Contents

Learning image-based transformations via convolutional neural networks: a review

Nicholas J. Tustison[1], Brian B. Avants[1], and James C. Gee[2]

[1]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

[2]Department of Radiology, University of Pennsylvania, Philadelphia, PA

Corresponding author:

Nicholas J. Tustison

ntustison@virginia.edu

Abstract

Recent methodological innovations in deep learning and associated advancements in computational hardware have significantly impacted the various core subfields of quantitative medical image analysis. The generalizability, computational efficiency, and open-source availability of deep learning algorithms and related software, particularly those utlizing convolutional neural networks, have produced paradigm shifts within the field. This impact is evident from topical prevalance in the literature, conference and workshop themes, and winning methodologies in relevant competitions. In this work, we review the various state-of-the-art approaches to learning and prediction and/or optimizing image transformations using convolutional neural networks. Although of primary importance within the quantitative imaging domain, image registration algorithmic development, in the context of these deep learning strategies, has received comparatively less attention than its counterparts (e.g., image segmentation). Nevertheless, significant progress has been made in this particular subfield which has been presented in various research venues. We contextualize these contributions within the broader scope of deep learning advancements and, in so doing, attempt to facilitate the leveraging and further development of such techniques within the medical imaging research community.

Key words: deep learning, diffeomorphisms, image registration, spatial normalization

1 Introduction

Determining the spatial correspondence between imaging domains is frequently a critical component in quantitative image analysis workflows. The trajectory of image registration theoretical and technological development has led to increasingly high quality transformational mappings that have significantly improved performance in related processing tasks (e.g., image segmentation via joint label fusion [1]) and imaging-based statistical analysis involving template-based normalization (e.g., voxel-based morphometry [2] and sparse canonical correlation analysis [3]). Several reviews [4–9] have charted this chronology and provided insight into related issues such as algorithmic classification, available implementations, evaluation strategies, and speculation concerning possible future directions of the field. While prescient in many respects, such speculation vis-à-vis the resurgence of deep learning is understandably limited due to its recent explosion in popularity and research focus.

The foundational concepts that form the basis for contemporary deep learning research dates back decades (e.g., [10]). Since this early seminal work, major developmental milestones include the *Neocognitron*, an early neural network for character recognition [11], and convolutional neural networks ("CNNs" or "ConvNets") utilized in speech [12] and visual signal processing [13], largely inspired by the visual cell types of the feline visual cortex [14]. Historical neural networks are differentiated from their modern progeny by the deep, or "hidden," layering that characterizes current architectures and is the reason for the extreme performance gains seen in the contemporary literature. The training of such architectures is made computationally tractable with gradient-based optimization using backpropagation (first performed in [13]) and the advent of GPU-based hardware [15]. Uptake by both industry and academia alike is further facilitated through the various neural network open-source software platforms (e.g., Tensorflow [16] and Keras [17]).

A key event in the widespread adoption of CNNs was the 2012 ImageNet Large Scale Visual Recognition Challenge for object classification [18]. The winning entry, a CNN-based architecture colloquially known as *AlexNet* [19], reduced the error rate by almost half over other entries . The following years' competitions were dominated by CNN variants such as VGG [20], GoogLeNet [21], and ResNet [22] with performance ultimately exceeding human performance in 2015 [23]. Additional competition outlets including conference-based venues (e.g., NeurIPS) and community-based platforms, such as Kaggle[1], continue to highlight the salience of CNNs as paradigmatic solutions to computational problems. This is in addition to the sheer number of formal research reports discussed in the same conferences and published in dedicated journals. Notable reviews by key figures in the field include those of Yann LeCun, Yoshua Bengio, Geoffrey Hinton [15], and Jürgen Schmidhuber [25].

Early CNN-based research tailored to medical imaging dates back to the 1990s with classification tasks providing the majority of use cases (e.g., lung nodule classification [26, 27] and breast tissue differentiation [28, 29]). Despite the early adoption by certain research groups, widespread uptake did not occur until much later. Several deep learning overviews specific to medical imaging have been presented in the recent research literature

- in editorial form [30];
- specific to generative adverserial networks (GANs) [31];
- focusing on MRI [32] and specific to neuro applications [33];
- for issues related to radiation therapy [34];
- concentrating on applications [35]; and
- as general reviews [36–40].

Despite the thorough treatment contained in these reviews, discussion of chronological adoption within the community is limited. Regardless, one can informally gauge this evolution from utilization of alternative machine learning techniques to predominately CNN-based approaches from the various competitions held simultaneously with medical imaging conferences. For example, the annual Multimodal Brain Tumor Segmentation (BraTS) Challenge has taken place under the auspices of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) since 2012 wherein large sets of training data are provided to the competitors who attempt to perform a voxelwise labeling of the constituent components of tumors from multimodal MR image data. The winning entries from the first two years employed random forest classifiers for segmentation [41]. Although variations of the traditional random forest scheme continued to be well represented in the 2014 Challenge, convolutional neural networks made an appearance [42]. By 2018, CNN-based pipelines were, by far, the most common [43] with specific preference being that of the

[1]Following the 2017 ImageNet challenge, in which the vast majority of teams surpassed the 5% classification error rate threshold, the ImageNet organizers ceded management to the Kaggle community which maintains a running performance assessment in ostensible perpetuity [24].

U-net architecture [44, 45] which, as we describe below, features prominently in image registration.

Conspicuously, coverage of the topic of deep learning-based image registration, relative to the related algorithmic categories of image classification and segmentation, has not been as extensive in the reviews mentioned above, despite its prominence in the broader research literature. This disparity seems to be similarly reflected in the quantity of published research for those respective categories [31, 38]. This review is meant to address this disparity and thus provide an overview of the current state-of-the-art of this burgeoning subfield. We first provide a description of key network components that are crucial to certain image registration architectures, or perhaps which might find utility in future architectures. Second, we discuss current approaches to the deep learning image registration categorized in terms of these basic architectural components.

2 Background on relevant network architectures

Prior to describing the various image registration algorithms that have been recently proposed in the literature which incorporate elements of deep learning, we first describe some basic architectural components specifically relevant to such a discussion which include:

- convolutional neural networks,
- spatial transformer networks,
- diffeomorphic transformer networks, and
- siamese networks.

Should we discuss the following?

- Deformable convolutional networks [46]
- U-net [44]

Since all but a small subset of components can be included for discussion, we defer the interested reader to the thorough cited earlier in addition to pertinent textbooks (e.g., [47]) for additional information.

