Yasir Ali Rasheed
15268
PhD (CS)
Supervised Transformation Estimation in Deep Learning

Sheeraz Ahmed
Department of Computer Science
Iqra National University,
Peshawar, Pakistan.
sheeraz.ahmad@inu.edu.pk

Yasir Ali Rasheed
Department of Computer Science,
Iqra National University,
Peshawar, Pakistan
Yasirkhan2@hotmail.com

Sheeraz Ahmed
Department of Computer Science
Iqra National University,
Peshawar, Pakistan.
sheeraz.ahmad@inu.edu.pk

Yasir Ali Rasheed
Department of Computer Science,
Iqra National University,
Peshawar, Pakistan
Yasirkhan2@hotmail.com

Sheeraz Ahmed
Department of Computer Science
Iqra National University,
Peshawar, Pakistan.
sheeraz.ahmad@inu.edu.pk

Yasir Ali Rasheed
Department of Computer Science,
Iqra National University,
Peshawar, Pakistan
Yasirkhan2@hotmail.com

Sheeraz Ahmed
Department of Computer Science
Iqra National University,
Peshawar, Pakistan.
sheeraz.ahmad@inu.edu.pk

Yasir Ali Rasheed
Department of Computer Science,
Iqra National University,
Peshawar, Pakistan
Yasirkhan2@hotmail.com

Abstract: Establishing image correspondence through robust image registration is important for many medical tasks such as fusion of image, creation of organ atlas and monitoring of increasing tumor is a very difficult task. Since the start of the current Deep Learning (DL) renaissance, the medical imaging research community has developed DL-based approaches and achieved state-of-the-art in many applications including image registration. The main scope of this survey is to provide a comprehensive review for the quick adoption of DL for image registration applications over the past few years in order highlight the challenges faced by the practitioners. This survey summarizes the past few years advancements in DL based medical image registration focusing on research challenges and relevant innovations. It also sheds light on the future research directions and identifies how to proceed further in this field.

Keywords—Deep Learning (DL), diffeomorphic, unimodal, convolutional

I. INTRODUCTION

Image registration is the process of changing various image datasets into a single system having matching image contents having an important uses in medicine field. Registration is needed while analyzing images obtained from different viewports at different instances or different sensors. [1]. Image registration was done manually by clinicians till now, in which quality of manual alignments mostly depends upon the ability of user, which can be clinically harmful. In order to address the issues related to manual registration, automatic registration has been developed. Even though different techniques for automatic registration of image have been discovered prior to the DL renaissance, it DL reshaped image registration research [2]. After the success of AlexNet in the ImageNet challenge of 2012 [3], DL has state-of-the-art performance in many computer vision tasks including: detection of object, extraction of features, segmentation, classification, denoising, and reconstruction of image [4].

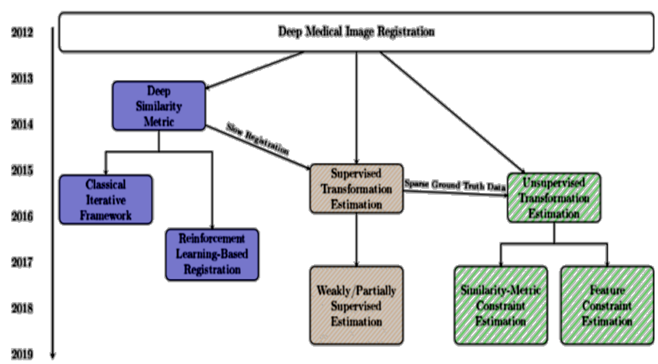

Figure 1: An overview of DL based medical image registration.

Figure 1 shows different categories of various DL based registration methods while Figure 2 shows the growing interest in DL-based registration methods by the number of research publications in past few years. The trends in Figures 1 and 2 shows that there is a lot of potential in this field and it is quickly removing the problems related to DL based medical image registration and many people have developed many successful applications [3, 5]. The red line in Figure 2 is a trend line for medical imaging based approach; blue line is the trend line for DL based medical image registration approach while dotted line is used for extrapolation.

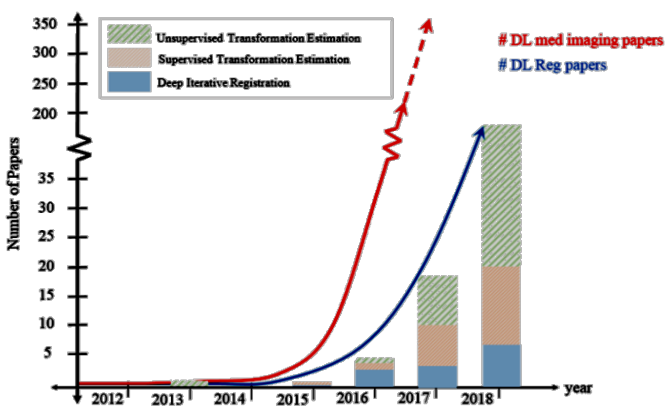

Figure 2: An overview of the number of DL based image registration and DL based medical imaging works.

