTV5IL5.
68. Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, van Ginneken B, et al. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med Image Anal* 2014;18:359-73.
69. Horn BKP, Schunck BG. Determining Optical-Flow. *Artif Intell* 1981;17:185-203.
70. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal* 1998;2:243-60.
71. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010;29:196-205.
72. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 2011;54:2033-44.
73. Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans Med Imaging* 2002;21:1421-39.
74. Viola P, Wells WM 3rd. Alignment by maximization of mutual information. *Int J Comput Vis* 1997;24:137-54.
75. Kim J, Fessler JA. Intensity-based image registration using robust correlation coefficients. *IEEE Trans Med Imaging* 2004;23:1430-44.

76. Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning. *IEEE Trans Biomed Eng* 2016;63:1505-16.
77. Simonovsky M, Gutiérrez-Becker B, Mateus D, Navab N, Komodakis N. A deep metric for multimodal registration. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*; Cham: Springer International Publishing; 2016:10-8.
78. Sedghi A, O'Donnell LJ, Kapur T, Learned-Miller E, Mousavi P, Wells WM 3rd. Image registration: Maximum likelihood, minimum entropy and deep learning. *Med Image Anal* 2021;69:101939.
79. Haskins G, Kruecker J, Kruger U, Xu S, Pinto PA, Wood BJ, Yan P. Learning deep similarity metric for 3D MR-TRUS image registration. *Int J Comput Assist Radiol Surg* 2019;14:417-25.
80. Czolbe S, Krause O, Feragen A. DeepSim: Semantic similarity metrics for learned image registration. *arXiv* 2020: 2011.05735.
81. So RWK, Chung ACS. A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using Bhattacharyya Distances. *Pattern Recognition* 2017;62:161-74.
82. Niethammer M, Kwitt R, Vialard FX. Metric Learning for Image Registration. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2019;2019:8455-64.
83. Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T, Comaniciu D. An Artificial Agent for Robust Image Registration. *Proceedings of the AAAI Conference on Artificial Intelligence* 2017;31:4168-75.
84. Ma K, Wang J, Singh V, Tamersoy B, Chang YJ, Wimmer A, Chen T. Multimodal Image Registration with Deep Context Reinforcement Learning. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*; Cham: Springer International Publishing; 2017:240-8.
85. Krebs J, Mansi T, Delingette H, Zhang L, Ghesu FC, Miao S, Maier AK, Ayache N, Liao R, Kamen A. Robust non-rigid registration through agent-based action learning. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Cham: Springer International Publishing; 2017:344-52.
86. Hu J, Luo Z, Wang X, Sun S, Yin Y, Cao K, Song Q, Lyu S, Wu X. End-to-end multimodal image registration via reinforcement learning. *Med Image Anal* 2021;68:101878.
87. Cao X, Yang J, Zhang J, Nie D, Kim M, Wang Q, Shen D. Deformable image registration based on similarity-steered CNN regression. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*; Cham: Springer International Publishing; 2017:300-8.
88. Teng X, Chen Y, Zhang Y, Ren L. Respiratory deformation registration in 4D-CT/cone beam CT using deep learning. *Quant Imaging Med Surg* 2021;11:737-48.
89. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: Fast predictive image registration - A deep learning approach. *Neuroimage* 2017;158:378-96.
90. Rohé MM, Datar M, Heimann T, Sermesant M, Pennec X. SVF-Net: Learning deformable image registration using shape matching. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*; Cham: Springer International Publishing; 2017:266-74.
91. Wang J, Zhang M. Deep Learning for Regularization Prediction in Diffeomorphic Image Registration. *arXiv* 2020: 2011.14229.
92. Sokooti H, Vos BD, Berendsen F, Ghafoorian M, Yousefi S, Lelieveldt B, Išgum I, Staring M. 3D Convolutional Neural Networks Image Registration Based on Efficient Supervised Learning from Artificial Deformations. *arXiv* 2019: 1908.10235.
93. Sokooti H, De Vos B, Berendsen F, Lelieveldt BP, Išgum I, Staring M. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*; Cham: Springer International Publishing; 2017:232-9.
94. Eppenhof KAJ, Pluim JPW. Pulmonary CT Registration Through Supervised Learning With Convolutional Neural Networks. *IEEE Trans Med Imaging* 2019;38:1097-105.
95. Eppenhof KAJ, Lafarge MW, Moeskops P, Veta M, Pluim JPW. Deformable image registration using convolutional neural networks. *Medical Imaging 2018 Image Processing*. SPIE 2018. doi: 10.1117/12.2292443.
96. Guo H, Kruger M, Xu S, Wood BJ, Yan P. Deep adaptive registration of multi-modal prostate images. *Comput Med*

