scientific reports

OPEN

A deep learning based framework for the registration of three dimensional multi-modal medical images of the head

Kh Tohidul Islam✉, Sudanthi Wijewickrema & Stephen O'Leary

Image registration is a fundamental task in image analysis in which the transform that moves the coordinate system of one image to another is calculated. Registration of multi-modal medical images has important implications for clinical diagnosis, treatment planning, and image-guided surgery as it provides the means of bringing together complimentary information obtained from different image modalities. However, since different image modalities have different properties due to their different acquisition methods, it remains a challenging task to find a fast and accurate match between multi-modal images. Furthermore, due to reasons such as ethical issues and need for human expert intervention, it is difficult to collect a large database of labelled multi-modal medical images. In addition, manual input is required to determine the fixed and moving images as input to registration algorithms. In this paper, we address these issues and introduce a registration framework that (1) creates synthetic data to augment existing datasets, (2) generates ground truth data to be used in the training and testing of algorithms, (3) registers (using a combination of deep learning and conventional machine learning methods) multi-modal images in an accurate and fast manner, and (4) automatically classifies the image modality so that the process of registration can be fully automated. We validate the performance of the proposed framework on CT and MRI images of the head obtained from a publicly available registration database.

Image registration is a spatial transformation process which brings different images into a single coordinate system. This enables direct comparison and integration of data obtained from multiple sources. Image registration has wide application in numerous fields, such as remote sensing, agriculture, infrared physics, and biomedical image analysis[1–3]. In the field of medical image analysis, it enables the integration of information from different temporal points and/or imaging modalities. For example, MRI (magnetic resonance imaging) scans of a patient at different times can show details of the growth of a tumor, comparison of pre- and post-operative scans can indicate the effects of a surgery, and using images from different imaging modalities (for example CT (computerized tomography), MRI (magnetic resonance imaging), and PET (positron emission tomography)) can provide additional information that can be used to improve the diagnosis of diseases. Figure 1 illustrates the process of 3D medical image registration.

Image registration methods can be broadly classified into intensity-based and feature-based methods depending on the matching technique used[4,5]. Assuming one of the images is fixed and the other is moving, both types of methods rely on determining the transform that minimises a cost function that defines the dissimilarity between the fixed image and the transformed moving image[6]. In intensity-based registration, the (dis)similarity of the two images is defined in terms of the correlation between pixel/voxel intensities. In contrast, in feature-based registration methods, image features (for example, edges, lines, contours, and point-based features) are extracted prior to the registration process and the correspondence between these features is used to define the (dis)similarity[5]. Feature-based image registration has the advantage of lower processing time, as only a smaller number of features are used in the (dis)similarity calculation when compared to each pixel (or voxel) used in intensity-based registration[7].

Based on the nature of the transformation, registration methods can be classified into rigid-body and non-rigid registration. Rigid-body registration involves a combination of rotation and translation in order to bring the images into the same coordinate system[8]. However, if the images have geometric differences that cannot

Department of Surgery (Otolaryngology), Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC 3010, Australia. ✉email: kh.tohidulislam@gmail.com

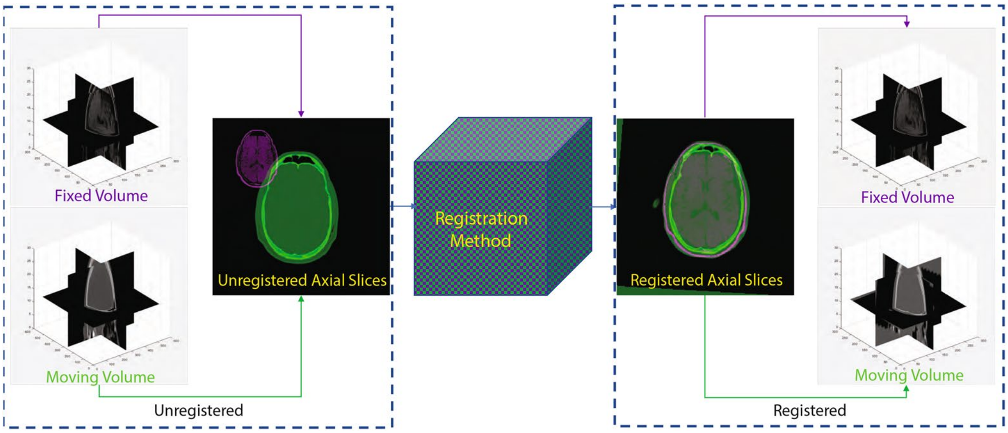


Figure 1. The process of 3D medical image registration.

be accounted for by using rotation and translation, rigid-body registration may not be sufficient. In such situations non-rigid (non-linear, deformable, or elastic) registration is used[9]. Demons registration[10], viscous fluid registration[11], B-spline registration[12], and finite element model (FEM) registration[13] are some common methods used in non-rigid image registration. The anatomical differences present in inter-patient (between different patients) medical images necessitate the use of non-rigid registration[14]. In contrast, rigid-body registration has been found to be adequate for intra-patient image registration (where the images are of the same patient taken at different times and/or using different imaging techniques)[15,16].

Over the last few decades, many methods have been introduced for medical image registration that typically use different cost functions and techniques. Mathematical optimisation as well as machine learning (both conventional methods and deep learning techniques) have been used for this purpose[17]. While the former is typically unsupervised, the latter can be either unsupervised or supervised. Optimization procedures minimise a distance function (or maximise a function that denotes similarity) between the fixed image and the moving image with respect to the registration (or transformation) parameters in an iterative process. Unsupervised machine learning techniques can also be used to optimise an objective function iteratively to find the registration parameters[18]. A limitation of these approaches is that the process has to be run from scratch for each new image pair. They are also computationally expensive and time-consuming[19,20]. In supervised machine learning, the model learns how a registration is performed through a training process involving pre-registered fixed and moving images. Although the training process could be resource intensive, once a model has been trained, the time taken for a registration is typically lower than that required for mathematical optimisation and unsupervised machine learning[21].

When employing traditional Machine learning methods for feature-based registration, features are required to be extracted prior to the registration process. The decision of which features to be used in the registration typically depend on the application/user. In contrast, deep learning techniques are able to learn the ideal features for the registration directly from the input data as part of the training process[6]. As such, many deep learning techniques have been introduced for medical image registration tasks[22–24].

Here, we explore the use of feature-based deep learning for 3D rigid-body registration of intra-patient, multi-modal (CT and MRI) images of the head. To this end, we first discuss a method of generating synthetic data to augment an existing registration dataset. Second, we describe a technique for ground truth generation for registration databases. Third, we introduce a registration method that incorporates the advantages of both conventional machine learning and deep learning. Fourth, we discuss a method to identify image modality in order to avoid manual intervention in selecting fixed and moving images. We validate the performance of our algorithms through experimental analyses on CT and MRI images of the head obtained from a publicly available medical image registration database.

Related work

There are several ways in which deep learning has been employed in feature-based supervised registration of 3D multi-modal images. Predominantly, researchers have trained deep regression models to predict the registration parameters[18,25,26]. Deep learning has also been used as a method of pre-processing, for example, to determine control points, which were then used to determine the registration parameters[27]. Another method that has utilised deep learning for this purpose was to train a model to predict the image of one modality given that of another modality of the same individual and registering this image with the original image of the corresponding modality, thereby reducing the problem to a single mode registration[28,29].

Chee and Wu[20] designed a deep learning model called AIRNet (affine image registration network) to predict the affine transformation parameters to register two 3D images. A twelve-element vector (a flattened version of the affine transformation matrix) was used as the output. First, they used an encoder (adapted from 2D DenseNet)[30] to extract features from the image pair. Then they concatenated these features and used them as input to several fully connected layers that performed the registration task in the form of regression. Other works, such as Chee and Wu[20] and Kori and Krishnamurthi[31] used a similar approach to predict affine transformation parameters by adapting a 2D VGG-19 network[32].

