Chapter 14

Image Registration: Fundamentals and Recent Advances Based on Deep Learning

Min Chen, Nicholas J. Tustison, Rohit Jena, and James C. Gee

Abstract

Registration is the process of establishing spatial correspondences between images. It allows for the alignment and transfer of key information across subjects and atlases. Registration is thus a central technique in many medical imaging applications. This chapter first introduces the fundamental concepts underlying image registration. It then presents recent developments based on machine learning, specifically deep learning, which have advanced the three core components of traditional image registration methods—the similarity functions, transformation models, and cost optimization. Finally, it describes the key application of these techniques to brain disorders.

Key words Image registration, Alignment, Atlas

1 Introduction

In medical image analysis, the *correspondence* between important features or analogous anatomy in two images is an important piece of information that can be used to study disease. Knowing the correspondences between spatial locations allows for comparisons between specific anatomical structures in the images. This allows us to answer questions such as "Is this structure larger in subject A than in subject B?" or "Is that structure malformed relative to the average population?" Likewise, knowing correspondences across time allows us to study changes in rates of disease processes. For example, "Is a disease causing the structure to grow or shrink over time?" or "How does the rate of change compare to an healthy individual?"

Correspondences between images also provide the ability to transfer information, which can be used as prior knowledge for tasks such as segmentation. Knowing the boundary for a specific anatomical structure in image A allows the image to be used as an atlas for finding those same boundaries in other images. If the correspondences between images A and B are known, then the

Olivier Colliot (ed.), *Machine Learning for Brain Disorders*, Neuromethods, vol. 197, https://doi.org/10.1007/978-1-0716-3195-9_14,

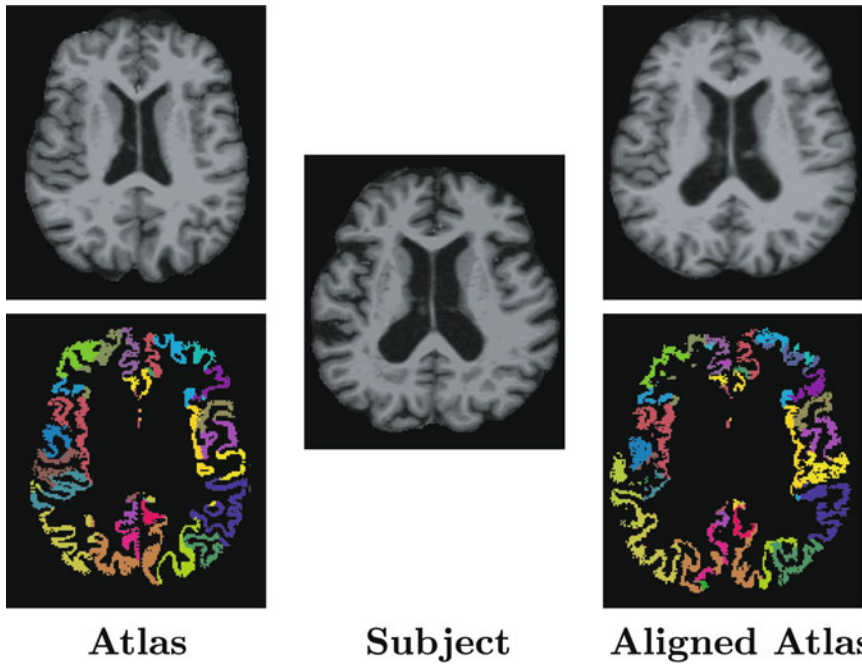

Fig. 1 Shown is an example of an atlas alignment using image registration between two different brain magnetic resonance images. The atlas image (top left) is transformed (top right) to be aligned with the fixed subject image (center). The transformation allows the anatomical labels from the atlas (bottom left) to be directly transferred (bottom right) to label the subject image

boundary in image A can be transferred through the correspondences and used as an approximate starting point for finding the analogous boundaries in image B (called the fixed image).

In the field of medical imaging and computer vision, the task of computing and aligning correspondences between different images is referred to as *image registration*. Given two images, image registration algorithms use image features such as image intensities or structures in the images to find a transformation that best aligns the correspondences between the two images. In Fig. 1, we show an example where such an algorithm is used to align the image intensities between two different brain images. We see that this alignment allows the anatomical labels on an atlas image to be directly transferred to the fixed image.

While the primary concept of image registration is simple, finding the solution is not so straightforward. The subject has been studied extensively for the past 40 years [1], and there is still little of consensus on the best general approach for the problem. We often cannot determine what are the correct correspondences between two images. In addition, we rarely know the exact way to model the transformation that best aligns those correspondences. We see from the example in Fig. 1 that aligning the intensity correspondences does not accurately align all of the anatomical correspondences between the images.

The number of varieties and applications of image registration that have been presented to date is tremendous [2, 3]. In this chapter, we will only discuss a limited subset of these techniques, specifically methods that have been developed in recent years that leverages machine learning (and in particular, deep CNNs) to solve the problem. We will start by providing a brief introduction to the fundamental building blocks of traditional image registration techniques and then delve into how various pieces of these designs have been developed and improved upon using machine learning models.

2 Fundamentals of Image Registration

The main goal of an image registration algorithm is to take a *moving image* and transform it to be spatially or temporally aligned with a target *fixed image*. The algorithm is generally defined by two parts: the type of transformation allowed to be performed on the moving image (the *transformation model*) and a definition of good alignment (the *similarity cost function*) between the two images. The algorithm is often iterative, in which case there is also an *optimizer*, which searches for how to adjust the transformation to best minimize the cost function. This is typically performed by estimating a transformation using the model, applying it to the moving image, and then evaluating the cost function between the transformed moving image and the fixed image. This cost then informs the algorithm on how to estimate a more accurate transformation for the next iteration. The process is repeated and optimized until either the moving and fixed images are considered aligned (i.e., a local minimum is reached in the cost function) or a maximum iteration count is exceeded. Figure 2 summarizes this iterative framework as a block diagram. Figure 3 shows several examples of registration results when using different transformation models to register between two MR images of the brain.

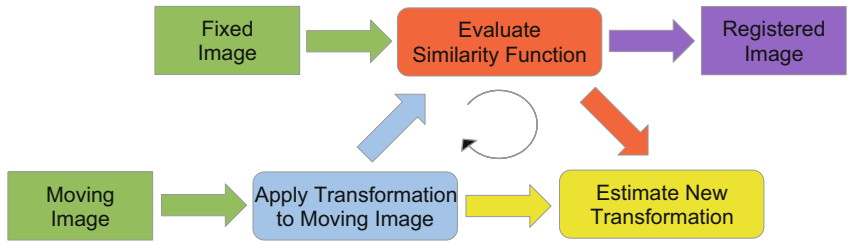

Fig. 2 Block diagram of the general registration framework. The coloring represents the main pieces of the framework: the input images (green), the output image (purple), the similarity cost function (orange), the transformation model (blue), and the optimizer (yellow)