This article is a comprehensive survey of DL-based medical image registration, identify the challenges faced by the practitioners and discuss the directions of the future research to address these challenges. DL is a class of machine learning that uses neural networks with a large number of layers for representations of data [5]. It is important to understand different types of neural networks used for a many applications, various newly invented architectures to deal with engineering problems, and different strategies used to training the neural networks. For this purpose, this DL introduction is divided into three parts: Types of Neural Network, Architectures of Network and Training Paradigms & Strategies. It is notable that to build the networks described in this section, many publicly accessible libraries are available [6]. In depth discussion of DL based medical image analysis and a range of DL research directions is not the scope of this article. Comprehensive review articles that survey the application of DL to medical image analysis [7], reinforcement learning [8], and the application of GANs to medical image analysis [9] are recommended for interested readers.

Table 1: Deep Iterative Registration Methods Overview.

Ref	Learning	Transform	Modality	ROI	Model
[10]	Metric	Deformable	CT	Lung	FCN
[11]	Metric	Deformable	CT/MR	Head	5-Layer DNN
[12]	Metric	Deformable	CT	Thorax	9-Layer CNN
[13]	Metric	Rigid	MR/US	Prostate	14-Layer CNN
[14]	RL Agent	Deformable	MR	Prostate	8-Layer CNN
[15]	RL Agent	Rigid	CT/ CBCT	Spine/ Cardiac	8-Layer CNN
[16]	RL Agent	Rigid	MR/CT	Spine	Dueling Network
[17]	Metric	Rigid	MR/US	Fetal Brain	LSTM/ STN
[18]	Multiple RL Agent	Rigid	X-ray/ CT	Spine	Dilated FCN
[19]	Metric	Rigid	MR/US	Abdominal	5-Layer CNN
[20]	Metric	Deformable	MR	Brain	5-Layer CNN
[21]	Metric	Deformable	MR	Brain	2-Layer CAE

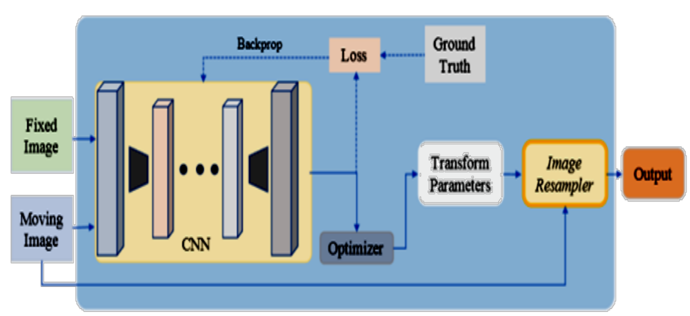

Figure 3: A visualization of the registration pipeline in an intensity-based registration framework.

II. DEEP ITERATIVE REGISTRATION

Automated intensity-based image registration requires a metric that balances the match between a moving and a static images and an optimization algorithm that updates the transformation parameters to maximize the match between these images. Prior to the DL renaissance, a number of manually crafted metrics were frequently used for such registration applications, including: sum of squared differences (SSD), cross-correlation (CC), mutual information (MI) [22], normalized cross correlation (NCC), and normalized mutual information (NMI). Initial applications of DL to medical image registration are direct extensions of the intensity-based registration framework [5, 23]. Later numerous groups used a reinforcement learning paradigm over and over again to estimate a transformation [1, 11, 22, 24] because this application is more relevant to how practitioners register..

2.1. Deep Similarity based Registration

In this section, methods that use DL to learn a matching metric are studied. This similarity metric is inserted into a classical intensity-based registration framework with a specified interpolation strategy, transformation model, and optimization algorithm. This overall framework is conceptualized in Figure 3. The solid lines represent flow of data which is needed during training and testing whereas dottedlines represent flow of data that are only needed during the training. Note that same is the case for the remainder of the figures in this article.

Overview of Works Manually crafted similarity metrics performance was quite well in the unimodal registration case,DL has been used to learn superior metrics. In this section, first we will describe the methods that use DL to increase the performance of unimodal intensity based registration pipelines before multimodal registration.

2.1.1.1. Unimodal Registration Authors in [23, 25] were the first to use DL in order to get an application-specific matching metric for registration. They extracted the features used for unimodal, deformable registration of 3D brain MR volumes using a Convolutional Stacked Autoencoder (CAE). Afterwards they have done the registration using gradient descent to enhance the NCC of the two sets of features which improved the registration techniques of diffeomorphic demons [26] and HAMMER [5].