Sloan et al.[33] introduced a deep learning regression model to predict rigid-body transformation parameters of intra-patient T1 and T2 MRI images of the head. Initially, they trained their network using a series of synthetic mono-modal images, where the fixed and moving images were identical but for some changes in rotation and translation. Later on they re-trained their method to register multi-modal images. They used two different networks to predict the transformation parameters as two different pipelines using a CNN and a full convolution network (FCN). In these methods, they first extracted features from the fixed and moving images using the feature extraction layers of the networks and then concatenated these features before sending them through several convolutional and fully connected layers. They showed experimentally that their method performed well compared to other similar methods they considered. However, it remains to be seen how it performs when registering images of modalities with larger intensity differences such as MRI and CT.

Liu et al.[29] proposed a synthetic image generation based approach using deep learning for multi-modal medical image registration. They established that if an input image of one modality could be predicted from that of another modality, and this predicted image is then registered with the input image of the same modality, the registration process can be simplified into a mono-modal one. To this end, they used a CNN with 10 convolutional layers combined with ReLU and batch normalization layers in order to learn the complex feature mapping of an image of the input modality to its corresponding output modality in an image-to-image regression approach[34]. They used T1, T2 and PD (proton density) weighted MRI images to train their CNN model by using the sum of squared differences (SSD) as the cost function. Once this image was generated, they performed a mono-modal registration on that and the original image of the same modality using mathematical optimization. Liu et al.[29] compared their image generation method with other methods and stated that their method can produce more detailed synthetic images than others, which in turn can make the conventional registration task easier. Since their method was only tested on MRI images of different types, it is not clear how it will perform in registration between images of modalities with larger differences.

Zou et al.[27] implemented interest/control point and feature extraction based deep learning models for rigid-body medical image registration. First, they used a FCN to perform a pixel-level interest/control point calculation. Then, they used a CNN for feature detection and matching. Next, they used the random sample consensus (RANSAC) algorithm to filter outliers and determined the transformation matrix with the most inliers by iteratively fitting transforms[35]. They compared their performance with a traditional feature-based registration method: scale-invariant feature transform (SIFT) for MRI and CT images and showed their method performed well[36].

Miao et al.[26] used a CNN regression model to predict rigid body transformations. Their model was trained on synthetic images generated using a manual transformation. They showed experimentally that their model outperformed traditional intensity-based registration methods with respect to accuracy and computational efficiency. Similarly, Zheng et al.[37] also used a pre-trained VGG-type CNN architecture[32] to perform multi-modal medical image registration. They used their CNN as a regression model to predict parameters for rigid-body transformations. Initially, their model was trained on a large number of synthetic images and then fine-tuned using a small number of image sets. Their model outperformed similar existing state-of-the-art methods in terms of both accuracy and computational efficiency.

Proposed method

We consider four different aspects related to multi-modal 3D medical image registration and introduce methods for each. First, we discuss how to augment an existing registration database with synthetic images, as obtaining large databases for medical image registration is often impractical. Second, we address the problem of ground truth generation. In the absence of ground truth data generated by human experts, which may require considerable time commitments from them, researchers have used alternative methods such as using a validated algorithm[38,39]. Here, we introduce a method of increasing the accuracy of such validated algorithms when generating ground truth data for multi-modal medical image registration. Third, we introduce a registration method that comprises two steps: learning of image features best suited for the registration task (using deep learning) and determination of the registration parameters (using regression). Fourth, we discuss a method of identifying the image modality. This is an important task because the type of image determines whether it is used as the fixed or moving image, which is typically done manually. Identifying the image modality prior to registration enables the process to be fully automated. An overview of the proposed framework is shown in Fig. 2.

Data augmentation.

Data augmentation is a common practice in machine learning performed in order to increase the size of existing databases. It is typically done by generating synthetic data from the original data available in the database. The method used in generating synthetic images from the original dataset and the number of synthetic images to be generated are two important concerns in data augmentation. Many methods have been introduced in response to the former question, such as, random cropping[40], mixing images[41], generative adversarial networks[42], neural style transfer[43], and geometric transformations[44,45]. In our application, our focus is rigid body registration. As such, we generated synthetic images by rotating each image by a random angle (in the range of $[-15^0 \quad 15^0]$) around a random axis and translating it by a random distance (in the range of $[-5 \quad 5]$) along the coordinate axes. Not as much focus has been garnered by how many synthetic images should be generated for optimal training, and typically researchers have used arbitrary numbers that have performed well for their applications[40,45,46]. We employed an iterative training and testing procedure to identify the amount of synthetic data required to achieve best results. To this end, we generated an increasing number of synthetic data per original image (from 4 to 124 in steps of 5) and selected the number that provided best performance results.

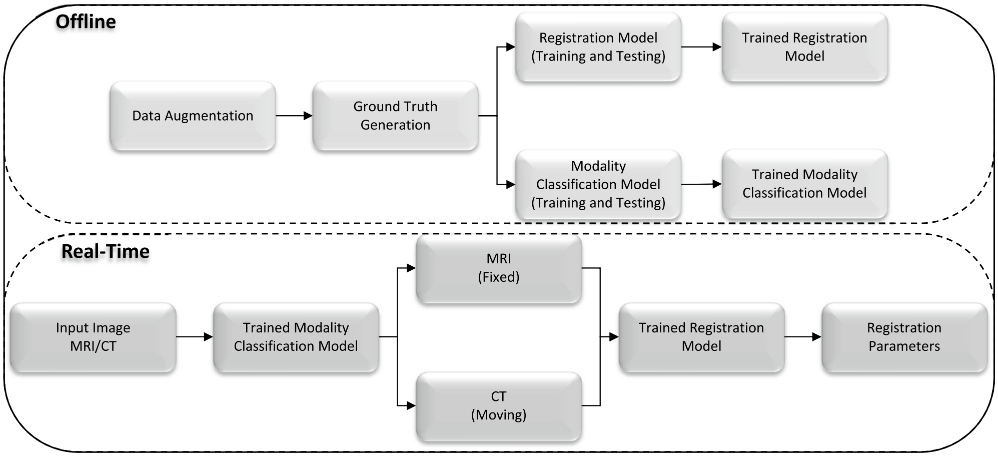

Offline
Data Augmentation
Ground Truth Generation
Registration Model (Training and Testing)
Trained Registration Model
Modality Classification Model (Training and Testing)
Trained Modality Classification Model
Real-Time
Input Image MRI/CT
Trained Modality Classification Model
MRI (Fixed)
CT (Moving)
Trained Registration Model
Registration Parameters

Figure 2. Overview of the proposed registration framework.

Generation of ground truth data.

The accuracy of a model generated using supervised learning is largely dependent on the quality of the ground truth data used in its training. As such, ground truth generation is an essential part of any supervised learning problem. In medical image processing, this is typically done manually by human experts such as radiologists[47]. However, this is a time consuming, expensive, and cumbersome process. To avoid human involvement, some researchers have used validated algorithms to generate ground truth data[20,48]. In registration problems, this typically takes the form of mathematical optimisation[38]. However, one of the issues in mathematical optimisation in registration is that if the orientations of the two images are not reasonably similar, it is possible that the process would result in a local (and not global) minimum.

To reduce the impact of this issue, we introduce a method that realigns the fixed and moving images of the head so that their orientations are closer together at the start of the optimisation. To this end, we use a characteristic inherent to most medical images of the human body: symmetry[49]. As the human head is roughly ellipsoidal, principal component analysis (PCA)[50] can be used as a simple, yet effective method of calculating symmetry. For images that have symmetry but are not ellipsoidal in nature, more complex symmetry calculation methods[51–53] can be used.