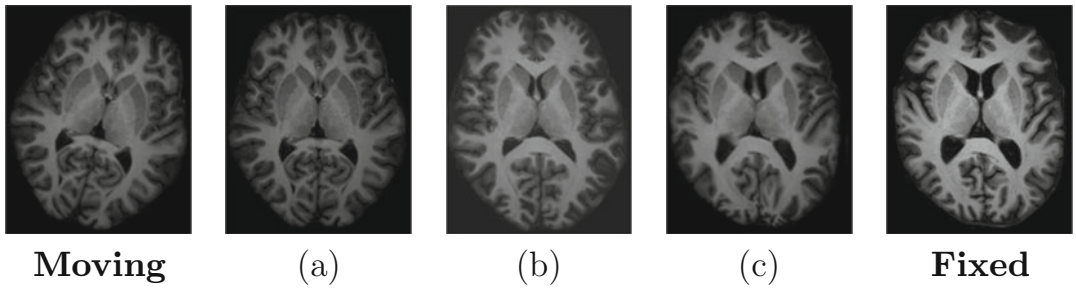

Fig. 3 Shown are examples of registration results between a moving and fixed MR image of the brain from two different subjects, using a **(a)** rigid, **(b)** affine, and **(c)** deformable registration

2.1 Registration as a Minimization Problem

To describe the general registration problem, we begin by using functions $\mathcal{S}(\mathbf{x}')$ and $\mathcal{T}(\mathbf{x})$ to represent the moving and fixed images, where $\mathbf{x}' = (x', y', z')$ and $\mathbf{x} = (x, y, z)$ describe 3D coordinates in the moving and fixed image domains ($\mathbb{D}_{\mathcal{S}}$ and $\mathbb{D}_{\mathcal{T}}$, respectively), and $\mathcal{S}(\mathbf{x}')$ and $\mathcal{T}(\mathbf{x})$ are the intensities of each image at those coordinates. The primary goal of image registration is to estimate a transformation $\mathbf{v} : \mathbb{D}_{\mathcal{T}} \rightarrow \mathbb{D}_{\mathcal{S}}$, which maps corresponding locations between $\mathcal{S}(\mathbf{x}')$ and $\mathcal{T}(\mathbf{x})$. This is generally represented as a *pullback* vector field, $\mathbf{v}(\mathbf{x})$, where the vectors are rooted in the fixed domain and point to locations in the moving domain. The field is applied to $\mathcal{S}(\mathbf{x}')$ by pulling moving image intensities into the fixed domain. This produces the registration result, a transformed moving image, $\tilde{\mathcal{S}}$, defined as

$$\tilde{\mathcal{S}}(\mathbf{x}) = \mathcal{S} \circ \mathbf{v}(\mathbf{x}) = \mathcal{S}(\mathbf{v}(\mathbf{x})) \ , \quad \forall \mathbf{x} \in \mathbb{D}_{\mathcal{T}} \ , \tag{1}$$

which has coordinates in the fixed domain.

The typical registration algorithm aims to find $\mathbf{v}$ such that the images $\tilde{\mathcal{S}}$ and $\mathcal{T}$ are as similar as possible while constraining $\mathbf{v}$ to be smooth and continuous so that the transformation is physically sensible. This can be performed by minimizing a cost function $\mathcal{C}(\cdot, \cdot)$ that evaluates how well aligned $\mathcal{S} \circ \mathbf{v}(\mathbf{x})$ and $\mathcal{T}(\mathbf{x})$ are to each other, and forcing $\mathbf{v}$ to follow a specific transformation model. Together we can describe this problem as a standard minimization problem,

$$\arg\min_{\mathbf{v}} \ \mathcal{C}(\mathcal{S} \circ \mathbf{v}, \mathcal{T}), \tag{2}$$

where the transformation $\mathbf{v}$ is the parameter being optimized.

2.2 Types of Registration

Registration algorithms are generally categorized by the transformation model used to constrain $\mathbf{v}$ and the cost function $\mathcal{C}$ to evaluate similarity. The optimization approach, while important, does not usually characterize the algorithm and is often chosen to best complement the other two components of the algorithm. In this section, we cover several standard models and cost functions

that are regularly used in medical imaging. However, the actual number of registration varieties in the current literature is extensive and outside the scope of this chapter. Several literature reviews on image registration exist for a more comprehensive understanding of the subject [2, 3].

2.2.1 Types of Transformation Models

The transformation model used to constrain $\mathbf{v}$ in the registration algorithm is generally chosen to match the problem at hand. For example, suppose we know that the moving and fixed image is of the same person, and their only difference is caused by a turn of the head in the scanner. In such a case, we would want to use a registration algorithm that restricts $\mathbf{v}$ to only perform translations and rotations in order to limit the possible transformation to what we expect has occurred. However, if the two images are of different people, then we might consider a more fluid transformation that can nonlinearly align parts of the anatomy. Here we will discuss two main archetypes of transformation models that are regularly used in medical imaging.

Global Transformation Models

One common choice for the transformation model is to represent $\mathbf{v}$ entirely through a global transformation on the image coordinate system. Here $\mathbf{v}$ is described by a single linear transformation matrix M and a translation vector $\mathbf{t} = (t_x, t_y, t_z)$:

$$\mathbf{v}(\mathbf{x}) = M\mathbf{x} + \mathbf{t}\ . \tag{3}$$

The transformation matrix M determines the restrictiveness of the model, which is often referred to as the model's *degrees of freedom* (dof). Algorithms that only allow translations and rotations (6 dof[1]) are referred to as *rigid* registrations. In such cases, M is the product of three rotation matrices (one for each axis):

$$M_{\text{rigid}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where θ_x, θ_y, and θ_z determine the amount of rotation around each axis. If global scaling is also allowed (7 dof in total), then the algorithm becomes a *similarity* registration, and M_{rigid} is multiplied with an additional scaling matrix:

$$M_{\text{similarity}} = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{bmatrix} M_{\text{rigid}}\ , \tag{5}$$

[1] Here, dof are given for the 3D case since the vast majority of medical images are 3D.

where s determines the amount of scaling. Finally, adding individual scaling and shearing (12 dof in total) allows for an *affine* registration. Here the scaling matrix is modified to have independent terms s_x, s_y, and s_z for each axis, and a shear matrix is included in the product:

$$M_{\text{affine}} = \begin{bmatrix} 1 & h_{xy} & h_{xz} \\ h_{yx} & 1 & h_{yz} \\ h_{zx} & h_{zy} & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} M_{\text{rigid}} \ , \tag{6}$$

where three pairs of shear terms describe the direction and magnitude of shearing in each axis (h_{yx} and h_{zx} for the x-axis; h_{xy} and h_{zy} for the y-axis; h_{xz} and h_{yz} for the z-axis).

The main application of these models is to account for registration problems where the moving and fixed images differ by very limited transformations. Rigid registration is regularly used to align images of the same subject, allowing for more accurate longitudinal analysis. It is also applied to images from different subjects to remove global misalignment, such as movement or shifts in position while still maintaining the physical structure in the images. Similarity and affine registrations are used when the images are expected to have differences in size or large regional transformations. In medical imaging, they offer a way to normalize different subjects in order to remove effects that are often considered unrelated to the disease being studied, such as the size of the head. In addition, affine registrations can be used to provide an initialization for more fluid registrations by removing large sweeping differences, and allowing the subsequent algorithm to focus on aligning more detailed differences. Figure 3a, b provides examples of results from rigid and affine registrations between brain MRIs from two different subjects.

Deformable Model

The main disadvantage of using only a transformation matrix to represent $\mathbf{v}$ is its inability to account for local differences between the moving and fixed images. To perform such alignments, a *deformable* registration is necessary, where the transformation is individually defined at each point in the image using a vector field:

$$\mathbf{v}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}) \ . \tag{7}$$

The vector field $\mathbf{u}$ is referred to as a *displacement field* and is generally restricted to be smooth and continuous to ensure the overall deformation is regularized so that the object is transformed in a physically sensible way.