2.1 Convolutional neural networks

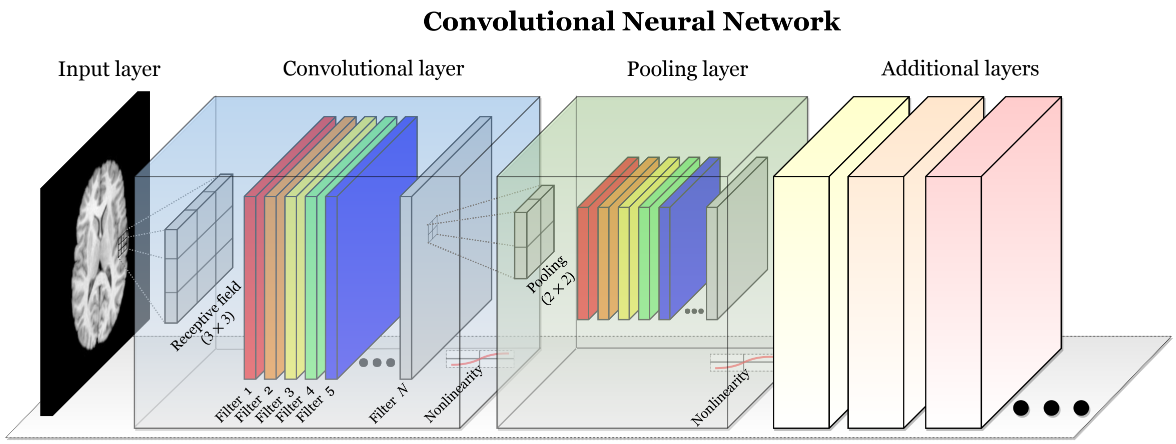

Figure 1: The basic elements of the CNN. The convolutional layer comprises several filters which are optimized in terms of their responses to various features found in the input layer. Pooling is used to extract salient features and reduce computational complexity and passed on to subsequent layers.

The grid-like informational content of certain data structures, such as 2-D and 3-D images, is perfectly suited to CNN-based training. The major elements of CNNs are localized convolutions, connections, and pooling [15]. As indicated by its name, the distinguishing characteristic of CNNs is the use of convolution instead of matrix operations in one or more of its constituent layers [47] where the output are feature maps. These feature maps are typically generated in an hierarchical fashion synthesizing simple geometric features at the base convolutional layers (lines, corners, etc.) progressing to more abstract features at the apical layers. The localized connections and weight-sharing provide a form of regularization while simultaneously reducing memory requirements [47]. The size of the convolution kernel, known as the "receptive field," determines the degree of localized connections. Finally, the accompanying pooling layers are used to subsample the convolutional feature maps in a way that statistically summarizes voxel neighborhoods within the feature maps. An illustration of a bare-bones CNN configuration is provided in Figure 1 which depicts the core components of convolution and max pooling. Architectural novelty derives from innovative arrangements of these core (and other) network components and the connections between them.

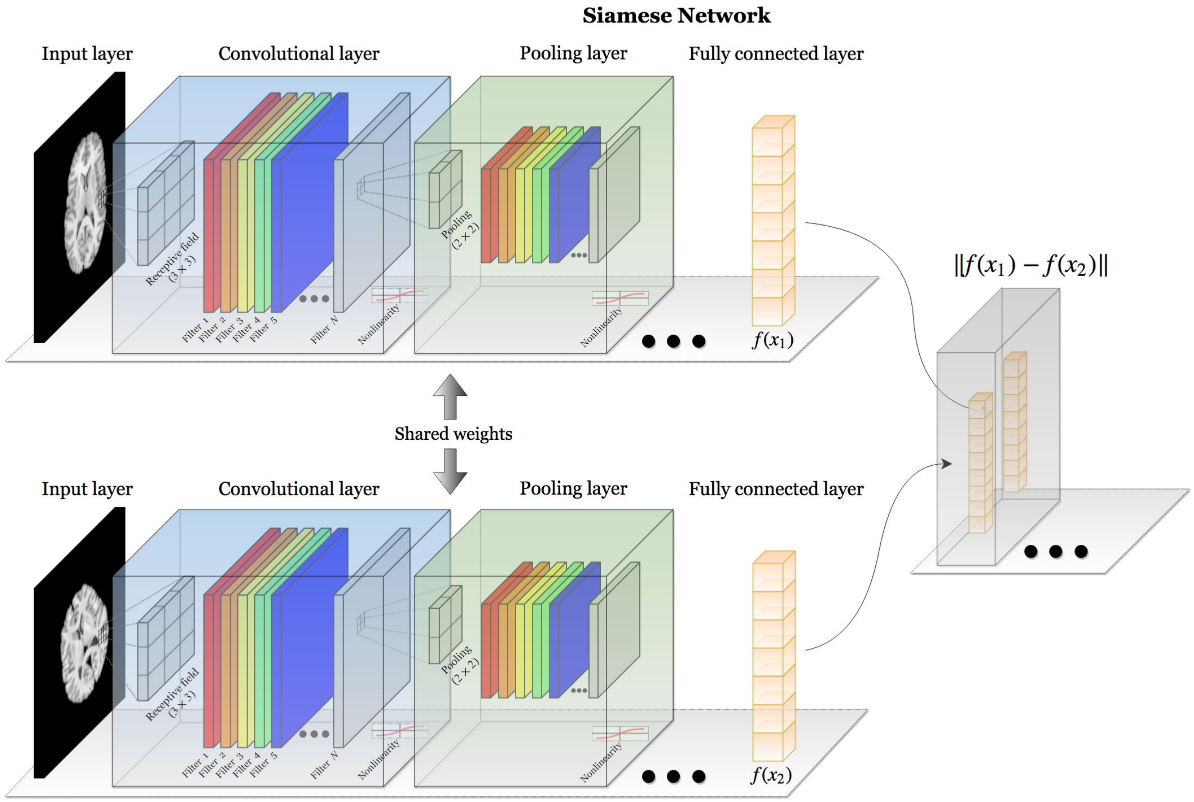

Figure 2: Illustration of a siamese architecture. Two identical convolutional branches, which traditionally share weights, are used to learn a similarity distance while simultaneously optimizing the input encoding captured in the fully connected layers.

2.2 Siamese networks

Any discussion of the evolution of image registration as a field would be incomplete without mention of the various similarity measures which have been proposed over the years (e.g., mutual information [6]). As we report below, these same similarity measures are often incorporated into CNNs as the loss function (or some component thereof). However, with deep learning, there are additional possibilities involving generating similarity functions using the input data without having to resort to basic statistical relationships between voxels and possibly their surrounding neighborhoods or explicitly designed features such as SIFT [48]. In terms of architectural elements, such learning is possible through so-called siamese networks [49, 50] illustrated in Figure 2. Siamese networks have identical input branches which feed into a decision layer involving some form similarity measure, oftentimes calculated from the fully connected encoding of input images.