In recent times, the researchers using end-to-end capacity [12], calculated registration error for the deformable registration of 3D thoracic CT scans (inhale-exhale). They used a 3D CNN to estimate the error map for given inhale-exhale pairs of thoracic CT scans. As mentioned in aforesaid method, only known features were used for this task.

Although the manually crafted descriptors performed better than the CNN-based descriptors, for lungs CT registration, paper [10] suggested to combinally use the CNN-based descriptors and manually crafted MRF-based self-similarity descriptors and best performance results were achieved. It is obvious that, in the unimodal registration case, DL may not outperform manually crafted methods while it can be used to get complementary information.

2.1.1.2. Multimodal Registration In multimodal case the benefits of the application of DL to intensity based registration are easier to understand, where manually crafted similarity metrics are rarely successful.

Recently[11], a stacked denoising autoencoder has been used to learn matching metric that evaluates the quality of the strict alignment of CT and MR images. They demonstrated that their metric improved the NMI-optimization-based and local cross correlation (LCC)-optimization-based for their application.

In an attempt to clearly visualize the image similarities in the multimodal case, authors [5] used a CNN to determine the difference between aligned 3D T1 and T2 weighted brain MR volumes. Given this matching metric, gradient descent was used in order to iteratively update the parameters that defines deformation field. This method has been successful in improving MI-optimization-based registration and paved the way for deep-intensity based multimodal registration.

Additionally, researchers in [27] performed the rigid registration of 3D US/MR (modalities with an even greater appearance difference than MR/CT) abdominal scans by using a 5-layer neural network to learn a similarity metric that is then optimized by Powells method. This approach also improved MI-optimization-based registration. Researchers [13] learned a similarity metric for multimodal rigid registration of MR and transrectal US (TRUS) volumes by using a CNN to predict target registration error (TRE). Instead of using aforementioned methods, they used an evolutionary algorithm to explore the solution space prior to using a traditional optimization algorithm because of the learned metric’s lack of convexity. This registration framework outperformed MIND-optimization-based [23] and MI-optimization-based registration. In contrast to the above methods, authors [22] used LSTM spatial co-transformer networks to iteratively register **MR** and **US** volumes group-wise. The recurrent spatial co-transformation took place in three phases: image warping, residual parameter prediction, parameter composition. Previous multimodal image similarity quantification method used self-similarity context descriptors therefore this method understands the image similarities better than previous one. [25].

2.1.1. Discussion and Assessment

Recently it is revealed that in multimodal medical image registration, the neural networks have ability to assess the similarities of image.. According to the results obtained from the methods mentioned in this section, DL can be successfully applied to challenging registration tasks. Furthermore for real time registration these iterative techniques are difficult to use.

III. Supervised Transformation Estimation

Although the previously described approaches were initially successful but as the transformation estimation in these methods was repetitive thus slowing down the registration [13]. This is especially true in cases of deformable registration where the solution space is more dimensional. [7]. This helped in the development of networks that can predict the transformation that corresponds to maximum matching in one step. However, fully supervised transformation estimation (exclusively using ground truth data to describe the loss function) has many problems that are discussed in this section. A visualization of supervised transformation estimation is given in Figure 5 and a description of notable works is given in Table 2.

3.1 Fully Supervised Transformation Estimation

In this section, the survey of the methods that use full supervision for single-step registration is performed. The use of neural network to perform registration as opposed to an iterative optimizer significantly speeds up the registration process.

3.1.1 Overview of works

Many registration applications involve deformable transformation models that normally not allow the usage of traditional convolutional neural networks due to computational expense linked with using FC-layers to make predictions in highly dimensional solution spaces [1]. The networks that are used to predict deformation fields are fully convolutional, the dimensionality of the solution space associated with a deformation field does not introduce additional computational constraints [4].

3.1.1.1 Rigid Registration Authors in [28] were the first to use DL to calculate rigid transformation parameters. They used a CNN to calculate the transformation matrix linked with the rigid registration of 2D/3D X-ray attenuation maps and 2D X-ray images. Hierarchical regression is adopted in which 6 transformation parameters are divided into 3 groups. In this approach, by changing the associated data, ground truth data was synthesized. Same is the case for the next three approaches described here. As compared to MI, CC, and gradient correlation (GC)-optimization-based registration approaches, this approach performed well in terms of both accuracy and computational efficiency. Better computational efficiency is obtained with the use of a forward pass through a neural network instead of an optimization algorithm to perform the registration.

The researchers [1] used a CNN to calculate the transformation parameters used to rigidly register 3D brain MR volumes. To train the network in this framework, affine image registration network (AIRNet), the MSE between the predicted and ground truth affine transforms were used and results showed that this was better than MI-optimization-based registration for both the unimodal and multimodal cases.