In order to calculate the PCA axes, we first resampled each image (fixed and moving) so that the scale of all three dimensions of the image were the same. This was done in order to preserve the real-world shape of the head. We then used Otsu's two-level global thresholding[54] method to remove the background points and extract a point cloud representing the head. We then calculated the principal components for the point cloud. Next, we reoriented the image so that its main axes were aligned along the principal components and the center of the image was at the mean location of the point cloud. This rigid body transformation can be represented by $\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$, where R is the rotation matrix and t is the vector of translation parameters. Once the fixed and moving images were brought to their 'symmetry' orientations using the transformations T_{Fix} and T_{Mov} respectively, we registered these two images using One Plus One Evolutionary (intensity-based) optimisation[55]. In the optimisation process, we used a growth factor of 1.05×10^{00}, epsilon of 1.5×10^{-06}, initial radius of 6.25×10^{-03}, maximum iterations of 100, number of spatial samples of 500, and number of histogram bins of 50. We considered all pixel values in the calculations. We then calculated the complete transformation that registered the original moving and fixed images as $T_{Comp} = T_{Fix}^{-1} T_{Reg} T_{Mov}$, where T_{Reg} is the rigid body transformation resulting from registration of the symmetry aligned images. Figure 3 illustrates the process of ground truth generation.

Registration of multi-modal images.

We trained a CNN for the registration of multi-modal images and used it to extract the features best suited for our task. Then, we trained a regression model on these features to predict the registration parameters. The architecture of the CNN used here is shown in Fig. 4. The image input layer used was 256 × 256 × 26 × 1. To speed up the training process and to reduce the sensitivity to network initialization, we used batch normalization layers after every convolution layer[56]. A Leaky Rectified Linear Unit (ReLU) layer was used after every batch normalization layer as a threshold operation to each element of the input, where any value less than zero was multiplied by a fixed scalar of 0.1[57]. Max pooling (3D) layers were used to perform down-sampling by dividing the 3D input into 3D pooling regions and computing the maximum of each region. A global average pooling layer was used before a fully connected layer to perform down-sampling by computing the mean of the height, width, and depth dimensions of the input and reduce the size of the activation without sacrificing performance. The final layer of the network was a regression output layer which computed the half-mean-squared-error loss. The output parameters were the three Euler angles representing the rotations around the three coordinate axes and the translations along the same. To train the network, we separated 30% of the data for validation and used stochastic gradient descent with a momentum (SGDM) optimizer of 0.9 and Max Epochs of 30[58]. We set the initial learning rate to 0.001 and dropped it to 0.0001 after 20 epochs.

Once the 3D DCNN was trained for registration, we used it to extract the features from the MRI and CT images. To this end, we activated the last convolution layer of the 3D DCNN (chosen using trial-and-error)

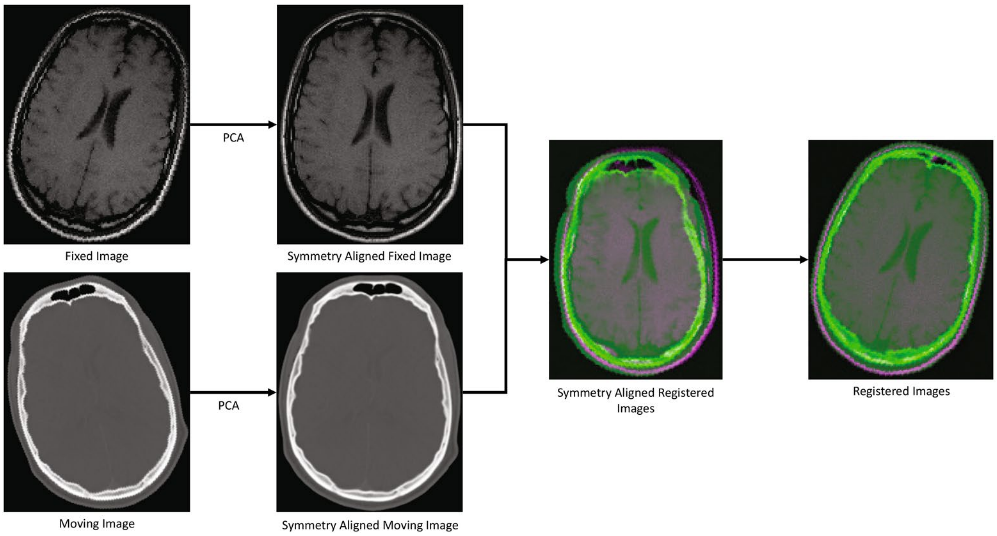

PCA
Fixed Image
Symmetry Aligned Fixed Image
PCA
Moving Image
Symmetry Aligned Moving Image
Symmetry Aligned Registered Images
Registered Images

Figure 3. Ground truth generation process.