Deformable registration can be loosely divided between algorithms that use parametric or nonparametric transformation models to represent $\mathbf{v}$. Parametric registrations use a set number of parameters to control basis functions, such as splines [4] or radial basis

functions [5], to construct and interpolate **v**. The algorithm optimizes these parameters to find the best **v** that minimizes the cost function. The transformations found under these models are often smooth and continuous by construction due to the basis functions used.

Nonparametric registrations are generally designed to create transformations that resemble physical motions such as elasticity [6], viscosity [7], diffusion [8], and diffeomorphism [9]. Rather than optimizing a set of parameters, the algorithm evolves the transformation at every iteration using forces imposed by the model. The strength and direction of these forces are determined by the cost function chosen and the constraints of the physical motion being modeled.

The primary application of deformable registration is to compute and align detailed correspondences between the moving and fixed images. This allows such registrations to be better suited for information transfer tasks, such as deforming anatomical labels in the moving image to match and label the same structures in the fixed image, and providing an initialization using various atlases and priors. In addition, the displacement field learned in the registration represents relative spatial change between correspondences in the moving and fixed image. Hence, it can be used to analyze morphology and shape differences between individuals [10, 11]. Figure 3c shows an example of a deformable registration performed using an adaptive bases algorithm after an affine alignment. Compared to the affine result, we see that the individual structures within the brain are now locally better aligned to match the same structures in the target brain.

2.2.2 Types of Cost Functions

The purpose of the similarity cost function is to quantify how closely aligned the transformed moving image and fixed images are to each other. Since it drives the optimization of the transformation model, the characteristics of the cost function determine what kind of images can be aligned, the degree of accuracy, and the ease of optimization. In this section, we will mainly discuss the three most popular intensity-based cost functions, which are available in most algorithms. Naturally, a large number of cost functions have been proposed in the literature, and a more complete list can be found here [2].

Sum of Square Differences.

Sum of square differences (SSD), or equivalently mean squared error (MSE), between image intensities is one of the most basic and earliest cost functions used for evaluating the similarity between two images. It consists simply of subtracting the intensity difference at each voxel between two images, squaring the difference, and then summing across all the voxels in the entire image. This can be described using

$$\mathcal{C}_{\mathrm{SSD}}(\mathcal{T},\tilde{\mathcal{S}}) = \sum_{\mathbf{x}\in\mathbb{D}_{\mathcal{T}}} \left(\mathcal{T}(\mathbf{x}) - \tilde{\mathcal{S}}(\mathbf{x})\right)^2. \tag{8}$$

The advantage of SSD is that it is computationally efficient, requiring only roughly three or four operations per voxel. In addition, it is very localized, since each voxel between the moving and fixed pair is calculated independently and then summed. This allows non-overlapping regions of the image to be calculated and optimized in parallel. In addition, this provides high local acuity, which allows small spatial differences between the images to be resolved by the cost function.

The main drawback of using SSD is that it is highly dependent on the absolute intensity values in the image. If correspondences in two images do not have exactly the same intensity range, the cost function will fail to register them correctly. As a result, SSD is very susceptible to errors in the presence of artifacts, intensity shifts, and partial voluming in the images.

Normalized Cross Correlation

The cross correlation (CC) function is a concept borrowed from signal processing theory for comparing the similarity between waveforms. It requires vectorizing the image (reshaping the 3D image grir into a single vector), subtracting the mean of each image, and then computing the dot product between the image vectors. The value is then divided by the magnitude of both mean subtracted vectors. This can be described by

$$\mathcal{C}_{\mathrm{CC}}(\mathcal{T},\tilde{\mathcal{S}}) = \left\langle \frac{(\mathcal{T}-\mu_{\mathcal{T}})}{||\mathcal{T}-\mu_{\mathcal{T}}||}, \frac{(\tilde{\mathcal{S}}-\mu_{\tilde{\mathcal{S}}})}{||\tilde{\mathcal{S}}-\mu_{\tilde{\mathcal{S}}}||} \right\rangle \tag{9}$$

$$= \frac{\sum_{\mathbf{x}\in\mathbb{D}_{\mathcal{T}}}((\mathcal{T}(\mathbf{x})-\mu_{\mathcal{T}})(\tilde{\mathcal{S}}(\mathbf{x})-\mu_{\tilde{\mathcal{S}}}))}{||\tilde{\mathcal{S}}-\mu_{\tilde{\mathcal{S}}}||\ ||\mathcal{T}-\mu_{\mathcal{T}}||}, \tag{10}$$

where $\mu_{\mathcal{T}}$ and $\mu_{\tilde{\mathcal{S}}}$ are the mean intensities of each image, and $||\cdot||$ indicate the ℓ_2 norm of the vectorized image intensities.

The primary advantage of CC over SSD is that it is robust to relative intensity shifts in the image, while SSD is not. This is due to the normalization using the image mean and magnitude, and the reliance on multiplication of voxel pairs instead of absolute differences. In the absence of an intensity shift, NCC can be shown to be equivalent to SSD as a cost function for optimization.

The drawback of CC is that both the mean and magnitude require a calculation over the entire image; hence, NCC loses much of the parallelization potential of SSD. In addition, the gradient on the function is more complicated to evaluate, which makes it a more difficult problem to optimize.

Mutual Information

Mutual information (MI) is a probabilistic measure of similarity derived from information theory. Using mutual information for image registration was originally presented in [12], and since then, it has become one of the most widely used registration cost functions [3]. Its success largely comes from its probabilistic nature, which gives it robustness to noise and shifts in intensity. In addition, the measure avoids evaluating direct intensity differences and instead looks at how the intensities between the two images are interdependent. This makes it a very robust measure for evaluating similarity between images with different modalities.

Mutual information is described from an information theory perspective. Hence, we start with a discrete random variable $\mathcal{A}$, with $P_{\mathcal{A}}(a)$ representing the probability of the value a occurring in $\mathcal{A}$. The Shannon entropy [13] of this variable is defined by

$$H(\mathcal{A}) = -\sum_{a} P_{\mathcal{A}}(a) \log(P_{\mathcal{A}}(a)) \ . \tag{11}$$

If the random variable represents image intensity values, then this entropy measures how well a given intensity value in the image can be predicted. Similarly, for a second random variable $\mathcal{B}$ and joint probability distribution $P_{\mathcal{A},\mathcal{B}}(a, b)$, the joint entropy is

$$H(\mathcal{A}, \mathcal{B}) = -\sum_{a,\, b} P_{\mathcal{A},\mathcal{B}}(a, b) \log\big(P_{\mathcal{A},\mathcal{B}}(a, b)\big) \ , \tag{12}$$

which represents how well a given pair of intensity value in the images can be predicted. Using these terms, the mutual information is given by

$$\mathrm{MI}(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \ , \tag{13}$$

which becomes

$$C_{\mathrm{MI}}(\mathcal{T}, \tilde{\mathcal{S}}) = -\left(H(\mathcal{T}) + H(\tilde{\mathcal{S}}) - H(\mathcal{T}, \tilde{\mathcal{S}})\right) \ , \tag{14}$$

within the context of our registration problem. Since MI increases when the images are more similar, we negate the measure in order to fit our minimization framework.