For atlas construction, the authors [27] used a deep residual regression network, a correction network and a bivariant geodesic distance based loss function to rigidly register T1 and T2 weighted 3D fetal brain MRs. The residual network is used to initially register the image volumes before forward pass through the correction network allowed for an enhancement of the capture range of the registration. This method was used for slice-to-volume and volume-to-volume registration methods. They authenticate the efficacy of their geodesic loss term and outperformed NCC-optimization-based registration.

Furthermore, researchers [26] proposed the integration of a Pair-wise Domain Adaptation module (PDA) into a pre-trained CNN that performs the rigid registration of pre-operative 3D X-Ray images and intra-operative 2D X-ray images using a limited amount of training data. Domain adaptation was used to address the inconsistency between synthetic data that was used to train the deep model and real data.

Researchers [29] used a CNN to revert the rigid transformation parameters for the registration of T1 and T2 weighted brain MRs. In this work both unimodal and multimodal registration were examined . The parameters that makeup the convolutional layers and were used to take out low-level features in all images were only shared in the unimodal case. These parameters were gathered separately in the multimodal case which outperformed MI-optimization-based image registration.

3.1.1.2 Deformable Registration In contrast to previous section, methods that use both real and synthesized ground truth labels will be discussed here. First we will discuss the methods that use clinical and publicly available ground truth labels for training. This order reflects the fact that simulation of realistic deformable transformations is more difficult than realistic rigid transformations.

This approach uses a U-net like architecture [12] along with a large diffeomorphic metric mapping to provide a basis, used the initial momentum values of the pixels of the image volumes as the network input, and evolved these values to obtain the predicted deformation field. This approach is better than the semi-coupled dictionary learning based registration [8].

The next year researchers [6] used a U-net [12] network to calculate the deformation field used to register 3D cardiac MR volumes. Mesh segmentations are used to calculate the reference transformation for a given pair of image and SSD between the prediction, also for loss function, ground truth is used . This method is better than LCC Demons based registration [30].

In same year, a CNN was used by authors [9] for mapping parts of input image a pair of 3D brain MR volumes to their respective displacement vector. To perform the registration, the total displacement vectors for a given image constitutes the deformation field. Furthermore, they used the similarities between input image patches to guide the learning process. They also used equalized active-points guided sampling strategy that makes them more likely to sample patches with higher gradient magnitudes and displacement values are more likely to be used for training and the results were better than SyN [31] and Demons [26] based registration methods.

Recently, authors [30] used a CNN to carry out the deformable registration of abdominal MR images to compensate for respiratory defects. This approach achieved registration results that are better than using non-motion optimized registrations and locally attached registrations. The researchers [25] justified the uncertainty associated with poor registration of 3D T1 and T2 weighted brain MRs using a low-rank Hessian estimation of the variable gaussian distribution of the transformation parameters. Real as well as synthetic data was used to evaluate this method..

Just as DL practitioners use random changes to increase the diversity of their dataset, paper [29] used random DVFs to expand their dataset. To predict a deformation field, a multi-scale CNN is used. This deformation is used for intra-subject registration of 3D chest CT images. This method used late fusion rather than early fusion, in which the patches are patched and used as the input to the network. The performance of this method is equal to B-Spline based registration [29].

Such an approach has significant, but limited ability to increase the size and range of datasets and these were the limitations that encouraged the development of more sophisticated ground truth generation. The other approaches explained in this section use simulated ground truth data for their applications.

Researchers [6] used a 3D CNN to perform the deformable registration of inhale-exhale 3D lung CT image volumes. A series of multi-scale, random transformations of aligned image pairs eliminate the need for manually annotated ground truth data while also maintaining realistic image appearance. In contrast with other methods, CNN can be trained using few medical images in a supervised capacity.

Contrary to the above the above researchers [13] generated ground truth data using Statistical Appearance Models (SAMs). They used a CNN to estimate the deformation field for the registration of 2D brain MRs and 2D cardiac MRs, and adapt FlowNet [24] for their application. It is proved that training FlowNet using SAM generated ground truth data, the performance results are better than CNNs trained using either randomly generated ground truth data or ground truth data obtained using the registration method described in [28].

In contrary to other methods in this section that generate ground truth data by using random transformations or manually crafted methods, the authors [31] used a CNN to learn plausible deformations for ground truth data generation. They assessed their approach on the 3D brain MR volumes in the ADNI dataset and results were better than the MI-optimization-based approach proposed in [2].

3.1.2 Discussion and Assessment

Supervised transformation estimation has allowed for real time, robust registration across applications with few limitations. The quality of the registrations using this framework depends on the quality of the ground truth registrations. Of course the quality of these labels depends on the skill of the practitioner. Since there are very few persons with expertise necessary to perform such registrations therefore, these labels are quite difficult to obtain. These limitations can be addressed by transformations of training data and the generation of synthetic ground truth data. Here it is necessary to make sure that the simulated data is abundantly similar to clinical data.