Zagoruyko et al. discuss multiple architectures for comparing images (and patches) via CNNs [51], some of which are well-represented in the works reviewed below. The most basic architecture involves arrangement of the image pair as two channels in the input layer of the network. This two-channel network is reportedly fast to train but can be more computationally burdensome for testing. Both siamese and psuedo-siamese networks, respectively differentiated by shared vs. non-shared weights in the identical input branches, are also used for deep learning-based image registration.

2.3 Spatial transformer networks

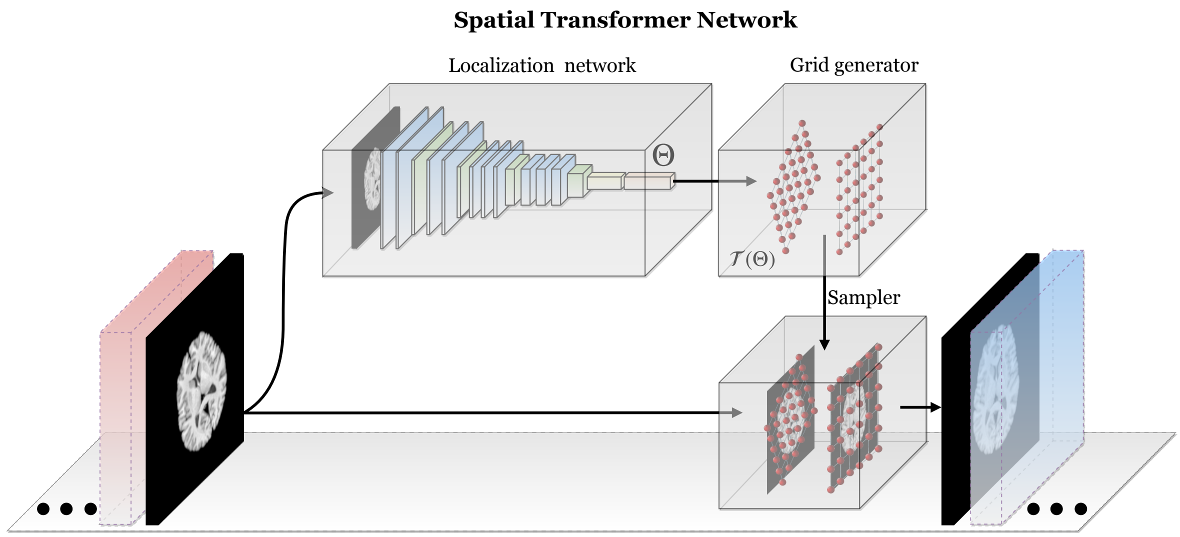

Figure 3: Diagrammatic illustration of the spatial transformer network. The STN can be placed anywhere within a CNN to provide spatial invariance for the input feature map. Core components include the localization network used to learn/predict the parameters which transform the input feature map. The transformed output feature map is generated with the grid generator and sampler.

In 2015 Jaderberg and his fellow co-authors described a powerful new module, known as the spatial transformer network (STN) [52] which figures prominently in many of the image registration approaches that we review below. Generally, STNs enhance CNNs by permiting 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 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 makes 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 Figure 3). 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 to facilitate, for example, classification. 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) [53] and the diffeomorphic transformer network [54]. We defer discussion of the latter to the next subsection but briefly describe the former. Two issues with STN include: 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 address 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 spirt of the inverse compositional Lucas-Kanade algorithm [55].

2.4 Deep diffeomorphic transformer networks

Although discussion of transform generalizability was included in the original STN paper [52], discussion was limited to affine, attention (scaling + translation), and thin-plate spline transforms which all fill the requirements of differentiability. This work was extended to diffeomorphic transforms in [54]. The computational load associated with generating traditional diffeomorphisms through velocity field integration [56] motivated the use of continuous piecewise affine-based (CPAB) transformations [57]. 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 [54]. Similar to the many publicly available implementations of STN, the authors provide their own Tensorflow implementation of the diffeomorphic transformer network.[2] The authors employ CUDA-based calculations for evaluating the CPAB gradients and transforms due to speed considerations.

2.5 Enhancing CNNs with CoordConv

Although not discussed let alone used in any of the papers reviewed below, the insight provided in [58] deserves consideration due to the subject matter of encoding spatial coordinates in CNN layers and its relevance to image registration. The authors describe a perplexing issue encountered during the course of their research. Reducing the core issue to toy examples, the authors demonstrate that training CNNs to regress cartesian coordinates from sparse, feature map pixel encodings (and vice versa) is highly problematic for conventional CNNs. In order to remedy this deficiency, the authors propose *CoordConv* which involves the modification of the conventional CNN layer with the concatenation of additional coordinate channels to the input. By explicitly encoding spatial information at each grid point in the input layer of the CNN, the authors improve performance not only in the toy examples but also in detection with the MNIST data set and in reinforcement learning scenarios involving video game play. Although not explicitly tested in the image registration problem domain, it is possible that such straightforward modifications to current architectures would substantially improve performance.

[2]https://github.com/SkafteNicki/ddtn

2.6 Generative adverserial networks

Goodfellow and colleagues introduced generative adverserial networks (GANs) in 2014 [59] which immediately demonstrated significant potential for generative modeling. Since their introduction, GANs have increasingly found traction in addressing many types of deep learning problems in the medical imaging domain [31] including image registration. GANs are a special type of network composed of two adverserial sub-networks 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 have exploded in the literature (see the GAN Zoo[3]). Initial extensions included architectural modifications for improved stability in training which have since become standard (e.g., deep convolutional GANs [60]).

3 Image registration with deep learning

The following overview of deep learning image registration methods is loosely categorized based on the discussion of network architectures given in the previous section. Specifically, we first discuss early work in which transformations were derived from CNN-based identification and localization of corresponding features in image pairs. We then review "two channel" approaches in which fixed and moving images are concatenated channelwise in the input layer. This leads to an overview of methods involving the related siamese and pseudo-siamese architectures. The final category concerns those adverserial approaches employing GANs. Other methods which do not fit in any of the above categories are also included. Each method is listed in Table 1 with a graphical summary provided in