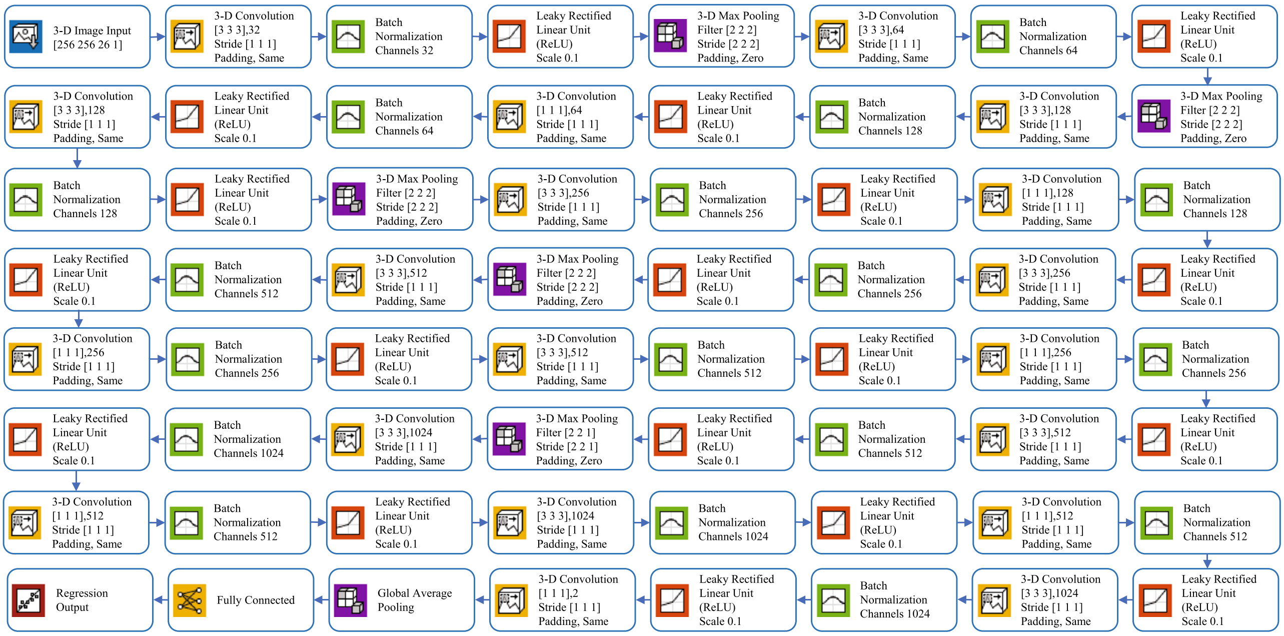

3-D Image Input [256 256 26 1]
3-D Convolution [3 3 3],32 Stride [1 1 1] Padding, Same
Batch Normalization Channels 32
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Max Pooling Filter [2 2 2] Stride [2 2 2] Padding, Zero
3-D Convolution [3 3 3],64 Stride [1 1 1] Padding, Same
Batch Normalization Channels 64
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Max Pooling Filter [2 2 2] Stride [2 2 2] Padding, Zero
3-D Convolution [3 3 3],128 Stride [1 1 1] Padding, Same
Batch Normalization Channels 128
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],64 Stride [1 1 1] Padding, Same
Batch Normalization Channels 64
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],128 Stride [1 1 1] Padding, Same
Batch Normalization Channels 128
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Max Pooling Filter [2 2 2] Stride [2 2 2] Padding, Zero
3-D Convolution [3 3 3],256 Stride [1 1 1] Padding, Same
Batch Normalization Channels 256
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],128 Stride [1 1 1] Padding, Same
Batch Normalization Channels 128
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],256 Stride [1 1 1] Padding, Same
Batch Normalization Channels 256
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Max Pooling Filter [2 2 2] Stride [2 2 2] Padding, Zero
3-D Convolution [3 3 3],512 Stride [1 1 1] Padding, Same
Batch Normalization Channels 512
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],256 Stride [1 1 1] Padding, Same
Batch Normalization Channels 256
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],512 Stride [1 1 1] Padding, Same
Batch Normalization Channels 512
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],256 Stride [1 1 1] Padding, Same
Batch Normalization Channels 256
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],512 Stride [1 1 1] Padding, Same
Batch Normalization Channels 512
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Max Pooling Filter [2 2 1] Stride [2 2 1] Padding, Zero
3-D Convolution [3 3 3],1024 Stride [1 1 1] Padding, Same
Batch Normalization Channels 1024
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],512 Stride [1 1 1] Padding, Same
Batch Normalization Channels 512
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],1024 Stride [1 1 1] Padding, Same
Batch Normalization Channels 1024
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],512 Stride [1 1 1] Padding, Same
Batch Normalization Channels 512
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [3 3 3],1024 Stride [1 1 1] Padding, Same
Batch Normalization Channels 1024
Leaky Rectified Linear Unit (ReLU) Scale 0.1
3-D Convolution [1 1 1],2 Stride [1 1 1] Padding, Same
Global Average Pooling
Fully Connected
Regression Output

Figure 4. Architecture of the proposed 3D deep convolutional neural network.

which extracted 128 features for each image (see Fig. 4). We then concatenated the features of the MRI and CT images and used them as input to a regression Artificial Neural Network (ANN). The number of input neurons of the ANN was 256 (128 each for MRI and CT image features). This ANN had one hidden layer containing 10 neurons (chosen using trial-and-error) and the final output was connected to 6 output neurons representing the 6 transformation parameters which we aimed to predict. This ANN was a feedforward network with hyperbolic tangent sigmoid transfer functions in the hidden layer and linear transfer functions in the output layer[59]. We selected Bayesian regularization backpropagation[60] as our training algorithm after comparing its performance with that of two others: Levenberg-Marquardt[61] and scaled conjugate gradient[62]. The final architecture of the proposed method (3D DCNN for feature extraction and ANN for regression) is shown in Fig. 5.

Identification of image modality. In order to make the proposed method fully automated, we developed a model to identify which of the input images was the fixed image and which was the moving image, and input them to the DCNN in the correct order. To this end, we used the same 3D DCNN architecture described above but replaced the fully connected layer with a softmax layer and the regression layer with a classification output layer. We used the same training strategy as above but changed the loss function to cross-entropy.

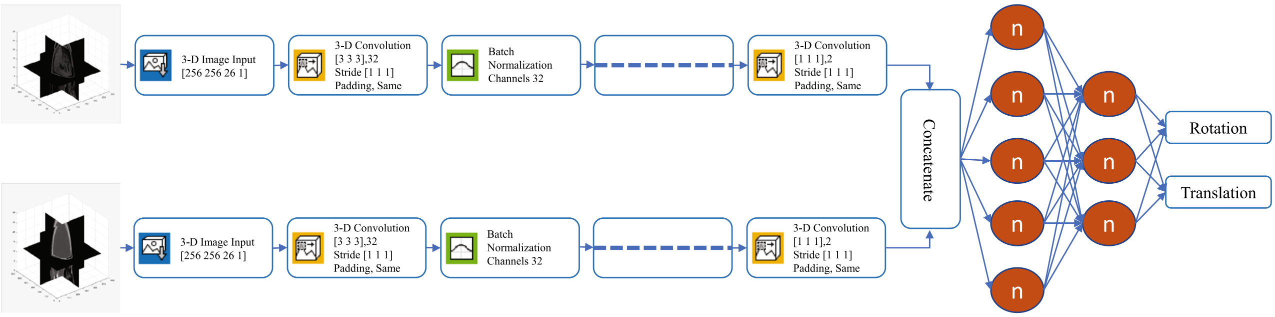


Figure 5. Complete architecture of the proposed method. For the full architecture of the DCNN, refer to Fig. 4.

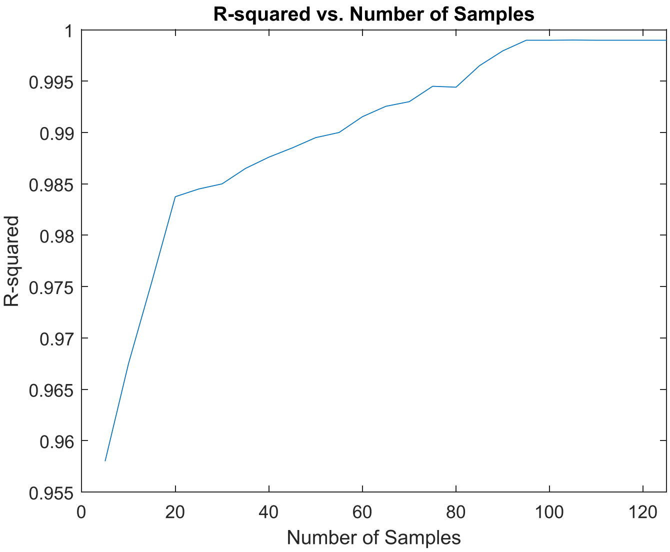


Figure 6. Change in network performance with the number of synthetic data.

Experimental results

Experimental setup. All methods were implemented using the MATLAB academic framework, including the Deep Learning Toolbox and Image Processing Toolbox. A Hewlett-Packard Z6 G4 Workstation model computer powered by Intel Xeno Silver 4108 CPU (1.80 GHZ) with 16 GB of physical memory and 5 GB of graphics memory (NVIDIA QuADro P2000 GPU) was used for running the experiments. The operating system used was 64-bit Microsoft Windows 10 Education. Also, part of this work (feature extraction) was carried out on the Spartan High Performance Computing (HPC) system[63].

We used the publicly available multi-modal 3D medical images provided by West et al.[64] as part of The Retrospective Image Registration Evaluation (RIRE) dataset (Dataset can be downloaded from: http://www.insight-journal.org/rire/download_data.php) in our experiments. This dataset contains 3D multi-modal images which were collected at the same time. Here, we used T1 weighted MRI images from this dataset as the fixed images and the corresponding CT images as the moving images. First, we removed artefacts from the data manually as a pre-processing step. This was done because, unlike in manual ground truth generation where artefacts would be ignored by the human expert, in automatic methods they introduce a measure of error into the results. We then resized all images to 256 × 256 × 26. Next, we split the dataset into training and test sets (60% and 40% respectively). As we used a freely available de-identified database, no ethics approval was required to carry out our experiments.