Intuitively, mutual information describes how dependent the intensities in one image are on the other. We see that, when the images are entirely independent, the joint entropy becomes the sum of the individual entropies and the mutual information is zero. On the other hand, when the images are entirely dependent (i.e., $\mathbf{v}$ maps $\mathcal{S}$ exactly to $\mathcal{T}$), then the joint entropy becomes the entropy of the fixed image and the mutual information is maximized. In practice, the entropy and joint entropies are calculated empirically from histograms (and joint histograms) of the intensities in the images.

Since the range of entropy is sensitive to the size of the image, it is common to use a normalized variant of the measure called normalized mutual information (NMI) [14]:

$$\mathrm{NMI}(\mathcal{T},\tilde{\mathcal{S}}) = \frac{H(\mathcal{T}) + H(\tilde{\mathcal{S}})}{H(\mathcal{T},\tilde{\mathcal{S}})} \ . \tag{15}$$

We see that this measure ranges from one to two, where two indicates a perfect alignment. Hence, we must again negate the measure when using it as a cost function to fit our minimization framework.

The main drawback of mutual information comes from its probabilistic nature. The measure relies on an accurate estimate of the probability density of the image intensities. As a result, its effectiveness decreases significantly when working with small regions within the image, where there is not enough intensity samples to accurately estimate such densities. Likewise, the measure is ineffective when facing areas of the image that have poor statistical consistency or lack clear structure [15]. Examples of this include cases where there is overwhelming noise or conversely, when the area has very homogeneous intensities and provides very little information. As a result, mutual information must be calculated over a relatively large region of the image, which reduces the measure's local acuity and diminishes its ability to handle small changes between the moving and fixed images. Lastly, as mentioned before, mutual information is almost entirely calculated from counts of intensity pairs, where the actual intensity value does not matter. While this is useful for addressing multimodal relationships, it also introduces inherent ambiguity into the measure. Given a moving and fixed image, their intensities can be paired in multiple ways to give the exact same mutual information after the transformation. Hence, the measure depends heavily on having a good initialization where the objects being registered are aligned well enough to give the correct intensity pairings at the start of the optimization. Otherwise, mutual information can cause the algorithm to align intensity pairs that incorrectly represent the correspondence between the images, resulting in registration errors [16].

3 Learning-Based Models for Registration

From the previous sections, we can see that there are numerous avenues where machine learning models can potentially be employed to address specific parts of the registration problem. We can build models to estimate the similarity between images, find anatomical correspondences in images, speed up the optimization, or even learn to estimate the transformations directly. As with most learning models, these techniques can be very broadly categorized into supervised and unsupervised techniques.

Supervised image registration within the context of machine learning entails utilizing sufficiently large training data sets of input

moving and fixed image pairs with their corresponding transformations. These data are used to train a model to learn those transformation parameters based on features discovered through the training process. The loss function quantifies the discrepancy between the predicted and input transformation parameters. For example, BIR-Net [17] presents a network for learning-based deformable registration using a dual supervision strategy where the loss is taken between the ground truth deformation field and the predicted field, in addition to the dissimilarity between the warped and fixed image. To prevent slow learning and overfitting, a hierarchical loss function is applied at various levels in the frontal part of the network. DeepFLASH [18] uses the fact that the entire optimization of large deformation diffeomorphic metric mappings (LDDMM) with geodesic shooting can be efficiently carried out in a low-dimensional bandlimited space. This motivates conversion of the velocity fields into the Fourier domain. However, neural networks that operate on complex values are inefficient and not straightforward. The method decomposes the registration framework into separable real and imaginary components and proposes the use of a dual-net that handles the real and imaginary parts separately.

One of the primary challenges with employing supervised models for image registration is that registration problems rarely have ground truth transformation data between the images. Beyond simple rigid transformations, it is too laborious and complex of a task to ask human graders to manually generate full 3D transforms between images. Instead, the desired transformations used in the training data are often obtained using outputs from traditional image registration algorithms or synthetically derived data sets, both of which can limit the capabilities of the model.

Given this limitation, more focus has been directed toward unsupervised learning-based registration approaches, which are more closely related to their traditional analogs in that they lack the use of input transformation data. Optimization is driven via loss functions which incorporate intensity-based similarity quantification in learning the correspondence between the fixed and moving images. This is conceptually analogous to the classic neural network example of unsupervised learning –the autoencoder (cf [19])– where differences between the input and the network-generated predicted version of the input are used to learn latent features characterizing the data. In the case of unsupervised image registration, the optimal transformation is that which maximizes the similarity cost function between the input, specifically the fixed image, and the network-generated predicted version of the input, specifically the warped moving image as determined by the concomitantly derived transform. Direct analogs to iterative methods can be seen in approaches such as [20], which presents a recursive cascade network where the moving image is warped iteratively to fit the

fixed image. Each subnetwork is implemented as a convolutional neural network which predicts the deformation field from the current warped image and the fixed image.

In the following sections, we will provide an overview of several key methodological archetypes in the advancement of image registration that has been made possible through the application of machine learning models. As with other parts of this chapter, it is outside of our scope to attempt to provide a comprehensive coverage of such a broad topic. Instead, we opt to lean toward more contemporary deep neural network-driven approaches, which have arisen from recent widespread adoption of deep learning models in medical image analysis. However, we encourage interested readers to explore several published review articles that can provide a more historical survey of this topic [2, 21].

3.1 Feature Extraction

Much of the early work incorporating machine learning into solving image registration problems involved the detection of corresponding features and then using that information to determine the correspondence relationship between spatial domains. These included training models to find key landmarks [22] or segmentation of structures [23], and fitting established transformations models to provide a full transformation between the images. Unsurprisingly, adaptions of these ideas carried through to deep learning approaches. For example, at the start of the current era of deep learning in image-related research, the authors of [24] proposed point correspondence detection using multiple feed-forward neural networks, each of which is trained to detect a single feature. These neural networks are relatively simple consisting of two hidden layers each with 60 neurons where the output is a probability of it containing a specific feature at the center of a small image neighborhood. These detected point correspondences are then used to estimate the total affine transformation with the RANSAC algorithm [25]. Similarly, *DeepFlow* [26] uses CNNs to detect matching features (called *deep matching*) which are then used as additional information in the large displacement optical flow framework [27]. A relatively small architecture, consisting of six layers, is used to detect features at different convolution sizes which are then matched across scales. Two algorithms for more traditional computer vision applications are proposed in [28] and [29] where both are based on the VGG architecture [30] for 2D homography estimation. The former framework includes both a regression network for determining corner correspondence and a classification network for providing confidence estimates of those predictions. The work in [29], which is publicly available, uses image patch pairs in the input layer and the ℓ_1 photometric loss between them to remove the need for direct supervision. Finally, in the category of feature learning, Wu et al. use nested auto-encoders (AE) to map patchwise image content to learned feature vectors [31]. These

patches are then subsampled based on the importance criteria outlined in [32] which tends toward regions of high informational content such as edges. The AE-based feature vectors at these image patches are then used to drive a HAMMER-based registration [33] which is inherently a feature-based, traditional image registration approach.