[3]https://github.com/hindupuravinash/the-gan-zoo

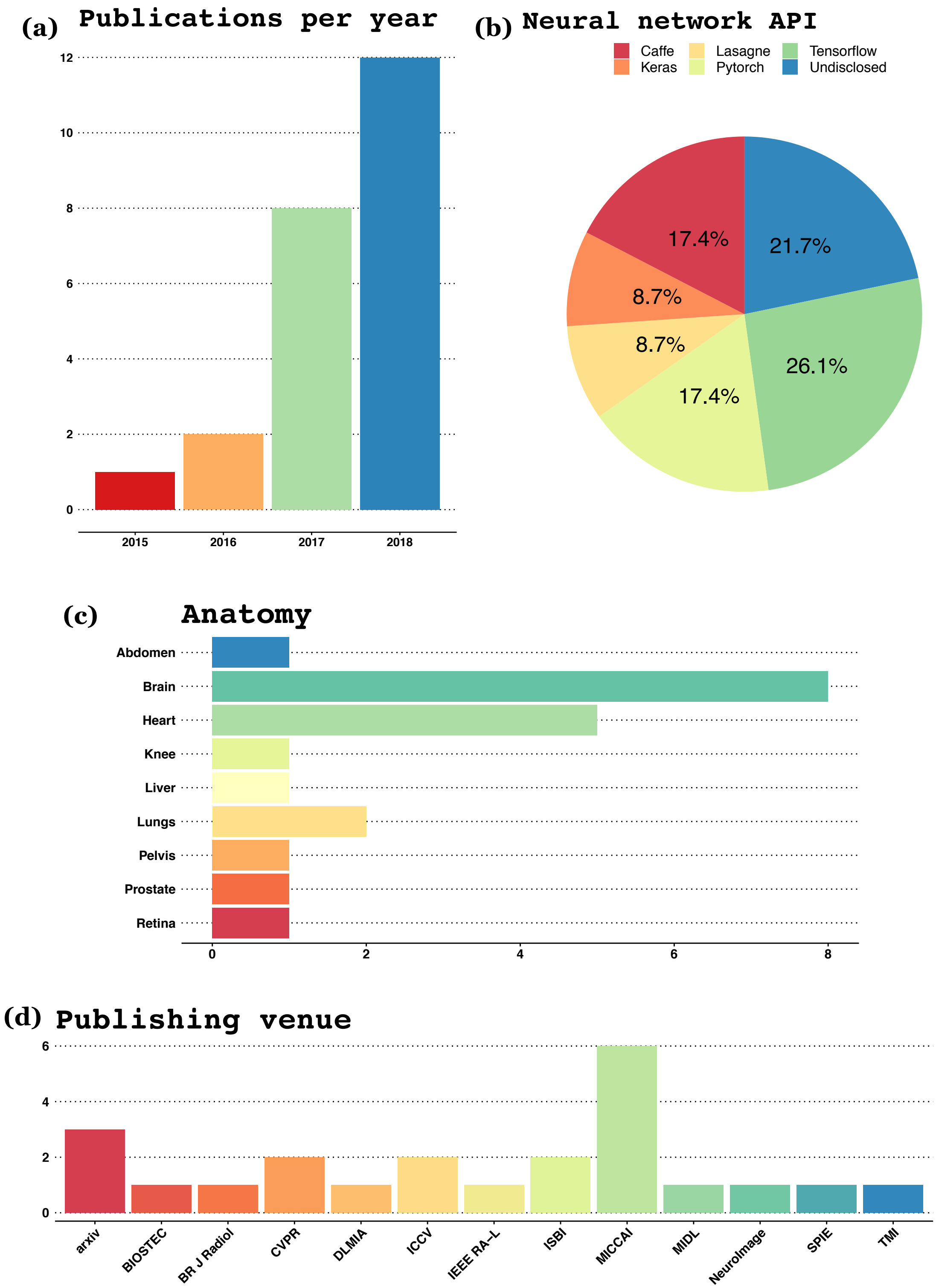

Figure 4: Graphical overview of the works reviewed including (a) publications per year, (b) choice of neural network API, (c) anatomy (where applicable), and (d) publishing venue.

Figure 4.

Table 1: Deep learning-based image registration methods organized in terms of basic network architecture.

Reference	Year	Dim.	Supervised?	†Transform	‡Similarity
			Feature localization		
Sergeev et al. [XX]	2012	3-D	Supervised	Affine	—
Simonovsky et al. [XX]	2016	3-D	Supervised	Deformable	—
Weinzaepfel et al. [XX]	2013	3-D	Supervised	Deformable	—
Wu et al. [XX]	2016	3-D	Supervised	Deformable	—
			Two channel		
Balakrishnan et al. [XX]	2018	3-D	Unsupervised	Deformable	CC
Cao et al. [XX]	2017	3-D	Supervised	Deformable	NCC
Dalca et al. [XX]	2018	3-D	Supervised	Diffeomorphic	MSQ
DeTone et al. [XX]	2016	2-D	Supervised	Homography	MSQ
deVos et al. [XX]	2017	2-D	Supervised	B-spline	NCC
Eppenhof et al. [XX]	2018	3-D	Supervised	TPS	None
Hu et al. [XX]	2018	3-D	Supervised	Affine + Deformable	Multiscale Dice
Lv et al. [XX]	2018	3-D	Supervised	Deformable	—
Nguyen et al. [XX]	2018	2-D	Unsupervised	Homography	L1
Rohe et al. [XX]	2017	3-D	Supervised	Diffeomorphic	None
Shan et al. [XX]	2017	2-D	Unsupervised	Deformable	MSQ
			Siamese/pseudo-siamese		
Dosovitskiy et al. [XX]	2015	2-D	Supervised	Optical flow	Correlation
Nowruzi et al. [XX]	2017	2-D	Supervised	Homography	MSQ
Rocco et al. [XX]	2017	2-D	Supervised	Affine + TPS	Learned
Sloan et al. [XX]	2018	2-D	Supervised	Rigid	None
Sokooti et al. [XX]	2017	3-D	Supervised	Deformable	Learned
Yang et al. [XX]	2018	3-D	Supervised	Diffeomorphic	None
Zhang et al. [XX]	2018	3-D	Supervised	IC Deformable	MSQ
			Generative adverserial networks		
Fan et al. [XX]	2018	3-D	Supervised	Deformable	Learned
Mahapatra et al. [XX]	2017	3-D	Supervised	Deformable	NMI/SSIM
Hu et al. [XX]	2018	3-D	Supervised	Deformable	
			Other		
Miao et al. [XX]	2016	2-D/3-D	Supervised	Rigid	
Sheikhjafari et al. [XX]	2018	2-D	Unsupervised	Deformable	AbsDiff

†TPS: thin-plate spline, IC: inverse consistent

‡CC: cross correlation, NCC: normalized CC, MSQ: mean squares, NMI: normalized mutual information, SSIM: structural similarity index, L1: L1 pixelwise photometric loss

3.1 Image registration via feature localization

Much of the early work incorporating deep learning into solving image registration problems involved the detection of corresponding features and then using that information to determine the correspondence relationship between the fixed and moving image pair. For example, just at the start of the current era of deep learning in image-related research, [61] 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 using the RANSAC algorithm [62]. Similarly, DeepFlow [63] uses CNNs to detect matching features (i.e., *deep matching*) which are then used as additional information in the large displacement optical flow framework [64]. A relatively small architecture is employed consisting of six layers and used to detect features at different convolution sizes and then matched across scales.