Data augmentation. As discussed above, we performed random rigid body transformations to generate different numbers of synthetic images in different steps ($n = 4, \ldots, 124$ in steps of 5). To avoid bias, we allocated each augmented image to the same subset (training or test) as that of the original image it was generated from. We then trained and tested the DCNN regression model discussed above on the resulting datasets. To observe the performance of the regression models we generated regression plots for each step and calculated the R^2 value (which describes the proportion of the variance of the predicted parameters explained by the regression model). As can be seen from the results shown in Fig. 6, the accuracy of the model plateaued at around $n = 94$. Thus, we chose $n = 99$ as the number of synthetic images generated for each original image in the dataset.

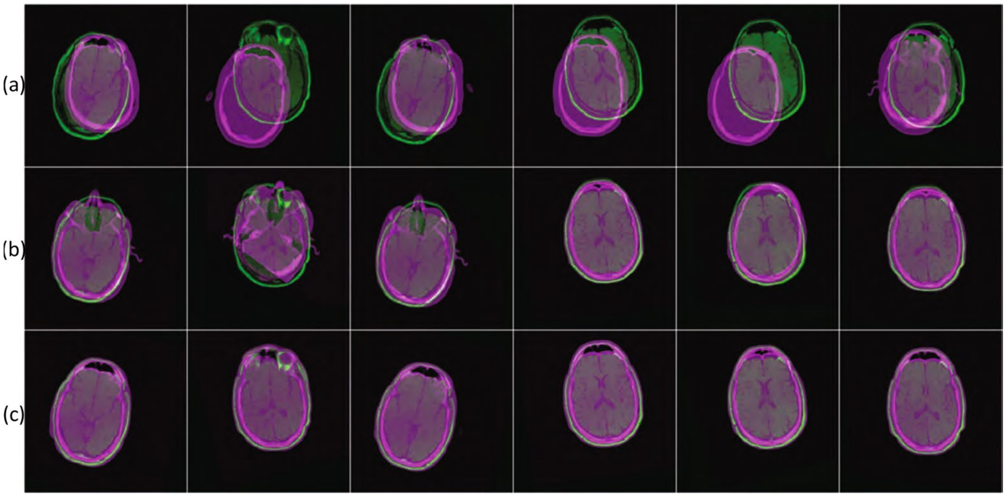


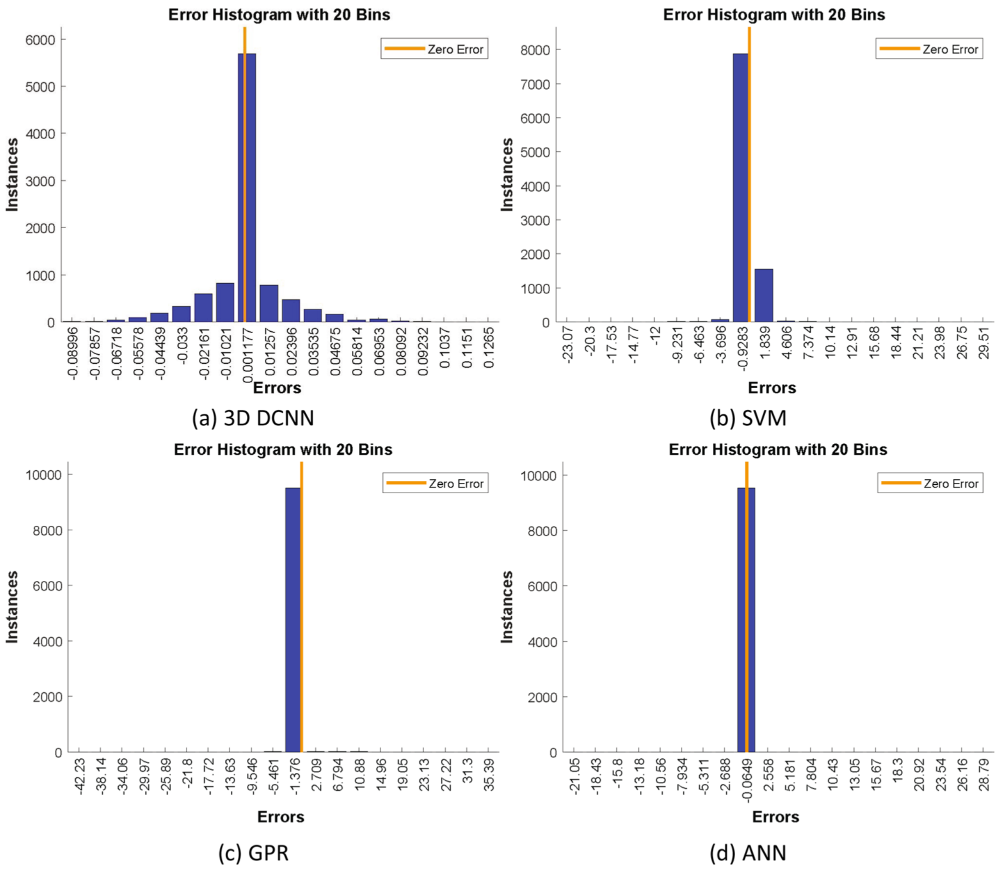
Figure 7. Example of registration without and with symmetry alignment. The rows show: (**a**) the fixed and moving images, (**b**) the results of the mathematical optimisation process without symmetry alignment, and (**c**) the results of the registration with symmetry alignment.

Generation of ground truth data. In the absence of manually obtained registration data, it is hard to determine the accuracy of the ground truth generation process using objective measures. Therefore, we conducted a manual assessment by a radiologist to evaluate the accuracy of the proposed ground truth generation process. To this end, we randomly selected 200 image pairs from the 1600 image pairs of our augmented dataset. We then displayed the axial, sagittal, and coronal slices of the registered image pairs obtained with and without the proposed modifications as methods A and B. The allocation of the method to A or B was randomised to remove bias. The assessor was asked to identify the best registration and given 3 options to choose from: A, B, or Both. The results of this assessment showed that 39.50% of the time, the 2 processes were similar, 28.00% of the time mathematical optimisation without symmetry alignment was better, and 32.50% of the time, mathematical optimisation with symmetry alignment was better. We show some visual results of this process in Fig. 7. Note from these results that in some cases (especially when the orientations of the fixed and moving images were close) the mathematical optimisation process performs well with or without symmetry alignment. However, in other instances (for example, where the orientations of the input images were vastly different), using symmetry alignment provided better registration results.

Multi-modal image registration. *Selection of regression model.* In order to select the best regression model for our application, we compared the performance of several models. First, we used the full 3D DCNN discussed above that was trained to predict registration parameters. Then, we used this DCNN as a feature extractor and combined it with several other regression models: Support vector machines (SVM), Gaussian process regression (GPR), and ANN. As SVM and GPR models were designed to predict one response at a time, we combined 6 prediction models to predict the 6 registration parameters. The resulting R^2 values for the regression models were: 0.958, 0.985, 0.990, and 0.998 for DCNN, SVM, GPR, and ANN respectively. We selected ANN as the proposed regression model as it gave the best R^2 value. Figure 8 shows the histograms of errors for the different regression models.