- Imaging Graph 2020;84:101769.
97. Fu Y, Lei Y, Wang T, Patel P, Jani AB, Mao H, Curran WJ, Liu T, Yang X. Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching. *Med Image Anal* 2021;67:101845.
 98. Hu S, Zhang L, Li G, Liu M, Fu D, Zhang W. Infant Brain Deformable Registration Using Global and Local Label-Driven Deep Regression Learning. In: Suk H-I, Liu M, Yan P, Lian C, editors. *Machine Learning in Medical Imaging* 2019;11861:106-14.
 99. Li B, Niessen WJ, Klein S, de Groot M, Ikram MA, Vernooij MW, Bron EE. A hybrid deep learning framework for integrated segmentation and registration: Evaluation on longitudinal white matter tract changes. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Cham: Springer International Publishing; 2019:645-53.
 100. Xu Z, Niethammer M. DeepAtlas: Joint Semi-supervised Learning of Image Registration and Segmentation. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Cham: Springer International Publishing; 2019:420-9.
 101. Zhu Z, Cao Y, Qin C, Rao Y, Lin D, Dou Q, Ni D, Wang Y. Joint affine and deformable three-dimensional networks for brain MRI registration. *Med Phys* 2021;48:1182-96.
 102. Estienne T, Vakalopoulou M, Battistella E, Carré A, Henry T, Lerousseau M, Robert C, Paragios N, Deutsch E. Deep learning based registration using spatial gradients and noisy segmentation labels. *arXiv* 2020: 2010.10897.
 103. Hering A, van Ginneken B, Heldmann S. mlvrnet: Multilevel variational image registration network. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Cham: Springer International Publishing; 2019:257-65.
 104. Fan J, Cao X, Yap PT, Shen D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Med Image Anal* 2019;54:193-206.
 105. Ahmad S, Fan J, Dong P, Cao X, Yap PT, Shen D. Deep Learning Deformation Initialization for Rapid Groupwise Registration of Inhomogeneous Image Populations. *Front Neuroinform* 2019;13:34.
 106. Ha IY, Hansen L, Wilms M, Heinrich MP. Geometric deep learning and heatmap prediction for large deformation registration of abdominal and thoracic CT. *International Conference on Medical Imaging with Deep Learning*; 2019.
 107. Onieva JO, Marti-Fuster B, de la Puente MP, Estépar RSJ. Diffeomorphic lung registration using deep cnns and reinforced learning. In: Stoyanov D, Taylor Z, Kainz B, Maicas G, Beichel RR, Martel A, et al., editors. *Image Analysis for Moving Organ, Breast, and Thoracic Images* 2018;11040:284-94.
 108. Hu Y, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Vercauteren T, Noble JA, Barratt DC. Adversarial deformation regularization for training image registration neural networks. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*; Cham: Springer International Publishing; 2018:774-82.
 109. Cao X, Yang J, Wang L, Xue Z, Wang Q, Shen D. Deep learning based inter-modality image registration supervised by intra-modality similarity. In: Shi Y, Suk HI, Liu M, editors. *Machine Learning in Medical Imaging*; Cham: Springer International Publishing; 2018:55-63.
 110. Estienne T, Vakalopoulou M, Christodoulidis S, Battistella E, Lerousseau M, Carre A, Klausner G, Sun R, Robert C, Mougiakakou S, Paragios N, Deutsch E. U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Cham: Springer International Publishing; 2019:310-9.
 111. Balakrishnan G, Zhao A, Sabuncu MR, Gutttag J, Dalca AV. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Trans Med Imaging* 2019. [Epub ahead of print]. doi: 10.1109/TMI.2019.2897538.
 112. Kuang D. Cycle-Consistent training for reducing negative Jacobian determinant in deep registration networks. In: Burgos N, Gooya A, Svoboda D, editors. *International Workshop on Simulation and Synthesis in Medical Imaging*; Cham: Springer; 2019:120-9.
 113. Kim B, Kim J, Lee JG, Kim DH, Park SH, Ye JC. Unsupervised deformable image registration using cycle-consistent cnn. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Cham: Springer International Publishing;