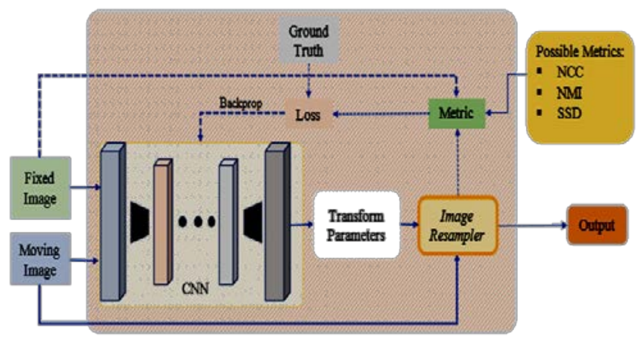

Figure 4: A visualization of deep single step registration.

3.2 Dual/Weakly Supervised Transformation Estimation

Dual supervision refers to the use of both ground truth data and some metric that quantifies image similarity to train a model. Alternatively, weak supervision refers to using the overlap of segmentations of corresponding anatomical structures to design the loss function.

3.2.1 Overview of works

Researchers [27] used hierarchical, dual-supervised learning to forecast the deformation field for 3D brain MR registration. They modify the traditional-Net architecture [12] through "gap-filling" (i.e., inserting convolutional layers after the Utype ends) and coarse-to-fine guidance. This approach took advantage of the similarities between predictive and ground truth changes and warped and fixed images to train the network.. The architecture used in this method improved the U-Net architecture and the dual supervision strategy is verified by removing the image similarity loss function term. A visualization of dual supervised transformation estimation is given in Figure 4.

Alternatively, in paper [25] a framework is used that was inspired by the GAN [29] to perform the rigid registration of 3D MR and TRUS volumes. Here the generator was trained to estimate a rigid transformation. However, the discriminator was trained to differentiate between images that were aligned using both ground truth transformations and the predicted transformations. In this method, both Euclidean distance to ground truth and an adversarial loss term are used to construct the loss function. Adversarial transformation estimation is given in Figure 5.

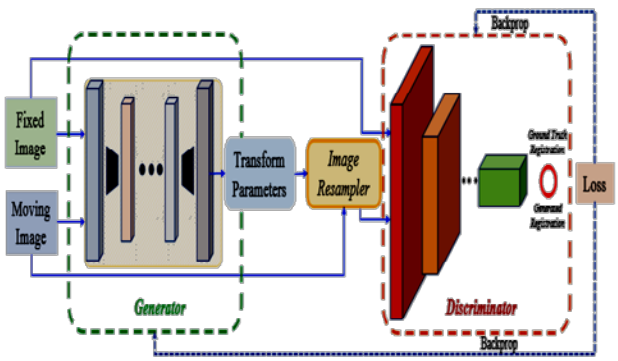

Figure 5: A visualization of an adversarial image registration framework.

Contrary to above methods that used dual supervision, authors [3, 32] in recent times used label similarity to train their network in order to perform MR-TRUS registration. Initially they used two neural networks: local-net and global-net to estimate the global affine transformation with 12 degrees of freedom and the local dense deformation field respectively [32]. The local-net uses the concatenation of the transformation of the moving image given by the global-net and the fixed image as its input. But later on [3], they combined these networks into end-to-end framework and thus outperformed NMI-optimization-based and NCC based registration. Figure 6 shows the weakly supervised transformation estimation. In another work researchers [33] simultaneously maximized label similarity and minimized an adversarial loss term to predict the deformation for MR-TRUS registration. This term of regularity forces the predicted transformation which resulted in the generation of a realistic image. Using the adversarial loss as a regularization term is likely to successfully force the transformation to be realistic given proper hyper parameter selection. The performance of this registration framework was less as compared to the performance of their aforementioned previous registration framework. However, they showed that adversarial regularization is better than standard bending energy based regularization. Similarly authors [4] built upon the progress made regarding both dual and weak supervision by introducing a label and similarity metric based loss function for cardiac motion tracking via the deformable registration of 2D cine-MR images. Both segmentation overlap and edge based normalized gradient fields distances were used to construct the loss function in this approach. Their method outperformed a multilevel registration approach similar to the one proposed in [27].

3.2.2 Discussion and Assessment

Direct transformation estimation marked a major breakthrough for DL based image registration. With full monitoring, encouraging results have been obtained. At the same time, however, these techniques require a large number of detailed illustrated images for training.. Partially/weakly supervised transformation estimation methods reduced the limitations associated with the trustworthiness and expense of ground truth labels but still require manually annotated data (e.g. ground truth and / or segmentations). In the multimodal case, the weak supervision allows for similarity quantification. In addition, partial supervision allows for the aggregation of methods that can be used to assess the quality of a predicted registration. It is a growing research field.