Complete registration process. We compared the performance of the proposed method with similar existing methods (Chee and Wu[20], Zheng et al.[37], Miao et al.[26], Sloan et al.[33], (CNN) Sloan et al.[33], (FCN) Liu et al.[29], and Zou et al.[27]) To this end, we generated the registration parameters from the trained networks, applied them to the moving image to get the registered image and then compared that to the ground truth (generated as discussed above). The metrics of comparison (for ground truth image x and predicted image y) were: Dice Similarity Coefficient (DSC), Jaccard Similarity Coefficient (JSC), Registration Precision (R_p), Registration Sensitivity (R_s), Contour Matching Score (CMS), and Structural Similarity Index Measure (SSIM)[65,66]. Equations 1 to 6 show the how these metrics were calculated. μ_x and μ_y represent the averages of the images, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the cross-covariance. C_1 and C_2 are regularization constants for the luminance and contrast respectively. In addition, we considered the elapsed time ($E_{t(s)}$), the total processing time (in seconds), as a comparison metric.

$$DSC(\mathbf{x},\mathbf{y}) = \frac{2\left|\mathbf{x}\cap\mathbf{y}\right|}{|\mathbf{x}| + \left|\mathbf{y}\right|} \tag{1}$$



(a) 3D DCNN

(b) SVM

(c) GPR

(d) ANN

Figure 8. Histograms of errors for different regression models.

$$JSC(\mathbf{x},\mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x} \cup \mathbf{y}|} \tag{2}$$

$$R_p(\mathbf{x},\mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{y}|} \tag{3}$$

$$R_s(\mathbf{x},\mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x}|} \tag{4}$$

$$CMS(\mathbf{x},\mathbf{y}) = \frac{2 \times R_p(\mathbf{x},\mathbf{y}) \times R_s(\mathbf{x},\mathbf{y})}{R_p(\mathbf{x},\mathbf{y}) + R_s(\mathbf{x},\mathbf{y})} \tag{5}$$

$$SSIM(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

Table 1 shows the comparison of the proposed method with other existing methods on the augmented RIRE dataset. The results show the average performance on four different random training and testing subsets. As can be seen from the results, the proposed algorithm outperformed the other methods in all the metrics considered.

We further tested the proposed method against the performance of the above DNNs when used as feature extractors. To this end, we activated their last convolution layers, concatenated the features extracted from the MRI and CT images, and used this as the input to an ANN regression model (see section 'Registration of Multi-Modal Images' for details). Table 2 shows the results of this comparison. Registration results for some example image pairs are shown in Fig. 9.

Identification of image modality. To evaluate fixed and moving image classification performance, we used commonly utilized metrics (accuracy, sensitivity, specificity, precision, f-measure, and g-mean)[67–69]. These metrics are defined in Equations 7–12 with respect to values of the confusion matrix: true positives (*TP*), true

Methodology	DSC	JSC	R_p	R_s	CMS	SSIM	$E_{t(s)}$
Chee and Wu[20]	0.9825	0.9760	0.9790	0.9805	**0.9875**	0.9680	10.55
Zheng et al.[37]	0.9850	0.9625	0.9600	0.9810	0.9725	0.9610	8.30
Miao et al.[26]	0.9785	0.9625	0.9570	0.9685	0.9690	0.9520	40.80
Sloan et al.[33] (CNN)	0.9780	**0.9765**	0.9760	0.9825	**0.9875**	0.9520	08.85
Sloan et al.[33] (FCN)	**0.9885**	0.9645	0.9825	0.9860	0.9850	0.9650	09.55
Liu et al.[29]	0.9580	0.9350	0.9610	0.9780	0.9680	0.9365	43.85
Zou et al.[27]	0.9670	0.9560	0.9820	0.9855	0.9780	0.9485	14.10
Proposed	**0.9885**	**0.9765**	**0.9830**	**0.9870**	**0.9875**	**0.9685**	**02.80**

Table 1. MRI and CT images registration performance. Best result for each metric is shown in bold.

Methodology	DSC	JSC	R_p	R_s	CMS	SSIM	$E_{t(s)}$
Chee and Wu[20]	0.9835	0.9780	0.9820	0.9815	0.9870	0.9685	9.50
Zheng et al.[37]	0.9865	0.9650	0.9630	0.9825	0.9730	0.9650	7.50
Miao et al.[26]	0.9790	0.9700	0.9610	0.9710	0.9730	0.9550	35.00
Sloan et al.[33] (CNN)	0.9785	0.9780	0.9765	0.9850	0.9890	0.9530	07.35
Sloan et al.[33] (FCN)	0.9895	0.9650	0.9835	0.9870	0.9855	0.9660	08.85
Liu et al.[29]	0.9610	0.9385	0.9670	0.9790	0.9700	0.9385	38.50
Zou et al.[27]	0.9680	0.9565	0.9835	0.9880	0.9785	0.9500	12.80
Proposed	**0.9910**	**0.9820**	**0.9903**	**0.9890**	**0.9930**	**0.9700**	**02.50**

Table 2. MRI and CT images registration performance (when the existing DNNs are used as feature extraction and coupled with a regression ANN). Best result for each metric is shown in bold.

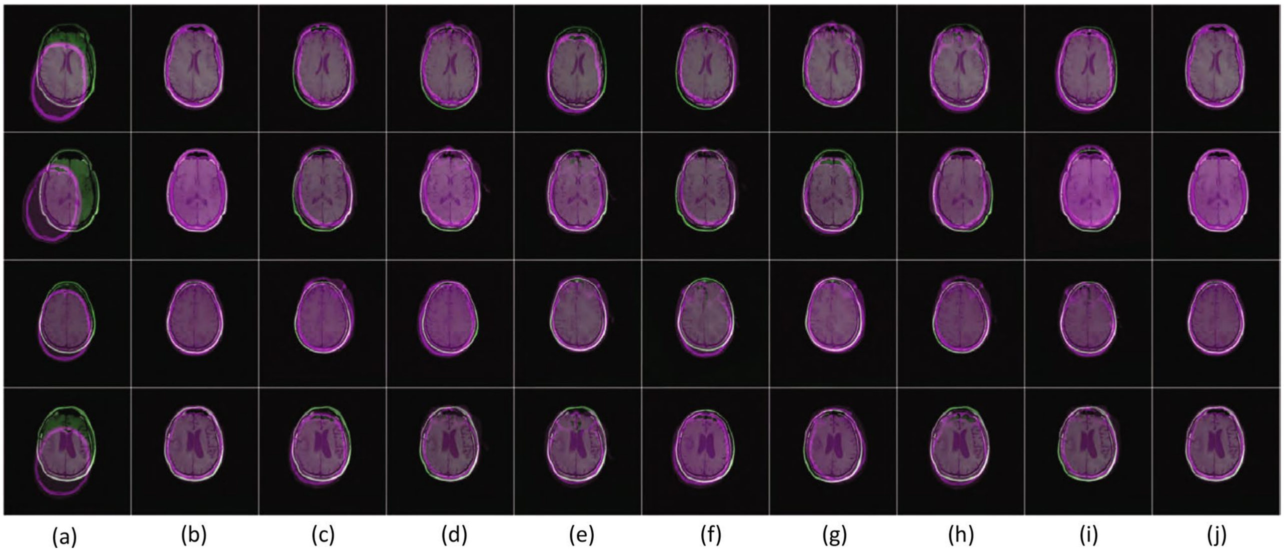


Figure 9. Examples of comparative performance of the registration methods. The columns show the axial views of: (**a**) input images, (**b**) ground truth, (**c**) Chee and Wu[20], (**d**) Zheng et al.[37], (**e**) Miao et al.[26], (**f**) Sloan et al.[33], (CNN) (**g**) Sloan et al.[33], (FCN) (**h**) Liu et al.[29], (**i**) Zou et al.[27], and (**j**) the proposed method.

negatives (*TN*), false positives (*FP*), and false negatives (*FN*) and results are shown in Table 3. We compared the performance of our method with some existing state-of-the-art pretrained architectures for image classification. The networks considered here were: AlexNet[70], VGG-19[32], SqueezeNet[71], GoogLeNet[72], and ResNet-101[73]. We modified the architectures of these networks so that they could accept 3D images as input.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

Architectures	Accuracy	Sensitivity	Specificity	Precision	F-measure	G-mean
AlexNet[70]	0.98520	0.97851	0.99500	0.99500	0.98625	0.98545
VGG-19[32]	0.98451	0.97632	0.99320	0.99530	0.98480	0.98785
SqueezeNet[71]	0.98475	0.97540	0.99885	0.99635	0.98645	0.98890
GoogLeNet[72]	0.99535	0.98965	0.99845	0.99845	0.99450	0.99480
ResNet-101[73]	0.99550	0.98975	0.99750	0.99750	0.99548	0.99565
Proposed	**0.99835**	**0.99565**	**0.99885**	**0.99885**	**0.99750**	**0.99750**

Table 3. Performance comparison for the identification of image modality. Best result for each metric is shown in bold.