A similarity measure for multimodal registration is formulated in terms of CNNs in the work of [65]. A two channel network is developed for input image patches (T1- and T2-weighted brain images). A B-spline image registration algorithm developed from the Insight Toolkit is used to leverage the output CNN-based similarity measure for comparison with an identical registration set-up employing mutual information. Finally, in the category of feature learning, Wu et al. use stacked auto-encoders (SAE) to map patchwise image content to learned feature vectors [66]. These patches are then subsampled based on the importance criteria outlined in [67] which tends towards regions of high informational content such as edges. The SAE-based feature vectors at these image patches are then used to drive a HAMMER-based registration [68] which is inherently a feature-based, traditional image registration approach.

3.2 Two channel architectures for image registration

3.2.1 Voxelmorph

Voxelmorph was first introduced in [69] which incorporates a U-net architecture with spatial transformer network. 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 is the voxelwise displacement field of the same size as the input (times three for each vector component). The loss function for training combines cross correlation and regularization. This was significantly updated in [70] to utilize stationary velocity fields for generating diffeomorphic transforms which employed novel scaling and squaring network layers. The underlying code has been made available[4] which has facilitated independent evaluations such as [71] to compare performance with traditional algorithms (i.e., IRTK [72], AIR [73], Elastix [74], ANTs [75], and NiftyReg [76]).

3.2.2 Homography estimation

Two algorithms for more traditional computer vision applications are proposed in [77] and [78] where both are based on the VGG architecture [20] for 2-D 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 [78], which is publicly available[5], uses patch pairs as input and the L1 photometric loss between them to remove the need for direct supervision. A spatial transform layer is also adapted for homography transformations.

3.2.3 Training loss on ground truth transformations

Instead of training with a loss function based on similarity measures between fixed and moving images, the works of [79, 80] formulate the loss in terms of the squared difference between ground-truth and predicted transformation parameters. In terms of network architecture, [79] employs a

[4]https://github.com/voxelmorph/voxelmorph
[5]https://github.com/tynguyen/unsupervisedDeepHomographyRAL2018

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. In contrast, [80] uses a smaller version of the VGG architecture to learn the $6 \times 6 \times 6$ thin-plate spline grid.

3.2.4 Training loss on similarity metrics

Intermodality transformations involving CT and MRI are learned by training on the intramodality image pairs in [81]. The basic U-net architecture using input patches of size $68 \times 68 \times 68$ incorporates a loss function combining normalized cross correlation (NCC) and explicit regularization. A related idea is developed in [82] which uses labeled data and intensity information during the training phase such that only unlabeled image data is required for prediction. The architecture is a densely connected U-net architecture with three types of residual shortcuts. They also use a multiscale Dice function with an explicit regularization term in the loss function for estimating both global and local transformations.

Thsi one is pretty unique. Need to go over this more thorougly (Fig. 1).

```
* STN using B-splines

* Input is concatenated fixed/moving images

* use average pooling instead of max pooling to reduce translation invariance.

* Uses spatially corresponding image patches in a convnet regressor
```

- DIRNet is trained by optimizing an image similarity metric
- Implementation in Theano and Lasagne
- Loss function normalized cross correlation
- Unsupervised — doesn't use training. Learns to register by optimizing a similarity metric. No iterative optimization—just does a single pass. –>

3.3 Siamese and pseudo-siamese architectures for image registration

3.3.1 Geodesic shooting with Quicksilver

The large deformation diffeomorphic metric mappings (LDDMM) framework for image matching derives from the theoretical foundations underlying diffeomorphic *flows* [83–85]. Such diffeomorphisms are sufficiently differentiable bijective mappings, or transformations, which have sufficiently differentiable inverses. Specifically, the set of possible diffeomorphic mappings, $\phi(\mathbf{x}, t)$ ($\mathbf{x} \in \Omega$, $t \in [0, 1]$), between two images, I and J can be described as the collection of *paths* connecting the two images on a manifold determined by the equation

$$\int_0^1 \|v(t)\|_L^2 dt + \int_\Omega |I \circ \phi^{-1}(x, 1) - J|^2 d\Omega. \tag{1}$$

v is a time-dependent smooth field dictated by the functional norm L and determines the mapping via the ordinary differential equation

$$\frac{d\phi(\mathbf{x}, t)}{dt} = v(\phi(\mathbf{x}, t), t), \phi(\mathbf{x}, 0) = \mathbf{Id}. \tag{2}$$

The optimal diffeomorphic transformation between I and J can be described as a geodesic [56] connecting the two images. Traditionally, computational approaches to determining this geodesic path involve discretization of the velocity field followed by numerical integration. This is performed for a given number of iterations where, presumably, convergence implies arrival at this geodesic (i.e., optimal) path. Alternatively, based on the work of [86], the Euler-Lagrange equations for Equation (1) can be written as a system incorporating a "momentum" term. It was further demonstrated that the initial momentum determined the entire geodesic path. This alternative perspective engendered a new approach to determining the diffeomorphic solution between two images, known as *geodesic shooting* (e.g., [56, 87]). Although initially formulated in terms of scalar momenta [87], a vector formulation was proposed in [88] which tends towards superior numerical behavior.

The supervised deep learning technique of Yang et al. [89], known as *Quicksilver*, leverages this geodesic shooting/vector momentum optimization approach for determining optimal diffeomorphic transformations. The network architecture consists of two parallel encoders for separate fixed/moving image patches ($15 \times 15 \times 15$ voxels) feature learning. The output is then concatenated and sent through three identical decoder branches (one for each dimension) which comprises the inverse operations as the single encoder branch. Thus, the output consists of the predicted vector momentum map which, as described above, determines the total transformation. In order to improve accuracy of the predicted momentum maps, a follow-on correction network is also proposed. This correction network, trained by inverting the mapping produced by the predicted momentum and computing the residual error, is meant to account for large deformations across patch boundaries. Of note, Quicksilver, written in PyTorch [90], is one of the handful of algorithms surveyed which has been made publicly available[6].

3.4 Adverserial image registration approaches

In order to constrain the mapping between moving and fixed images, the GAN-based approach outlined in [91] combines a content loss term (which includes subterms for normalized mutual information, structural similarity [92], and a VGG-based filter feature L2-norm between the two images) with a "cyclical" adverserial loss. This is constructed in the style of [93] who proposed this GAN extension, viz., 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 [94] 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 [95] whereby anatomical labels are used to drive registration during training only. The generator is constructed from an encoder/decoder architecture based on ResNet blocks [22]. The prediction framework includes both localized tissue deformation and the linear coordinate-system-changes assocated with the ultrasound imaging acquisition.

[6] https://github.com/rkwitt/quicksilver

In [96], the discriminator loss is based on quantification of how well two images are aligned where the negative case derives 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. The registration network consists of a U-net type architecture which takes two 3-D image patches from the image pair as input and produces a patchwise displacement field. The discriminator network takes an image pair as input and outputs the similarity probability.