3.2 Domain Adaptation

In contrast to detecting discrete corresponding feature points to drive the image registration, a number of learning models have been built to predict the intensity similarity between images, directly. These techniques have largely been focused on addressing intermodality alignment, which remains an open problem due to the complexities of establishing accurate correspondence when the intensities themselves do not necessarily correspond. Models have been developed to learn intermodal spatial relationships by extending traditional concepts of image similarity, such as in [34], where intermodality transformations involving CT and MRI are learned by training on the intramodality image pairs using a basic U-net architecture and incorporating a loss function combining normalized cross correlation (NCC) and explicit regularization for enforcing smoothness of the displacement field. A related idea is developed in [35] which uses labeled data and intensity information during the training phase such that only unlabeled image data is required for prediction. The latter architecture is a densely connected U-net architecture with three types of residual shortcuts [36]. For the loss function, the authors use a multiscale Dice function with an explicit regularization term for estimating both global and local transformations. Similarity functions can also be formulated directly using learning models, such as in [37] where a two-channel network is developed for input image patches (T1- and T2-weighted brain images), and likewise, the B-spline image registration algorithm developed from the Insight Toolkit [38], which leverages the output of a CNN-based similarity measure for comparison with an identical registration setup employing mutual information.

In recent years, intermodality registration has benefited from progress made in the field of *domain adaptation*, also referred to as *image synthesis* in earlier works. The general premise behind these frameworks is that learning-based models can be used to establish the latent relationship between the intensity domains between different modalities. This allows an image in one modality to be synthesized into the other modality, or alternatively both modalities can be moved into a third artificial modality that has shared features from both modalities. When applied to image registration, these synthesized modalities can then be used to convert multimodal registration problems into mono-modal problems that can be solved by leveraging the efficiency and accuracy of mono-modal registration techniques. [39]

Of particular note in this area are methods developed around generative adversarial networks (GANs), first introduced by Goodfellow and colleagues [40], which have increasingly found traction in addressing many types of deep learning problems in the medical imaging domain [41] including image registration. GANs are a special type of network composed of two adversarial subnetworks known as the *generator* (usually characterized by deconvolutional layers) and the *discriminator* (usually a CNN). These work in a minimax fashion to learn data distributions in the absence of extensive sample data. Seeded with a random noise image (e.g., sampled from a uniform or Gaussian distribution), the generator produces synthetic images which are then evaluated by the discriminator as belonging either to the true or synthetic data distributions in terms of some probability scalar value. This back-and-forth results in a generator network which continually improves its ability to produce data that more closely resembles the true distribution while simultaneously enhancing the discriminator's ability to judge between true and synthetic data sets. Since the original "vanilla" GAN paper, the number of proposed GAN extensions has exploded in the literature. Initial extensions included architectural modifications for improved stability in training which have since become standard (e.g., deep convolutional GANs [42]). Please refer to Chap. 5 for a more extensive coverage of GANs.

In order to constrain the mapping between moving and fixed images, the GAN-based approach outlined in [43] combines a content loss term (which includes subterms for normalized mutual information, structural similarity [44], and a VGG-based filter feature ℓ_2-norm between the two images) with a "cyclical" adversarial loss. This is constructed in the style of [45] who proposed this GAN extension, CycleGAN, to ensure that the normally underconstrained forward intensity mapping is consistent with a similarly generated inverse mapping for "image-to-image translation" (e.g., converting a Monet painting to a realistic photo or rendering a winter nature scene as its summer analog). However, in this case, the cyclical aspect is to ensure a regularized field through forward and inverse displacement consistency.

The work of [46] employs discriminator training between finite-element modeling and generated displacements for the prostate and surrounding tissues to regularize the predicted displacement fields. The generator loss employs the weakly supervised learning method proposed by the same authors in [47] whereby anatomical labels are used to drive registration during training only. The generator is constructed from an encoder/decoder architecture based on ResNet blocks [36]. The prediction framework includes both localized tissue deformation and the linear coordinate system changes associated with the ultrasound imaging acquisition.

In [48], the discriminator loss is based on quantification of how well two images are aligned where the negative cases derive from the registration generator and the positive cases consist of identical images (plus small perturbations). Explicit regularization is added to the total loss for the registration network which consists of a U-net type architecture that extracts two 3D image patches as input and produces a patchwise displacement field. The discriminator network takes an image pair as input and outputs the similarity probability.

3.3 Transformation Learning

Many of the methods described so far have been centered around using learning models to establish spatial correspondences between images, and then fitting traditional transformation models to align the images. An alternative approach is to directly learn and predict the transformation between images. Earlier work [49] employed CNN-based regression for estimation of 2D/3D rigid image alignment of 3D X-ray attenuation maps derived from CT and corresponding 2D digitally reconstructed (DRR) X-ray images. The transformation space is partitioned into distinct zones where each zone corresponds to a CNN-based regressor which learns transformation parameters in a hierarchical fashion. The loss function is the mean squared error on the transformation parameters.

A novel deep learning perspective was given in [50] where displacement fields are assumed to form low-dimensional manifolds and are represented in the proposed fully connected network as low-dimensional vectors. From the input vector, the network generates a 2D displacement field used to warp the moving image using bilinear interpolation. The absolute intensity difference is used to optimize the parameters of network and latent vectors. Instead of explicit regularization of the displacement field, the sum of squares of the network weights is included with the intensity error term in the loss function. Instead of training with a loss function based on similarity measures between fixed and moving images, the works of [51, 52] formulate the loss in terms of the squared difference between ground truth and predicted transformation parameters. In terms of network architecture, [51] employs a variant of U-net for training/prediction based on reference deformations provided by registration of previously segmented ROIs for cardiac matching where priority is alignment of the epicardium and endocardium. Displacement fields are parameterized by stationary velocity fields [53]. In contrast, [52] uses a smaller version of the VGG architecture to learn the parameters of a $6 \times 6 \times 6$ thin-plate spline grid.

In 2015, Jaderberg and his fellow co-authors described a powerful new module, known as the spatial transformer network (STN) [54][2] which features prominently now in many contemporary deep

[2] Note that these networks are different from transformers and visual transformers described in Chap. 6.

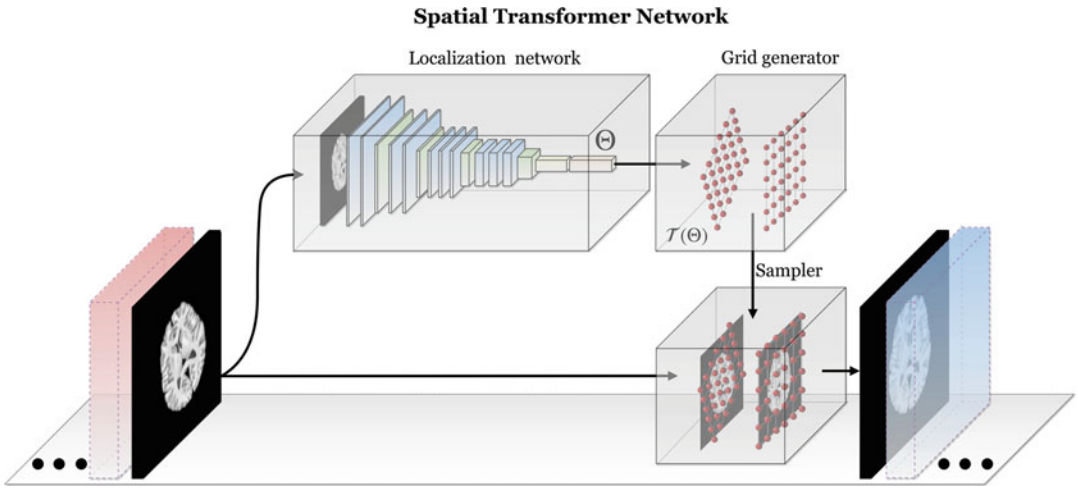

Fig. 4 Diagrammatic illustration of the spatial transformer network. The STN can be placed anywhere within a CNN to provide spatial invariance for the input feature map. Core components include the localization network used to learn/predict the parameters which transform the input feature map. The transformed output feature map is generated with the grid generator and sampler. ©2019 Elsevier. Reprinted, with permission, from [21]