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$F{-}Measure = 2 \times \left(\frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \right) \tag{11}$$

$$G{-}Mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \tag{12}$$

Conclusion

In this paper, we introduced a fully automated deep learning framework for 3D multi-modal medical image registration. To this end, we considered four aspects of image registration (data augmentation, ground truth generation, image registration, and identification of image modality) and introduced methods to address each. We validated the performance of the proposed methods on CT and MRI images of the head obtained from a publicly available database and showed that they outperformed similar existing methods. Although this framework was only tested on CT and MRI images of the head, it should be applicable to images of other organs and images of different modalities with minimal modifications (for example, more complex symmetry detection for data augmentation) and retraining.

Received: 13 August 2020; Accepted: 31 December 2020
Published online: 21 January 2021

References

1. Peng, H. *et al.* BrainAligner: 3d registration atlases of drosophila brains. *Nat. Methods* **8**, 493–498. https://doi.org/10.1038/nmeth.1602 (2011).
2. Wang, C.-W., Ka, S.-M. & Chen, A. Robust image registration of biological microscopic images. *Sci. Rep.* https://doi.org/10.1038/srep06050 (2014).
3. Kaur, B., Kaur, A. & Kaur, G. Applications of image registration: research cell?. *Int. J. Eng. Sci.* **17**, 192–198 (2016).
4. Losnegård, A. *et al.* Intensity-based volumetric registration of magnetic resonance images and whole-mount sections of the prostate. *Comput. Med. Imaging Graph.* **63**, 24–30. https://doi.org/10.1016/j.compmedimag.2017.12.002 (2018).
5. Cao, W., Lyu, F., He, Z., Cao, G. & He, Z. Multimodal medical image registration based on feature spheres in geometric algebra. *IEEE Access* **6**, 21164–21172. https://doi.org/10.1109/access.2018.2818403 (2018).
6. Haskins, G., Kruger, U. & Yan, P. Deep learning in medical image registration: a survey. *Mach. Vis. Appl.* **31**, 8. https://doi.org/10.1007/s00138-020-01060-x (2020).
7. McLaughlin, R. A. *et al.* A comparison of 2d–3d intensity-based registration and feature-based registration for neurointerventions. *Med. Image Comput. Comput. Assist. Interv. MICCAI.* https://doi.org/10.1007/3-540-45787-9_65 (2002).
8. Fookes, C. B. & Bennamoun, M. Rigid and non-rigid image registration and its association with mutual information: a review (2002).
9. Goshtasby, A., Staib, L., Studholme, C. & Terzopoulos, D. Nonrigid image registration: guest editors' introduction. *Comput. Vis. Image Underst.* **89**, 109–113. https://doi.org/10.1016/s1077-3142(03)00016-x (2003).
10. Peyrat, J.-M., Delingette, H., Sermesant, M., Xu, C. & Ayache, N. Registration of 4d cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons. *IEEE Trans. Med. Imaging* **29**, 1351–1368. https://doi.org/10.1109/tmi.2009.2038908 (2010).
11. Chang, H.-H. & Tsai, C.-Y. Adaptive registration of magnetic resonance images based on a viscous fluid model. *Comput. Methods Programs Biomed.* **117**, 80–91. https://doi.org/10.1016/j.cmpb.2014.08.004 (2014).
12. Rueckert, D. *et al.* Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* **18**, 712–721. https://doi.org/10.1109/42.796284 (1999).

13. Zhang, J., Wang, J., Wang, X. & Feng, D. The adaptive FEM elastic model for medical image registration. *Phys. Med. Biol.* **59**, 97–118. https://doi.org/10.1088/0031-9155/59/1/97 (2013).
14. Chen, M., Kanade, T., Pomerleau, D. & Schneider, J. 3-d deformable registration of medical images using a statistical atlas. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'99*, 621–630. https://doi.org/10.1007/10704282_67 (Springer Berlin Heidelberg, 1999).
15. Hahn, D. A. *et al.* A practical salient region feature based 3d multi-modality registration method for medical images. In Reinhardt, J. M. & Pluim, J. P. W. (eds.) *Medical Imaging 2006: Image Processing*. https://doi.org/10.1117/12.653071 (SPIE, 2006).
16. Muenzing, S. E., van Ginneken, B., Murphy, K. & Pluim, J. P. Supervised quality assessment of medical image registration: application to intra-patient CT lung registration. *Med. Image Anal.* **16**, 1521–1531. https://doi.org/10.1016/j.media.2012.06.010 (2012).
17. Zhang, Z. & Sejdić, E. Radiological images and machine learning: trends, perspectives, and prospects. *Comput. Biol. Med.* **108**, 354–370. https://doi.org/10.1016/j.compbiomed.2019.02 (2019).
18. de Vos, B. D. *et al.* A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **52**, 128–143. https://doi.org/10.1016/j.media.2018.11.010 (2019).
19. Li, H. & Fan, Y. Non-rigid image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:1709.00799 (2017).
20. Chee, E. & Wu, Z. Airnet: self-supervised affine registration for 3d medical images using neural networks. arXiv preprint arXiv:1810.02583 (2018).
21. Boveiri, H. R., Khayami, R., Javidan, R. & MehdiZadeh, A. R. Medical image registration using deep neural networks: a comprehensive review. arXiv preprint arXiv:2002.03401 (2020).
22. Wu, G. *et al.* Unsupervised deep feature learning for deformable registration of MR brain images. *Adv. Inf. Syst. Eng.*. https://doi.org/10.1007/978-3-642-40763-5_80 (2013).
23. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. https://doi.org/10.1016/j.media.2017.07.005 (2017).
24. Cheng, X., Zhang, L. & Zheng, Y. Deep similarity learning for multimodal medical images. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **6**, 248–252. https://doi.org/10.1080/21681163.2015.1135299 (2018).
25. Yang, X., Kwitt, R. & Niethammer, M. Fast predictive image registration. *Deep Learn. Data Label. Med. Appl.* https://doi.org/10.1007/978-3-319-46976-8_6 (2016).
26. Miao, S., Wang, Z. J. & Liao, R. A CNN regression approach for real-time 2d/3d registration. *IEEE Trans. Med. Imaging* **35**, 1352–1363. https://doi.org/10.1109/tmi.2016.2521800 (2016).
27. Zou, M. *et al.* Rigid medical image registration using learning-based interest points and features. *Comput. Mater. Contin.* **60**, 511–525 (2019).
28. Han, X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* **44**, 1408–1419. https://doi.org/10.1002/mp.12155 (2017).
29. Liu, X., Jiang, D., Wang, M. & Song, Z. Image synthesis-based multi-modal image registration framework by using deep fully convolutional networks. *Med. Biol. Eng. Comput.* **57**, 1037–1048. https://doi.org/10.1007/s11517-018-1924-y (2019).
30. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.243 (IEEE, 2017).
31. Kori, A. & Krishnamurthi, G. Zero shot learning for multi-modal real time image registration. arXiv preprint arXiv:1908.06213 (2019).
32. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
33. Sloan, J. M., Goatman, K. A. & Siebert, J. P. Learning rigid image registration—utilizing convolutional neural networks for medical image registration. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*. https://doi.org/10.5220/0006543700890099 (SCITEPRESS - Science and Technology Publications, 2018).
34. Santhanam, V., Morariu, V. I. & Davis, L. S. Generalized deep image to image regression. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.573 (IEEE, 2017).
35. Fischler, M. A. & Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395. https://doi.org/10.1145/358669.358692 (1981).
36. Lowe, D. Object recognition from local scale-invariant features. *Proc. Seventh IEEE Int. Conf. Comput. Vis.*https://doi.org/10.1109/iccv.1999.790410 *(1999).*
37. Zheng, J., Miao, S. & Liao, R. Learning CNNs with pairwise domain adaption for real-time 6dof ultrasound transducer detection and tracking from x-ray images. *Lect. Notes Comput. Sci.* https://doi.org/10.1007/978-3-319-66185-8_73 (2017).
38. Miri, M. S., Ghayoor, A., Johnson, H. J. & Sonka, M. Comparative study of multimodal intra-subject image registration methods on a publicly available database. In Gimi, B. & Krol, A. (eds.) *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*. https://doi.org/10.1117/12.2214209 (SPIE, 2016).
39. Wodzinski, M. & Müller, H. Learning-based affine registration of histological images. *Biomed. Image Registr.* https://doi.org/10.1007/978-3-030-50120-4_2 (2020).
40. Park, S., Baek Lee, S. & Park, J. Data augmentation method for improving the accuracy of human pose estimation with cropped images. *Pattern Recognit. Lett.* **136**, 244–250. https://doi.org/10.1016/j.patrec.2020.06.015 (2020).
41. Lin, W.-H., Zhong, J.-X., Liu, S., Li, T. & Li, G. ROIMIX: proposal-fusion among multiple images for underwater object detection. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp40776.2020.9053829 (IEEE, 2020).
42. Mahapatra, D. & Ge, Z. Training data independent image registration using generative adversarial networks and domain adaptation. *Pattern Recogn.* **100**, 107109. https://doi.org/10.1016/j.patcog.2019.107109 (2020).
43. Zanini, R. A. & Colombini, E. L. Parkinson's disease EMG data augmentation and simulation with DCGANs and style transfer. *Sensors* **20**, 2605. https://doi.org/10.3390/s20092605 (2020).
44. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data*. https://doi.org/10.1186/s40537-019-0197-0 (2019).
45. Kim, E. K., Lee, H., Kim, J. Y. & Kim, S. Data augmentation method by applying color perturbation of inverse PSNR and geometric transformations for object recognition based on deep learning. *Appl. Sci.* **10**, 3755. https://doi.org/10.3390/app10113755 (2020).
46. Uzunova, H., Wilms, M., Handels, H. & Ehrhardt, J. Training CNNs for image registration from few samples with model-based data augmentation. *Med. Image Comput. Comput. Assist. Interv. MICCAI* **2017**, 223–231. https://doi.org/10.1007/978-3-319-66182-7_26 (2017).
47. Fu, Y. *et al.* Deep learning in medical image registration: a review. arXiv preprint arXiv:1912.12318 (2019).
48. Zhang, X., Dong, H., Gao, D. & Zhao, X. A comparative study for non-rigid image registration and rigid image registration. arXiv preprint arXiv:2001.03831 (2020).
49. Vupputuri, A., Ashwal, S., Tsao, B. & Ghosh, N. Ischemic stroke segmentation in multi-sequence MRI by symmetry determined superpixel based hierarchical clustering. *Comput. Biol. Med.* **116**, 103536. https://doi.org/10.1016/j.compbiomed.2019.103536 (2020).
50. Pearson, K. L. I. I. on lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572. https://doi.org/10.1080/14786440109462720 (1901).