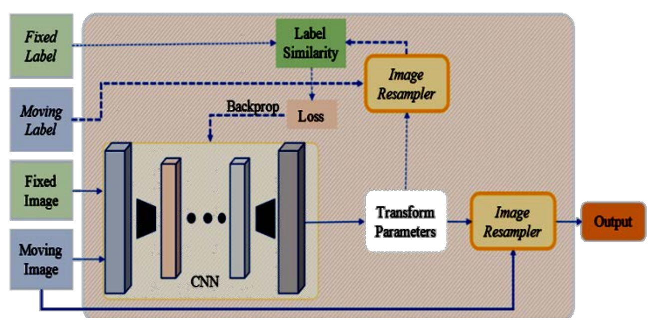

Figure 6: A visualization of deep single step registration where the agent is trained using label similarity.

IV. Research Trends and Future Directions

Figure 2 shows some emerging research trends.. It seems that DL-based medical image registration follows the observed trend of general application of DL to medical image analysis. Secondly, unsupervised transformation estimation methods have recently been given more attention by the research community. Also DL based methods are performing better than traditional optimization based techniques [6]. On the basis of observed research trends, it is hypothesized that the following research directions will receive more attention by the researchers.

4.1 Deep Adversarial Image Registration

We further speculate that GANs will be used more frequently in DL based image registration in the next few years. As described above, GANs can serve several different purposes in DL based medical image registration: using a discriminator as a learned similarity metric, ensuring that predicted transformations are realistic, and using a GAN to perform image translation to transform a multimodal registration problem into a unimodal registration problem.

GAN-like frameworks have been used in several works to directly train transformation predicting neural network. Several recent works [3, 25] use a discriminator to discern between aligned and misaligned image pairs. Although the training paradigm borrows from an unsupervised training strategy, the discriminator requires pre-aligned image pairs. Therefore, it will have limited success in multimodal or challenging unimodal applications where it is difficult to register images. Because discriminators are trained to assign all misaligned image pairs the same label, they will likely be unable to model a spectrum of misalignments. Despite this limitation, the application of GANs to medical image registration are still quite promising and will be described below.

Unconstrained deformation field prediction can result in warped moving images with unrealistic organ appearances. A common approach is to add the L2 norm of the predicted deformation field, its gradient, or its Laplacian to the loss function. However, the use of such regularization terms may limit the magnitude of the deformations that neural networks are able to predict, so authors [33] explored the use of a GAN-like framework to produce realistic deformations. Constraining the deformation prediction using a discriminator results in superior performance relative to the use of L2 norm regularization in that work.

Lastly, GANs can be used to map medical images in a source domain (e.g. MR) to a target domain (e.g. CT) [27, 30], regardless of whether or not paired training data is available [30]. This image appearance reduction technique would be advantageous because many unimodal unsupervised registration methods use similarity metrics that often fail in the multimodal case. If image translation is performed as a pre-processing step, then commonly used similarity metrics could be used to define the loss function of transformation predicting networks.

4.2 Raw Imaging Domain Registration

This article has focused on surveying methods performing registration using reconstructed images. However, we speculate that it is possible to incorporate reconstruction into an end-to-end DL based registration pipeline. In 2016, Wang [26] postulated that deep neural networks could be used to perform image reconstruction. Further, several works [4] recently demonstrated the ability of DL to map data points in the raw data domain to the reconstructed image domain. Therefore, it is reasonable to expect that registration pipelines that take raw data as input and output registered, reconstructed images can be developed within the next few years.

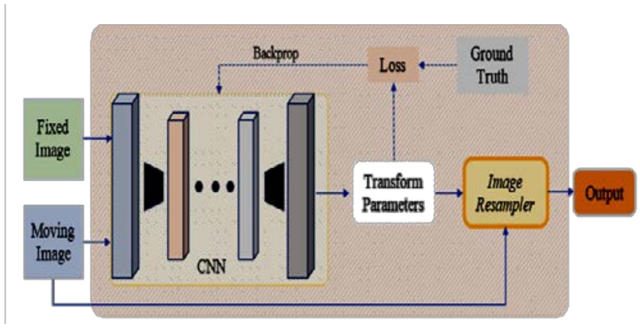

Figure 7: A visualization of supervised single step registration.