- 2019;166-74.
114. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal* 2019;52:128-43.
 115. Jiang Z, Yin FF, Ge Y, Ren L. A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration. *Phys Med Biol* 2020;65:015011.
 116. Lei Y, Fu Y, Wang T, Liu Y, Patel P, Curran WJ, Liu T, Yang X. 4D-CT deformable image registration using multiscale unsupervised deep learning. *Phys Med Biol* 2020;65:085003.
 117. Lei Y, Fu Y, Harms J, Wang T, Curran WJ, Liu T, Higgins K, Yang X. 4D-CT Deformable Image Registration Using an Unsupervised Deep Convolutional Neural Network. In: Nguyen D, Xing L, Jiang S, editors. *Artificial Intelligence in Radiation Therapy*; Cham: Springer International Publishing; 2019:26-33.
 118. Shao W, Banh L, Kunder CA, Fan RE, Soerensen SJC, Wang JB, Teslovich NC, Madhuripan N, Jawahar A, Ghanouni P, Brooks JD, Sonn GA, Rusu M. ProsRegNet: A deep learning framework for registration of MRI and histopathology images of the prostate. *Med Image Anal* 2021;68:101919.
 119. Shen Z, Han X, Xu Z, Niethammer M. Networks for joint affine and non-parametric image registration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2019.
 120. Duan L, Ni X, Liu Q, Gong L, Yuan G, Li M, Yang X, Fu T, Zheng J. Unsupervised learning for deformable registration of thoracic CT and cone-beam CT based on multiscale features matching with spatially adaptive weighting. *Med Phys* 2020;47:5632-47.
 121. Yan P, Xu S, Rastinehad AR, Wood BJ. Adversarial Image Registration with Application for MR and TRUS Image Fusion. In: Shi Y, Suk HI, Liu M, editors. *Machine Learning in Medical Imaging* 2018;11046:197-204.
 122. Elmahdy MS, Jagt T, Zinkstok RT, Qiao Y, Shahzad R, Sokooti H, Yousefi S, Incrocci L, Marijnen CAM, Hoogeman M, Staring M. Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer. *Med Phys* 2019;46:3329-43.
 123. Fan J, Cao X, Wang Q, Yap PT, Shen D. Adversarial learning for mono- or multi-modal registration. *Med Image Anal* 2019;58:101545.
 124. Fan J, Cao X, Xue Z, Yap PT, Shen D. Adversarial similarity network for evaluating image alignment in deep learning based registration. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*; Cham: Springer International Publishing; 2018:739-46.
 125. Huang Y, Ahmad S, Fan J, Shen D, Yap PT. Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain MR images. *Med Image Anal* 2021;67:101817.
 126. Kearney V, Haaf S, Sudhyadhom A, Valdes G, Solberg TD. An unsupervised convolutional neural network-based algorithm for deformable image registration. *Phys Med Biol* 2018;63:185017.
 127. Li H, Fan Y. Non-rigid image registration using self-supervised fully convolutional networks without training data. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*; 2018:1075-8.
 128. Stergios C, Mihir S, Maria V, Guillaume C, Marie-Pierre R, Stavroula M, Nikos P. Linear and deformable image registration with 3d convolutional neural networks. In: Stoyanov D, Taylor Z, Kainz B, Maicas G, Beichel RR, Martel A, et al., editors. *Image Analysis for Moving Organ, Breast, and Thoracic Images* 2018;11040:13-22.
 129. Sun L, Zhang S. Deformable mri-ultrasound registration using 3d convolutional neural network. In: Stoyanov D, Taylor Z, Aylward S, Tavares JMRS, Xiao Y, Simpson A, et al., editors. *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*; Cham: Springer International Publishing; 2018:152-8.
 130. Zhao S, Lau T, Luo J, Chang EI, Xu Y. Unsupervised 3D End-to-End Medical Image Registration With Volume Tweening Network. *IEEE J Biomed Health Inform* 2020;24:1394-404.
 131. Yu H, Zhou X, Jiang H, Kang H, Wang Z, Hara T, Fujita H. Learning 3D non-rigid deformation based on an unsupervised deep learning for PET/CT image registration. *SPIE Medical Imaging* 2019. doi: 10.1117/12.2512698.
 132. Ghosal S, Ray N. Deep deformable registration: Enhancing accuracy by fully convolutional neural net. *Pattern Recognit Lett* 2017;94:81-6.
 133. Sentker T, Madesta F, Werner R. GDL-FIRE4D: Deep Learning-Based Fast 4D CT Image Registration. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*; Cham:

- Springer International Publishing; 2018:765-73.
134. Guo Y, Wu X, Wang Z, Pei X, Xu XG. End-to-end unsupervised cycle-consistent fully convolutional network for 3D pelvic CT-MR deformable registration. *J Appl Clin Med Phys* 2020;21:193-200.
135. Vishnevskiy V, Gass T, Szekely G, Tanner C, Goksel O. Isotropic Total Variation Regularization of Displacements in Parametric Image Registration. *IEEE Trans Med Imaging* 2017;36:385-95.

Cite this article as: Xiao H, Teng X, Liu C, Li T, Ren G, Yang R, Shen D, Cai J. A review of deep learning-based three-dimensional medical image registration methods. *Quant Imaging Med Surg* 2021;11(12):4895-4916. doi: 10.21037/qims-21-175