51. Tuzikov, A. V., Colliot, O. & Bloch, I. Evaluation of the symmetry plane in 3d MR brain images. *Pattern Recogn. Lett.* **24**, 2219–2233. https://doi.org/10.1016/s0167-8655(03)00049-7 (2003).
52. Cicconet, M., Hildebrand, D. G. C. & Elliott, H. Finding mirror symmetry via registration and optimal symmetric pairwise assignment of curves. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. https://doi.org/10.1109/iccvw.2017.206 (IEEE, 2017).
53. Nagar, R. & Raman, S. 3dsymm: robust and accurate 3d reflection symmetry detection. *Pattern Recogn.* **107**, 107483. https://doi.org/10.1016/j.patcog.2020.107483 (2020).
54. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* **9**, 62–66. https://doi.org/10.1109/tsmc.1979.4310076 (1979).
55. Styner, M., Brechbuhler, C., Szckely, G. & Gerig, G. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans. Med. Imaging* **19**, 153–165. https://doi.org/10.1109/42.845174l (2000).
56. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015).
57. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML* **30**, 3 (2013).
58. Robbins, H. & Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951).
59. Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T. & Alkon, D. L. Accelerating the convergence of the back-propagation method. *Biol. Cybern.* **59**, 257–263. https://doi.org/10.1007/bf00332914 (1988).
60. MacKay, D. J. C. Bayesian interpolation. *Neural Comput.* **4**, 415–447. https://doi.org/10.1162/neco.1992.4.3.415 (1992).
61. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168. https://doi.org/10.1090/qam/10666 (1944).
62. Møller, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **6**, 525–533. https://doi.org/10.1016/s0893-6080(05)80056-5 (1993).
63. Meade, B., Lafayette, L., Sauter, G. & Tosello, D. Spartan HPC-cloud hybrid. *Deliv. Perform. Flex.* https://doi.org/10.4225/49/58ead90dceaaa (2017).
64. West, J. B. *et al.* Comparison and evaluation of retrospective intermodality image registration techniques. In Loew, M. H. & Hanson, K. M. (eds.) *Medical Imaging 1996: Image Processing*, vol. 2710, 332–347, https://doi.org/10.1117/12.237936. International Society for Optics and Photonics (SPIE, 1996).
65. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* **11**, 178–189. https://doi.org/10.1016/s1076-6332(03)00671-8 (2004).
66. Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612. https://doi.org/10.1109/tip.2003.819861 (2004).
67. Japkowicz, N. Why question machine learning evaluation methods. *AAAI Workshop Eval. Methods Mach. Learn.* **2006**, 6–11 (2006).
68. Powers, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2**, 37–63 (2011).
69. Olson, D. L. & Delen, D. *Advanced Data Mining Techniques* (Springer, Berlin, 2008).
70. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, 1097–1105 (Curran Associates, Inc., 2012).
71. Iandola, F. N. *et al.* Squeezenet: alexnet-level accuracy with 50x fewer parameters and < 0.5 MB model size. arXiv preprint arXiv:1602.07360 (2016).
72. Szegedy, C. *et al.* Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2015.7298594 (IEEE, 2015).
73. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.90 (IEEE, 2016).

Acknowledgements

The authors would like to thank Dr. Bridget Copson of the Department of Medical Imaging at St. Vincent's Hospital, Melbourne, Australia, for conducting the manual assessment of the ground truth generation technique. This work was supported by the University of Melbourne and the Australian Government under the Melbourne Research Scholarship (MRS) and Research Training Program (RTP) Scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Conceptualization, K.T.I.; Data curation, K.T.I. and S.W.; Formal analysis, K.T.I., S.W., and S.W.; Funding acquisition, S.W. and S.O.; Investigation, K.T.I. and S.W.; Methodology, K.T.I. and S.W.; Project administration, S.W. and S.O.; Resources, S.W. and S.O.; Software, K.T.I. and S.O.; Supervision, S.W. and S.O.; Validation, K.T.I., S.W., and S.O.; Visualization, K.T.I. and W.W.; Writing—original draft, K.T.I.; Writing—review and editing, K.T.I., S.W. and S.O. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.T.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.