4 Discussion

In the introduction, we mentioned the relative lack of development in the field image registration. It is natural to ask why this is the case. Is it possible that deep learning is reducing the algorithmic contribution of image registration and corresponding development? For example, prior to the introduction of segmentation

For years, image registration has been an extremely important tool for quantitative medical image analysis, in part, due to the role it played in incorporating spatial prior information for subsequent processing. However, deep learning is capable of doing many of those things without reference to

Although DDTN

Acknowledgments

We gratefully acknowledge the support of the NVIDIA Corporation with the donation of two Titan Xp GPUs used in support of this research.

References

1. Iglesias, J. E. and Sabuncu, M. R. “**Multi-Atlas Segmentation of Biomedical Images: A Survey**” *Med Image Anal* 24, no. 1 (2015): 205–219. doi:10.1016/j.media.2015.06.012

2. Ashburner, J. and Friston, K. J. “**Voxel-Based Morphometry–the Methods**” *Neuroimage* 11, no. 6 Pt 1 (2000): 805–21. doi:10.1006/nimg.2000.0582

3. Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C., and Grossman, M. “**Dementia Induces Correlated Reductions in White Matter Integrity and Cortical Thickness: A Multivariate Neuroimaging Study with Sparse Canonical Correlation Analysis**” *Neuroimage* 50, no. 3 (2010): 1004–16. doi:10.1016/j.neuroimage.2010.01.041

4. Brown, L. G. “**A Survey of Image Registration Techniques**” *ACM Comput. Surv.* 24, no. 4 (1992): 325–376. doi:10.1145/146370.146374, Available at http://doi.acm.org/10.1145/146370.146374

5. Maintz, J. B. and Viergever, M. A. “**A Survey of Medical Image Registration**” *Med Image Anal* 2, no. 1 (1998): 1–36.

6. Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. “**Mutual-Information-Based Registration of Medical Images: A Survey**” *IEEE Trans Med Imaging* 22, no. 8 (2003): 986–1004. doi:10.1109/TMI.2003.815867

7. Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., and Gopinath, K. “**Brain Functional Localization: A Survey of Image Registration Techniques**” *IEEE Trans Med Imaging* 26, no. 4 (2007): 427–51. doi:10.1109/TMI.2007.892508

8. Viergever, M. A., Maintz, J. B. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. W. “**A Survey of Medical Image Registration - Under Review**” *Med Image Anal* 33, (2016): 140–144. doi:10.1016/j.media.2016.06.030

9. Keszei, A. P., Berkels, B., and Deserno, T. M. “**Survey of Non-Rigid Registration Tools in Medicine**” *J Digit Imaging* 30, no. 1 (2017): 102–116. doi:10.1007/s10278-016-9915-8

10. Ivakhnenko, A. G. “**Polynomial Theory of Complex Systems**” *IEEE Transactions on Sys-*

tems, Man, and Cybernetics SMC-1, no. 4 (1971): 364–378.

11. Fukushima, K. "**Neocognitron: A Self Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position**" *Biol Cybern* 36, no. 4 (1980): 193–202.

12. Waibel, A. "**Phoneme Recognition Using Time-Delay Neural Networks**" *Meeting of the Institute of Electrical, Information and Communication Engineers (IEICE).* (1987):

13. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jacke, L. D. "**Backpropagation Applied to Handwritten Zip Code Recognition**" *Neural Computation* 1, no. 4 (1989): 541–551.

14. Hubel, D. H. and Wiesel, T. N. "**Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex**" *J Physiol* 160, (1962): 106–54.

15. LeCun, Y., Bengio, Y., and Hinton, G. "**Deep Learning**" *Nature* 521, no. 7553 (2015): 436–44. doi:10.1038/nature14539

16. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. "**TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**" (2015):

17. Chollet, F. and others. "**Keras**"

18. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. "**ImageNet Large Scale Visual Recognition Challenge**" *International Journal of Computer Vision* 115, no. 3 (2015): 211–252.

19. Krizhevsky, A., Sutskever, I., and Hinton, G. E. "**ImageNet Classification with Deep Convolutional Neural Networks**" *Proceedings of the 25th International Conference on Neural Information Processing Systems* (2012): 1097–1105. Available at http://dl.acm.org/citation.cfm?id=

2999134.2999257

20. Simonyan, K. and Zisserman, A. "**Very Deep Convolutional Networks for Large-Scale Image Recognition**" *CoRR* abs/1409.1556, (2014): Available at http://arxiv.org/abs/1409.1556

21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. "**Rethinking the Inception Architecture for Computer Vision**" *CoRR* abs/1512.00567, (2015): Available at http://arxiv.org/abs/1512.00567

22. He, K., Zhang, X., Ren, S., and Sun, J. "**Deep Residual Learning for Image Recognition**" *CoRR* abs/1512.03385, (2015): Available at http://arxiv.org/abs/1512.03385

23. He, K., Zhang, X., Ren, S., and Sun, J. "**Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015):

24. Available at https://www.kaggle.com/c/imagenet-object-localization-challenge

25. Schmidhuber, J. "**Deep Learning in Neural Networks: An Overview**" *Neural Networks* 61, (2015): 85–117.

26. Lo, S. C., Freedman, M. T., Lin, J. S., and Mun, S. K. "**Computer-Aided Detection of Mammographic Calcifications: Pattern Recognition with an Artificial Neural Network**" *Proc. spie: Medical imaging: Image processing* 1898, (1992): 859–869.

27. Lo, S. C., Freedman, M. T., Lin, J. S., and Mun, S. K. "**Automatic Lung Nodule Detection Using Profile Matching and Back-Propagation Neural Network Techniques**" *J Digit Imaging* 6, no. 1 (1993): 48–54.

28. Chan, H. P., Lo, S. C., Sahiner, B., Lam, K. L., and Helvie, M. A. "**Computer-Aided Detection of Mammographic Microcalcifications: Pattern Recognition with an Artificial Neural Network**" *Med Phys* 22, no. 10 (1995): 1555–67. doi:10.1118/1.597428

29. Sahiner, B., Chan, H. P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D., and Goodsitt, M. M. "**Classification of Mass and Normal Breast Tissue: A Convolution Neural Network Classifier with Spatial Domain and Texture Images**" *IEEE Trans Med Imaging* 15, no. 5 (1996):

598–610. doi:10.1109/42.538937

30. Greenspan, H., Ginneken, B. V., and Summers, R. M. "**Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique**" *IEEE Trans Med Imaging* 35, no. 5 (2016): 1153–1159.