V. Conclusion

In this article, the recent works that use DL to perform medical image registration have been examined. As each application has its own unique challenges, the creation of the DL based frameworks must be carefully designed. Many DL based medical image registration applications share similar challenges including the lack of a robust similarity metric for multimodal applications, in which there are significant image appearance differences and/or different fields of view (e.g. MR-TRUS registration) [13], the lack of availability of large datasets, the challenge associated with obtaining segmentations and ground truth registrations, and quantifying the uncertainty of a model's prediction. Furthermore, despite the sophistication of many of the methods discussed in this survey, resampling and interpolation are often not among the components of registration that are learned by the neural network. Recent successes have demonstrated the impact of the application of DL to medical image registration. This trend can be observed across medical imaging applications. Many future exciting works are sure to build on the recent progress that has been outlined in this paper.

Table 2: Supervised Transformation Estimation Methods. Gray rows use Diffeomorphisms.

Ref	Supervision	Transform	Modality	ROI	Model
[9]	Real Transforms	Deformable	MR	Brain	9-layer CNN
[1]	Synthetic Transforms	Rigid	MR	Brain	AIRNet
[6]	Synthetic Transforms	Deformable	CT	Lung	U-Net
[27]	Real Transforms + Similarity Metric	Deformable	MR	Brain	U-Net
[4]	Segmentations + Similarity Metric	Deformable	MR/US	Prostate	U-Net GAN
[34]	Segmentations + Adversarial Loss	Deformable	MR/US	Prostate	GAN
[35]	Segmentations	Deformable	MR/US	Prostate	30-layer FCN
[36]	Synthetic Transforms	Deformable	MR	Brain	GoogleNet
[37]	Real Transforms	Deformable	MR	Abdominal	CNN
[38]	Synthetic Transforms	Rigid	X-ray/ DDR	Bone	6-layer CNN
[39]	Real Transforms	Deformable	MR	Cardiac	SVF-Net
[40]	Synthetic Transforms	Rigid	MR	Brain	11-layer CNN ResNet-18
[41]	Synthetic Transforms	Rigid	MR	Brain	6-layer CNN 10-layer FCN
[42]	Synthetic Transforms	Deformable	CT	Chest	RegNet
[43]	Synthetic Transforms	Deformable	CT/US	Liver	DVFNet
[44]	Synthetic Transforms	Deformable	MR	Brain/ Cardiac	FlowNet
[45]	Synthetic Transforms + Adversarial Loss	Rigid	MR/US	Prostate	GAN
[46]	Real + Synthetic Transforms	Deformable	MR	Brain	FCN
[47]	Real Transforms	Deformable	MR	Brain	FCN
[48]	Synthetic Transforms	Rigid	X-ray	Bone	17-layer CNN PDA Module

REFERENCES

[1] Chee, E. and Wu, J. (2018). Airnet: Self-supervised affine registration for 3d medical images using neural networks. *arXiv preprint arXiv:1810.02583*.

[2] Ambinder, E. P. (2005). A history of the shift toward full computerization of medicine. *Journal of oncology practice*, 1(2):54–56.

[3] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B. C., Awwal, A. A. S., and Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on DL approaches. *arXiv preprint arXiv:1803.01164*.

[4] Hering, A., Kuckertz, S., Heldmann, S., and Heinrich, M. (2018). Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking. *arXiv preprint arXiv:1812.01859*.

[5] Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *DL*, volume 1. MIT press Cambridge.

[6] Eppenhof, K. A. and Pluim, J. P. (2018a). Pulmonary ct registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*.

[7] Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., and Zhang, Z. (2015). Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274*.

[8] Cao, T., Singh, N., Jojic, V., and Niethammer, M. (2015). Semi-coupled dictionary learning for deformation prediction. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 691–694. IEEE.

[9] Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., and Shen, D. (2017). Deformable image registration based on similarity-steered cnn regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–308. Springer.

[10] Blendowski, M. and Heinrich, M. P. (2018). Combining mrf-based deformable registration and deep binary 3d-cnn descriptors for large lung motion estimation in copd patients. *International journal of computer assisted radiology and surgery*, pages 1–10.

[11] Cheng, X., Zhang, L., and Zheng, Y. (2018). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):248–252.

[12] Eppenhof, K. A. J. and Pluim, J. P. (2018b). Error estimation of deformable image registration of pulmonary ct scans using convolutional neural networks. *Journal of Medical Imaging*, 5(2):024003.

[13] Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P. A., Wood, B. J., and Yan, P. (2019). Learning deep similarity metric for 3d mr-trus image registration. *International Journal of Computer Assisted Radiology and Surgery*, 14:417–425.

[14] Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F. C., Miao, S., Maier, A. K., Ayache, N., Liao, R., and Kamen, A. (2017). Robust non-rigid registration through agent-based action learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 344–352. Springer.

[15] Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., and Comaniciu, D. (2017). An artificial agent for robust image registration. In *AAAI*, pages 4168–4175.

[16] Ma, K., Wang, J., Singh, V., Tamersoy, B., Chang, Y.-J., Wimmer, A., and Chen, T. (2017). Multimodal image registration with deep context reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 240–248. Springer.