learning-based registration approaches. Generally, STNs enhance CNNs by permitting a flexibility which allows for an explicit spatial invariance that goes beyond the implicitly limited translational invariance associated with the architecture's pooling layers. In many image-based tasks (e.g., localization or segmentation), designing an algorithm that can account for possible pose or geometric variation of the object(s) of interest within the image is crucial for maximizing performance. The STN is a fully differentiable layer which can be inserted anywhere in the CNN to learn the parameters of the transformation of the input feature map (not necessarily an image) which renders the output in such a way so as to optimize the network based on the specified loss function. The added flexibility and the fact that there is no manual supervision or special handling required make this module an essential addition for any CNN-based toolkit.

An STN comprises three principal components: (1) a localization network, (2) a grid generator, and (3) a sampler (*see* Fig. 4). The localization network uses the input feature map to learn/regress the transformation parameters which optimize a specified loss function. In many examples provided, this amounts to transforming the input feature map to a quasi-canonical configuration. The actual architecture of the localization network is fairly flexible, and any conventional architecture, such as a fully connected network (FCN), is suitable as long as the output maps to the continuous estimate of the transformation parameters. These transformation parameters are then applied to the output of the grid generator which are simply the regular coordinates of the input

image (or some normalized version thereof). The sampler, or interpolator, is used to map the transformed input feature map to the coordinates of the output feature map.

Since Jaderberg's original STN formulation, extensions have been proposed such as the inverse compositional STN (IC-STN) [55] and the diffeomorphic transformer network [56]. Two issues with the STN include the following: (1) potential boundary effects in which learned transforms require sampling outside the boundary of the input image which can cause potential learning errors for subsequent layers and (2) the single-shot estimate of the learned transform which can compromise accuracy for large transformation distances. The IC-STN addresses both of these issues by (1) propagating transformation parameters instead of propagating warped input feature maps until the final transformation layer and (2) recurrent usage of the localization network for inferring transform compositions in the spirit of the inverse compositional Lucas-Kanade algorithm [57].

Although discussion of transform generalizability was included in the original STN paper [54], discussion was limited to affine, attention (scaling + translation), and thin-plate spline transforms which all comply with the requirement of differentiability. This work was extended to diffeomorphic transforms in [56]. The computational load associated with generating traditional diffeomorphisms through velocity field integration [58] motivated the use of continuous piecewise affine-based (CPAB) transformations [59]. The CPAB approach utilizes a tesselation of the image domain which translates into faster and more accurate generation of the resulting diffeomorphism. Although this does constrain the flexibility of the final transformation, the framework provides an efficient compromise for use in deep learning architectures. Analogous to traditional image registration, the deep diffeomorphic transformer layer can be placed in serial following an affine-based STN layer for a global-to-local total transformation estimation. This is demonstrated in the experiments reported in [56].

The development of the STN has led to a number of notable generalized deep learning-based registration approaches. *VoxelMorph*, first presented in [60], incorporates a U-net architecture with a STN where the input layer consists of the concatenated full fixed and moving image volumes resized and cropped to $160 \times 192 \times 224$ voxels. The output consists of the voxelwise displacement field of the same size as the input (times three—one for each vector component). The loss function for training combines cross correlation and a diffusion regularizer on the spatial gradients of the displacement field. This was extended to a generative approach in [61] to yield diffeomorphic transformations based on SVFs [53] using novel scaling and squaring network layers. The U-net architecture is used to estimate the distribution parameters of the velocity fields encapsulated by training data. A new imaging

pair can then be registered by sampling from this learned distribution, computing the resulting diffeomorphic transformation, and then warping the moving image. The underlying code has been made publically available which has facilitated independent evaluations such as [62] to compare performance with traditional algorithms (i.e., IRTK [63], AIR [64], Elastix [65], ANTs [66], and NiftyReg [67]). Other variations include CycleMorph [68], which uses a cycle-consistency objective to learn to produce the original image from the deformed image conditioned on the transformation. This prevents degeneracies in the learned registration fields and demonstrates the potential to preserve topologies by inducing cycle consistency on the images. Another generative image registration approach is that of [69] which uses a conditional variational autoencoder [70], an extension of the variational autoencoder [71] which permits incorporation of additional information for latent inference modeling. This multi-scale generative framework encodes the SVFs which are ultimately converted to the total transformation field in a similar fashion as [61].

3.4 Optimization and Equation Solving

A current limitation of traditional registration techniques is the computation cost associated with finding an iterative solution. Most existing registration methods do not scale linearly with image size; thus, as advancements in medical imaging lead to increasingly higher resolution data, the time scale to operate registration techniques can expand to hours, and possibly days, per registration. While not specific to image registration, one area of research that can help address this is the application of learning models to replace classic optimization and equation solving techniques. These can lead to dramatic speed up of existing registration techniques while maintaining the same transformation models. Examples of advancements in this area include the use of learning-based ODE solutions to perform diffeomorphic registration [72] and the use of deep learning to initialize classical optimization approaches, such as Newton's method [73].

4 Registration in the Study of Brain Disorders

This final section will explore how learning-based models have impacted several primary applications of image registration, particularly for the study of diseases. As before, this discussion is far from comprehensive, but more to demonstrate current trends in using machine learning models to advance common areas of registration-driven image analysis.

4.1 Spatial Normalization and Atlasing

Normative and disease-specific atlases play an important role in the characterization of a disease. By registering images from different subjects into a common atlas space (i.e., *spatial normalization*), we can remove typical variability between subjects, such as brain size, to allow for more sensitive detection of disease-driven differences between subjects. Learning-based registration can enable higher throughput registration during atlas construction [74], thus allowing more subjects to be included into the atlas and better encompassing the variability within a cohort. Various models have been proposed to embed these advantages directly into the network, such as [75], which uses a joint learning framework where image attributes are used to learn conditional templates, and an efficient deformation to these templates is jointly learned. In addition, learning models have been used to provide priors for the atlas [76] and establish groupwise correspondence within a cohort [77].

4.2 Label Transfer

As described in earlier sections, establishing correspondences between images via image registration allows for the transfer of spatially embedded data, such as structural annotations and segmentations, between different images and subjects. This method, colloquially referred to as *label transfer*, allows for automatic identification of anatomy in the image that may be relevant to a disease. While a natural application of learning models for label transfer is to simply replace traditional registration approaches with learning-based ones, there has also been more sophisticated integration of machine learning into these frameworks. Popular among these are joint techniques that aim to integrate and solve for both the segmentation and registration problem simultaneously in the same framework [78, 79]. For example, LT-Net [80] learns a multi-atlas registration using cycle consistency and a LSGAN objective [81] to discriminate synthesized images from real ones. Cycle consistency is applied in the image space (between the true atlas and the reconstructed atlas), the transformation space (a voxel warped from the forward transformation composed with the reversed transformation would end up in its starting point), and the segmentation label space. Learning models have also been shown to be effective for correcting systematic errors in both the registration and segmentation parts of the framework [82]. Other models have been proposed for replacing non-registration parts of the standard multi-atlas label transfer framework, such as the voting scheme [83].