31. Yi, X., Walia, E., and Babyn, P. "**Generative Adversarial Network in Medical Imaging: A Review**" *arXiv preprint* (2018):

32. Mazurowski, M. A., Buda, M., Saha, A., and Mustafa R. Bashir. "**Deep Learning in Radiology: An Overview of the Concepts and a Survey of the State of the Art with Focus on Mri**" *Journal of Magnetic Resonance Imaging* (2018):

33. Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., and Lladó, X. "**Deep Convolutional Neural Networks for Brain Image Analysis on Magnetic Resonance Imaging: A Review**" *Artif Intell Med* (2018): doi:10.1016/j.artmed.2018.08.008

34. Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., Summers, R. M., and Giger, M. L. "**Deep Learning in Medical Imaging and Radiation Therapy**" *Med Phys* (2018): doi:10.1002/mp.13264

35. Ker, J., Wang, L., Rao, J., and Lim, T. "**Deep Learning Applications in Medical Image Analysis**" *IEEE Access* 6, (2018): 9375–9389.

36. Suzuki, K. "**Overview of Deep Learning in Medical Imaging**" *Radiol Phys Technol* 10, no. 3 (2017): 257–273. doi:10.1007/s12194-017-0406-5

37. Shen, D., Wu, G., and Suk, H.-I. "**Deep Learning in Medical Image Analysis**" *Annu Rev Biomed Eng* 19, (2017): 221–248. doi:10.1146/annurev-bioeng-071516-044442

38. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M. van der, Ginneken, B. van, and Sánchez, C. I. "**A Survey on Deep Learning in Medical Image Analysis**" *Med Image Anal* 42, (2017): 60–88. doi:10.1016/j.media.2017.07.005

39. Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M. K. "**Medical Image Analysis Using Convolutional Neural Networks: A Review**" *J Med Syst* 42, no. 11

(2018): 226. doi:10.1007/s10916-018-1088-1

40. Biswas, M., Kuppili, V., Saba, L., Edla, D. R., Suri, H. S., Cuadrado-Godia, E., Laird, J. R., Marinhoe, R. T., Sanches, J. M., Nicolaides, A., and Suri, J. S. "**State-of-the-Art Review on Deep Learning in Medical Imaging**" *Front Biosci (Landmark Ed)* 24, (2019): 392–426.

41. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C. R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K. M., Jena, R., John, N. M., Konukoglu, E., Lashkari, D., Mariz, J. A., Meier, R., Pereira, S., Precup, D., Price, S. J., Raviv, T. R., Reza, S. M. S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C. A., Sousa, N., Subbanna, N. K., Szekely, G., Taylor, T. J., Thomas, O. M., Tustison, N. J., Unal, G., Vasseur, F., Wintermark, M., Ye, D. H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., and Van Leemput, K. "**The Multimodal Brain Tumor Image Segmentation Benchmark (Brats)**" *IEEE Trans Med Imaging* 34, no. 10 (2015): 1993–2024. doi:10.1109/TMI.2014.2377694

42. "**Conference Proceedings of the 3rd MICCAI BraTS Challenge**" (2014):

43. "**Pre-Conference Proceedings of the 7th MICCAI BraTS Challenge**" (2018):

44. Ronneberger, O., Fischer, P., and Brox, T. "**U-Net: Convolutional Networks for Biomedical Image Segmentation**" *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* 9351, (2015): 234–241.

45. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T., and Ronneberger, O. "**U-Net: Deep Learning for Cell Counting, Detection, and Morphometry**" *Nat Methods* 16, no. 1 (2019): 67–70. doi:10.1038/s41592-018-0261-2

46. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. "**Deformable Convolutional**

Networks” *arXiv preprint arXiv:1703.06211* (2017):

47. Goodfellow, I., Bengio, Y., and Courville, A. “**Deep Learning**” (2016):

48. Lowe, D. G. “**Distinctive Image Features from Scale-Invariant Keypoints**” *International Journal of Computer Vision* 60, no. 2 (2004): 91–110. doi:10.1023/B:VISI.0000029664.99615.94, Available at https://doi.org/10.1023/B:VISI.0000029664.99615.94

49. Bromley, J., Guyon, I., LeCun, Y., Sckinger, E., and Shah, R. “**Signature Verification Using a ”Siamese” Time Delay Neural Network**” *NIPS* (1994):

50. Chopra, S., Hadsell, R., and LeCun, Y. “**Learning a Similarity Metric Discriminatively, with Application to Face Verification**” *Proceedings of the international conference of computer vision* (2005):

51. Zagoruyko, S. and Komodakis, N. “**Learning to Compare Image Patches via Convolutional Neural Networks**” *Proceedings of the ieee conference on computer vision and pattern recognition* (2015):

52. Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. “**Spatial Transformer Networks**” (2015):

53. Lin, C.-H. and Lucey, S. “**Inverse Compositional Spatial Transformer Networks**” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017):

54. Detlefsen, N. S., Freifeld, O., and Hauberg, S. “**Deep Diffeomorphic Transformer Networks**” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018):

55. Baker, S. and Matthews, I. “**Lucas-Kanade 20 Years on: A Unifying Framework**” *International Journal of Computer Vision* 56, no. 3 (2004): 221–255. doi:10.1023/B:VISI.0000011205.11775.fd, Available at https://doi.org/10.1023/B:VISI.0000011205.11775.fd

56. Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. “**Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms**” *International Journal of Computer Vision* 61, no. 2 (2004): 139–157.

57. Freifeld, O., Hauberg, S., Batmanghelich, K., and Fisher, J. W. “**Transformations Based on**

Continuous Piecewise-Affine Velocity Fields” *IEEE Trans Pattern Anal Mach Intell* 39, no. 12 (2017): 2496–2509. doi:10.1109/TPAMI.2016.2646685

58. Liu, R., Lehman, J., Molino, P., Such, F. P., Frank, E., Sergeev, A., and Yosinski, J. “**An Intriguing Failing of Convolutional Neural Networks and the Coordconv Solution**” *arXiv preprint* (2018):

59. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. “**Generative Adversarial Nets**” *Advances in neural in- formation processing systems* (2014):

60. Radford, A., Metz, L., and Chintala, S. “**Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks**” *Proceedings of the international conference on learning representations* (2016):

61. Sergeev, S., Zhao, Y., Linguraru, M. G., and Kazunori Okada. “**Medical Image Registration Using Machine Learning-Based Interest Point Detector**” *Proceeings of the spie* (2012):

62. Fischler, M. A. and Bolles, R. C. “**Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography**” *Comm. ACM.* 24, no. 6 (1981): 381–395.