[17] Matthew, J., Hajnal, J. V., Rueckert, D., and Schnabel, J. A. (2018). Lstm spatial co-transformer networks for registration of 3d fetal us and mr brain images. In *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*, pages 149–159. Springer.

[18] Miao, S., Piat, S., Fischer, P., Tuysuzoglu, A., Mewes, P., Mansi, T., and Liao, R. (2017). Dilated fcn for multi-agent 2d/3d medical image registration. *arXiv preprint arXiv:1712.01651*.

[19] Sedghi, A., Luo, J., Mehrtash, A., Pieper, S., Tempany, C. M., Kapur, T., Mousavi, P., and Wells III, W. M. (2018). Semi-supervised deep metrics for image registration. *arXiv preprint arXiv:1804.01565*.

[20] Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (2016). A deep metric for multimodal registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer.

[21] Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., and Shen, D.

(2013). Unsupervised deep feature learning for deformable registration of mr brain images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer.

[22] Chollet, F. et al. (2015). Keras.

[23] Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, M., and Schnabel, J. A. (2012). Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis*, 16(7):1423–1435.

[24] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., and Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2758–2766.

[25] Heinrich, M. P., Jenkinson, M., Papież, B. W., Brady, M., and Schnabel, J. A. (2013). Towards realtime multimodal fusion for image-guided interventions using self-similarities. In *International conference on medical image computing and computer-assisted intervention*, pages 187–194. Springer.

[26] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[27] Fan, J., Cao, X., Yap, P.-T., and Shen, D. (2018b). Birnet: Brain image registration using dual-supervised fully convolutional networks. *arXiv preprint arXiv:1802.04692*.

[28] Ehrhardt, J., Schmidt-Richberg, A., Werner, R., and Handels, H. (2015). Variational registration. In *Bildverarbeitung für die Medizin 2015*, pages 209–214. Springer.

[29] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

[30] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8789–8797.

[31] Aants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26– 41.

[32] Ali, S. and Rittscher, J. (2019). Conv2warp: An unsupervised deformable image registration with continuous convolution and warping.

[33] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283.

[34] Hu, Y., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., Emberton, M., Vercauteren, T., Noble, J. A., and Barratt, D. C. (2018a). Adversarial deformation regularization for training image registration neural networks. *arXiv preprint arXiv:1805.10665*.

[35] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., et al. (2018c). Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*, 49:1–13.

[36] Ito, M. and Ino, F. (2018). An automated method for generating training sets for DL based image registration. In *The 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING*, pages 140–147. INSTICC, SciTePress.

[37] Lv, J., Yang, M., Zhang, J., and Wang, X. (2018). Respiratory motion correction for free-breathing 3d abdominal mri using cnn-based image registration: a feasibility study. *The British journal of radiology*, 91(xxxx):20170788.

[38] Miao, S., Wang, Z. J., Zheng, Y., and Liao, R. (2016b). Real-time 2d/3d registration via cnn regression. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1430–1434. IEEE.

[39] Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. (2017). Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Interve tion*, pages 266–274. Springer.

[40] Salehi, S. S. M., Khan, S., Erdogmus, D., and Gholipour, A. (2018). Real-time deep registration with geodesic loss. *arXiv preprint arXiv:1803.05982*.

[41] Sloan, J. M., Goatman, K. A., and Siebert, J. P. (2018). Learning rigid image registration - utilizing convolutional neural networks for medical image registration. In *11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 89–99. SCITEPRESS-Science and Technology Publications.

[42] Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B. P., Išgum, I., and Staring, M. (2017). Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer.

[43] 43. Sun, Y., Moelker, A., Niessen, W. J., and van Walsum, T. (2018). Towards robust ct-ultrasound registration using DL methods. *In Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, pages 43–51. Springer.

[44] Uzunova, H., Wilms, M., Handels, H., and Ehrhardt, J. (2017). Training cnns for image registration from few samples with model-based data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 223–231. Springer.

[45] Yan, P., Xu, S., Rastinehad, A. R., and Wood, B. J. (2018). Adversarial image registration with application for mr and trus image fusion. *arXiv preprint arXiv:1804.11024*.

[46] Yang, X. (2017). *Uncertainty Quantification, Image Synthesis and Deformation Prediction for Image Registration*. PhD thesis, The University of North Carolina at Chapel Hill.

[47] Yang, X., Kwitt, R., and Niethammer, M. (2016). Fast predictive image registration. In *DL and Data Labeling for Medical Applications*, pages 48–57. Springer.

[48] Zheng, J., Miao, S., Wang, Z. J., and Liao, R. (2018). Pairwise domain adaptation module for cnn-based 2-d/3-d registration. *Journal of Medical Imaging*, 5(2):021204.