4.3 Morphometry

Voxel-based [84] and tensor-based [85] morphometry is the analysis of the transformation result from an image registration to study the shape and structural characteristics of a disease. In these approaches, a disease cohort is spatially normalized into a common space and the warped images and resulting deformation fields from each registration are statistically compared on a voxel level to reveal

morphological characteristics in the cohort. Machine learning models offer new ways to analyze the resulting morphology, such as integrating them as part of a multivariate biomarker framework to detect a disease [86, 87].

5 Conclusion

Image registration is a core pillar of modern-day image analysis, allowing for the alignment and transfer of spatial information between subjects and imaging modalities. Learning-based models have marked improvements on core aspects of image registration, ranging from more accurate feature detection, to better intensity correspondences, particularly across modalities, to improving the speed and accuracy of the alignment.

Acknowledgements

The authors wish to acknowledge the staff and researchers at the Penn Image Computing and Science Laboratory (PICSL) for their support and expertise.

References

1. Anuta PE (1970) Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques. IEEE Trans Geosci Electron 8(4):353–368
2. Sotiras A, Davatzikos C, Paragios N (2013) Deformable medical image registration: a survey. IEEE Trans Med Imaging 32(7):1153–1190
3. Pluim JP, Maintz JA, Viergever MA (2003) Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22(8):986–1004
4. Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11(6):567–585
5. Rohde GK, Aldroubi A, Dawant BM (2003) The adaptive bases algorithm for intensity-based nonrigid image registration. IEEE Trans Med Imaging 22(11):1470–1479
6. Gee JC, Bajcsy RK (1998) Elastic matching: continuum mechanical and probabilistic analysis. Brain Warping 2
7. Christensen GE, Rabbitt RD, Miller MI (1996) Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5(10):1435–1447
8. Thirion JP (1998) Image matching as a diffusion process: an analogy with Maxwell's demons. Med Image Anal 2(3):243–260
9. Beg MF, Miller MI, Trouvé A, Younes L (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 61(2):139–157
10. Ashburner J, Friston KJ (2000) Voxel-based morphometry—the methods. NeuroImage 11(6):805–821
11. Davatzikos C, Genc A, Xu D, Resnick SM (2001) Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. NeuroImage 14(6):1361–1369
12. Wells III WM, Viola P, Atsumi H, Nakajima S, Kikinis R (1996) Multi-modal volume registration by maximization of mutual information. Med Image Anal 1(1):35–51
13. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
14. Studholme C, Hill DL, Hawkes DJ (1999) An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn 32(1):71–86
15. Andronache A, von Siebenthal M, Székely G, Cattin P (2008) Non-rigid registration of

multi-modal images using both mutual information and cross-correlation. Med Image Anal 12(1):3–15

16. Maes F, Vandermeulen D, Suetens P (2003) Medical image registration using mutual information. Proc IEEE 91(10):1699–1722
17. Fan J, Cao X, Yap PT, Shen D (2019) BIRNet: brain image registration using dual-supervised fully convolutional networks. Med Image Anal 54:193–206
18. Wang J, Zhang M (2020) Deepflash: an efficient network for learning-based medical image registration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4444–4452
19. Hinton GE, Zemel RS (1994) Autoencoders, minimum description length and Helmholtz free energy. In: Cowan JD, Tesauro G, Alspector J (eds) Advances in neural information processing systems. Morgan-Kaufmann, Burlington, pp 3–10
20. Zhao S, Dong Y, Chang EIC, Xu Y (2019) Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV)
21. Tustison NJ, Avants BB, Gee JC (2019) Learning image-based spatial transformations via convolutional neural networks: a review. Magn Reson Imaging 64:142–153
22. Ozuysal M, Calonder M, Lepetit V, Fua P (2009) Fast keypoint recognition using random ferns. IEEE Trans Pattern Anal Mach Intell 32(3):448–461
23. Powell S, Magnotta VA, Johnson H, Jammalamadaka VK, Pierson R, Andreasen NC (2008) Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. NeuroImage 39(1): 238–247
24. Sergeev S, Zhao Y, Linguraru MG, Okada K (2012) Medical image registration using machine learning-based interest point detector. In: Proceedings of the SPIE
25. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm ACM 24(6): 381–395
26. Weinzaepfel P, Revaud J, Harchaoui Z, Schmid C (2013) Deepflow: large displacement optical flow with deep matching. In: Proceedings of the IEEE international conference on computer vision, pp 1385–1392. https://doi.org/10.1109/ICCV.2013.175
27. Brox T, Malik J (2011) Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans Pattern Anal Mach Intell 33(3):500–513. https://doi.org/10.1109/TPAMI.2010.143
28. DeTone D, Malisiewicz T, Rabinovich A (2016) Deep image homography estimation. arXiv:160603798
29. Nguyen T, Chen SW, Shivakumar SS, Taylor CJ, Kumar V (2018) Unsupervised deep homography: a fast and robust homography estimation model. In: Proceedings of IEEE robotics and automation letters
30. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556, 1409.1556
31. Wu G, Kim M, Wang Q, Munsell BC, Shen D (2016) Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans Biomed Eng 63(7):1505–1516. https://doi.org/10.1109/TBME.2015.2496253
32. Wang Q, Wu G, Yap PT, Shen D (2010) Attribute vector guided groupwise registration. NeuroImage 50(4):1485–1496. https://doi.org/10.1016/j.neuroimage.2010.01.040
33. Shen D, Davatzikos C (2002) Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging 21(11):1421–1439. https://doi.org/10.1109/TMI.2002.803111
34. Cao X, Yang J, Zhang J, Nie D, Kim MJ, Wang Q, Shen D (2017) Deformable image registration based on similarity-steered cnn regression. In: Proceedings of the international conference on medical image computing and computer-assisted intervention 10433:300–308. https://doi.org/10.1007/978-3-319-66182-7_35
35. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, Wang G, Bandula S, Moore CM, Emberton M, Ourselin S, Noble JA, Barratt DC, Vercauteren T (2018) Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Anal 49:1–13. https://doi.org/10.1016/j.media.2018.07.002
36. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. CoRR abs/1512.03385. http://arxiv.org/abs/1512.03385, 1512.03385
37. Simonovsky M, Gutierrez-Becker B, Mateus D, Navab N, Komodakis N (2016) A deep metric for multimodal registration. In: Proceedings of the international conference on medical image computing and computer-assisted intervention