63. Weinzaepfel, P., Revaud, J., Harchaoui, Z., and Schmid, C. “**DeepFlow: Large Displacement Optical Flow with Deep Matching**” *2013 ieee international conference on computer vision* (2013): 1385–1392. doi:10.1109/ICCV.2013.175

64. Brox, T. and Malik, J. “**Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation**” *IEEE Trans Pattern Anal Mach Intell* 33, no. 3 (2011): 500–13. doi:10.1109/TPAMI.2010.143

65. Simonovsky, M., Gutierrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. “**A Deep Metric for Multimodal Registration**” *Proceedings of the international conference on medical image computing and computer-assisted intervention* (2016):

66. Wu, G., Kim, M., Wang, Q., Munsell, B. C., and Shen, D. “**Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning**”

IEEE Trans Biomed Eng 63, no. 7 (2016): 1505–16. doi:10.1109/TBME.2015.2496253

67. Wang, Q., Wu, G., Yap, P.-T., and Shen, D. “**Attribute Vector Guided Groupwise Registration**” *Neuroimage* 50, no. 4 (2010): 1485–96. doi:10.1016/j.neuroimage.2010.01.040

68. Shen, D. and Davatzikos, C. “**HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration**” *IEEE Trans Med Imaging* 21, no. 11 (2002): 1421–39. doi:10.1109/TMI.2002.803111

69. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. “**An Unsupervised Learning Model for Deformable Medical Image Registration**” (2018):

70. Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. “**Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration**” *Medical image computing and computer assisted intervention – miccai 2018* (2018): 729–738.

71. Nazib, A., Fookes, C., and Perrin, D. “**A Comparative Analysis of Registration Tools: Traditional Vs Deep Learning Approach on High Resolution Tissue Cleared Data**” *arXiv preprint* (2018):

72. Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. “**Nonrigid Registration Using Free-Form Deformations: Application to Breast Mr Images**” *IEEE Trans Med Imaging* 18, no. 8 (1999): 712–21. doi:10.1109/42.796284

73. Woods, R. P., Mazziotta, J. C., and Cherry, S. R. “**MRI-Pet Registration with Automated Algorithm**” *J Comput Assist Tomogr* 17, no. 4 (–): 536–46.

74. Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. W. “**Elastix: A Toolbox for Intensity-Based Medical Image Registration**” *IEEE Trans Med Imaging* 29, no. 1 (2010): 196–205. doi:10.1109/TMI.2009.2035616

75. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. “**A Reproducible Evaluation of Ants Similarity Metric Performance in Brain Image Registration**” *Neuroimage* 54, no. 3 (2011): 2033–44. doi:10.1016/j.neuroimage.2010.09.025

76. Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S. “**Fast Free-Form Deformation Using Graphics Processing Units**” *Comput*

Methods Programs Biomed 98, no. 3 (2010): 278–84. doi:10.1016/j.cmpb.2009.09.002

77. DeTone, D., Malisiewicz, T., and Rabinovich, A. "**Deep Image Homography Estimation**" *arXiv:1606.03798* (2016):

78. Nguyen, T., Chen, S. W., Shivakumar, S. S., Taylor, C. J., and Kumar, V. "**Unsupervised Deep Homography: A Fast and Robust Homography Estimation Model**" *Proceedings of ieee robotics and automation letters* (2018):

79. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. "**SVF-Net: Learning Deformable Image Registration Using Shape Matching**" *Medical image computing and computer assisted intervention – miccai 2017* (2017): 266–274.

80. Eppenhof, K. A. J., Lafarge, M. W., Moeskops, P., Veta, M., and Pluim, J. P. W. "**Deformable Image Registration Using Convolutional Neural Networks**" (2018):

81. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M.-J., Wang, Q., and Shen, D. "**Deformable Image Registration Based on Similarity-Steered Cnn Regression**" *Med Image Comput Comput Assist Interv* 10433, (2017): 300–308. doi:10.1007/978-3-319-66182-7_35

82. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., Ourselin, S., Noble, J. A., Barratt, D. C., and Vercauteren, T. "**Weakly-Supervised Convolutional Neural Networks for Multimodal Image Registration**" *Med Image Anal* 49, (2018): 1–13. doi:10.1016/j.media.2018.07.002

83. Trouvé, A. "**Diffeomorphic Groups and Pattern Matching in Image Analysis**" *Int. J. Computer Vision* 28, (1995): 213–221.

84. Christensen, G. E., Rabbitt, R. D., and Miller, M. I. "**Deformable Templates Using Large Deformation Kinematics**" *IEEE Trans Image Process* 5, no. 10 (1996): 1435–47. doi:10.1109/83.536892

85. Dupuis, P., Grenander, U., and Miller, M. I. "**Variational Problems on Flows of Diffeomorphisms for Image Matching**" *Quarterly of Applied Mathematics* LVI, (1998): 587–600.

86. Miller, M. I., Trouvé, A., and Younes, L. "**Geodesic Shooting for Computational Anatomy**"

J Math Imaging Vis 24, no. 2 (2006): 209–228. doi:10.1007/s10851-005-3624-0

87. Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. "**Diffeomorphic 3d Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation**" *Int J Comput Vis* 97, (2012): 229–241.

88. Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. "**A Vector Momenta Formulation of Diffeomorphisms for Improved Geodesic Regression and Atlas Construction**" *Proc IEEE Int Symp Biomed Imaging* 2013, (2013): 1219–1222. doi:10.1109/ISBI.2013.6556700

89. Yang, X., Kwitt, R., Styner, M., and Niethammer, M. "**Quicksilver: Fast Predictive Image Registration—A Deep Learning Approach**" *Neuroimage* 158, (2017): 378–396. doi:10.1016/j.neuroimage.2017.07.008

90. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. "**Automatic Differentiation in Pytorch**" *NIPS-w* (2017):

91. Mahapatra, D., Antony, B., Sedai, S., and Garnavi, R. "**Deformable Medical Image Registration Using Generative Adversarial Networks**" *Proceedings of ieee 15th international symposium on biomedical imaging* (2018):

92. Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. "**Image Quality Assessment: From Error Visibility to Structural Similarity**" *IEEE Trans Image Process* 13, no. 4 (2004): 600–12.

93. Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. "**Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks**" *Computer vision (iccv), 2017 ieee international conference on* (2017):

94. Hu, Y., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., Emberton, M., Vercauteren, T., Noble, J. A., and Barratt, D. C. "**Adversarial Deformation Regularization for Training Image Registration Neural Networks**" *Proceedings of the international conference on medical image computing and computer-assisted intervention* (2018):

95. Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., Emberton, M., Noble, J. A., Barratt, D. C., and Vercauteren, T. "**Label-Driven Weakly-Supervised Learning for Mul-**

timodal Deformable Image Registration" *Proceedings of ieee 15th international symposium on biomedical imaging* (2018):

96. Fan, J., Cao, X., Xue, Z., Yap, P.-T., and Shen, D. "**Adversarial Similarity Network for Evaluating Image Alignment in Deep Learning Based Registration**" *Med image comput comput assist interv* (2018):