38. Yoo TS, Metaxas DN (2005) Open science–combining open data and open source software: medical image analysis with the insight toolkit. Med Image Anal 9(6):503–6. https://doi.org/10.1016/j.media.2005.04.008
39. Chen M, Carass A, Jog A, Lee J, Roy S, Prince JL (2017) Cross contrast multi-channel image registration using image synthesis for mr brain images. Med Image Anal 36:2–14
40. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems
41. Yi X, Walia E, Babyn P (2018) Generative adversarial network in medical imaging: a review. Preprint
42. Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the international conference on learning representations
43. Mahapatra D, Antony B, Sedai S, Garnavi R (2018) Deformable medical image registration using generative adversarial networks. In: Proceedings of IEEE 15th international symposium on biomedical imaging
44. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
45. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE international conference on computer vision
46. Hu Y, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Vercauteren T, Noble JA, Barratt DC (2018) Adversarial deformation regularization for training image registration neural networks. In: Proceedings of the international conference on medical image computing and computer-assisted intervention
47. Hu Y, Modat M, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Noble JA, Barratt DC, Vercauteren T (2018) Label-driven weakly-supervised learning for multimodal deformable image registration. In: Proceedings of IEEE 15th international symposium on biomedical imaging
48. Fan J, Cao X, Xue Z, Yap PT, Shen D (2018) Adversarial similarity network for evaluating image alignment in deep learning based registration. In: Proceedings of the international conference on medical image computing and computer-assisted intervention
49. Miao S, Wang ZJ, Liao R (2016) A CNN regression approach for real-time 2D/3D registration. IEEE Trans Med Imaging 35(5):1352–1363. https://doi.org/10.1109/TMI.2016.2521800
50. Sheikhjafari A, Noga M, Punithakumar K, Ray N (2018) Unsupervised deformable image registration with fully connected generative neural network. In: Proceedings of medical imaging with deep learning
51. Rohé MM, Datar M, Heimann T, Sermesant M, Pennec X (2017) SVF-Net: learning deformable image registration using shape matching. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S (eds) Proceedings of the international conference on medical image computing and computer-assisted intervention. Springer International Publishing, Cham, pp 266–274
52. Eppenhof KAJ, Lafarge MW, Moeskops P, Veta M, Pluim JPW (2018) Deformable image registration using convolutional neural networks. In: Proceedings of the SPIE: medical imaging: image processing
53. Arsigny V, Commowick O, Pennec X, Ayache N (2006) A log-euclidean framework for statistics on diffeomorphisms. In: Proceedings of the international conference on medical image computing and computer-assisted intervention, vol 9, pp 924–31
54. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K (2015) Spatial transformer networks. In: Neural information processing systems
55. Lin CH, Lucey S (2017) Inverse compositional spatial transformer networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
56. Detlefsen NS, Freifeld O, Hauberg S (2018) Deep diffeomorphic transformer networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
57. Baker S, Matthews I (2004) Lucas-kanade 20 years on: a unifying framework. Int J Comput Vis 56(3):221–255. https://doi.org/10.1023/B:VISI.0000011205.11775.fd
58. Beg MF, Miller MI, Trouvé A, Younes L (2004) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 61(2):139–157
59. Freifeld O, Hauberg S, Batmanghelich K, Fisher JW (2017) Transformations based on continuous piecewise-affine velocity fields. IEEE Trans Pattern Anal Mach Intell 39(12):2496–2509. https://doi.org/10.1109/TPAMI.2016.2646685

60. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV (2018) An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition
61. Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR (2018) Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G (eds) Proceedings of the international conference on medical image computing and computer-assisted intervention. Springer International Publishing, Cham, pp 729–738
62. Nazib A, Fookes C, Perrin D (2018) A comparative analysis of registration tools: traditional vs deep learning approach on high resolution tissue cleared data. Preprint. arXiv
63. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18(8):712–721. https://doi.org/10.1109/42.796284
64. Woods RP, Mazziotta JC, Cherry SR (1993) MRI-PET registration with automated algorithm. J Comput Assist Tomogr 17(4):536–546
65. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW (2010) Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29(1):196–205. https://doi.org/10.1109/TMI.2009.2035616
66. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage 54(3):2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025
67. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S (2010) Fast free-form deformation using graphics processing units. Comput Methods Prog Biomed 98(3):278–284. https://doi.org/10.1016/j.cmpb.2009.09.002
68. Kim B, Kim DH, Park SH, Kim J, Lee JG, Ye JC (2021) Cyclemorph: cycle consistent unsupervised deformable image registration. Med Image Anal 71:102036
69. Krebs J, Mansi T, Mailhé B, Ayache N, Delingette H (2018) Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Proceedings of the 4th international workshop, DLMIA and 8th international workshop, ML-CDS
70. Sohn K, Lee H, Yan X (2015) Learning structured output representation using deep conditional generative models. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (eds) Advances in neural information processing systems 28. Curran Associates Inc., Red Hook, pp 3483–3491
71. Kingma DP, Welling M (2014) Auto-encoding variational bayes. In: Proceedings of the 2nd international conference on learning representations (ICLR)
72. Wu Y, Jiahao TZ, Wang J, Yushkevich PA, Hsieh MA, Gee JC (2022) Nodeo: a neural ordinary differential equation based optimization framework for deformable image registration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 20804–20813
73. Huang J, Wang H, Yang H (2020) Int-deep: a deep learning initialized iterative method for nonlinear problems. J Comput Phys 419:109675
74. Iqbal A, Khan R, Karayannis T (2019) Developing a brain atlas through deep learning. Nat Mach Intell 1(6):277–287
75. Dalca A, Rakic M, Guttag J, Sabuncu M (2019) Learning conditional deformable templates with convolutional networks. In: Wallach H, Larochelle H, Beygelzimer A, d' Alché-Buc F, Fox E, Garnett R (eds) Advances in neural information processing systems. Curran Associates Inc., Red Hook, vol 32. https://proceedings.neurips.cc/paper/2019/file/bbcbff5c1f1ded46c25d28119a85c6c2-Paper.pdf
76. Wu N, Wang J, Zhang M, Zhang G, Peng Y, Shen C (2022) Hybrid atlas building with deep registration priors. In: 2022 IEEE 19th international symposium on biomedical imaging (ISBI). IEEE, Piscataway, pp 1–5
77. Yang J, Küstner T, Hu P, Liò P, Qi H (2022) End-to-end deep learning of non-rigid group-wise registration and reconstruction of dynamic MRI. Front Cardiovasc Med 9
78. Sinclair M, Schuh A, Hahn K, Petersen K, Bai Y, Batten J, Schaap M, Glocker B (2022) Atlas-ISTN: joint segmentation, registration and atlas construction with image-and-spatial transformer networks. Med Image Anal 78:102383
79. Xu Z, Niethammer M (2019) Deepatlas: joint semi-supervised learning of image registration and segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 420–429

80. Wang S, Cao S, Wei D, Wang R, Ma K, Wang L, Meng D, Zheng Y (2020) LT-Net: label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)
81. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2794–2802
82. Wang H, Yushkevich PA (2013) Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. Front Neuroinform 7:27
83. Ding Z, Han X, Niethammer M (2019) Votenet: a deep learning label fusion method for multi-atlas segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 202–210
84. Ashburner J, Friston KJ (2000) Voxel-based morphometry–the methods. NeuroImage 11(6 Pt 1):805–821. https://doi.org/10.1006/nimg.2000.0582
85. Hua X, Leow AD, Parikshak N, Lee S, Chiang MC, Toga AW, Jack Jr CR, Weiner MW, Thompson PM, Initiative ADN, et al (2008) Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. NeuroImage 43(3):458–469
86. Pahuja G, Prasad B (2022) Deep learning architectures for Parkinson's disease detection by using multi-modal features. Comput Biol Med 105610
87. Huang H, Zheng S, Yang Z, Wu Y, Li Y, Qiu J, Cheng Y, Lin P, Lin Y, Guan J, et al (2022) Voxel-based morphometry and a deep learning model for the diagnosis of early alzheimer's disease based on cerebral gray matter changes. Cereb Cortex 33(3):754–763