International Journal of Engineering Research and Advanced Technology (IJERAT)

DOI: 10.31695/IJERAT.2025.3.1

E-ISSN : 2454-6135

Volume. 11, No.3
March – 2025

Matching Medical Images with Deep Learning Networks: A Survey

Ikhlas Watan Ghindawi[1] and Lamyaa Mohammed Kadhim[2]

[1] Department of Computer Science

[2] College of Dentistry

University of Al-Mustansiriyah

Bhagdad, Iraq

ABSTRACT

Many patients' lives are being saved by image-guided interventions, and the image registration issue must be regarded as the most difficult and complex problem to solve. However, the latest enormous advancements in machine learning (ML), which include the potential to deploy deep neural networks (DNNs) on modern many-core GPUs, have created a promising opportunity for tackling a variety of medical applications, including registration. The most recent research on medical image registration with DNNs is reviewed in detail in the presented work. All of the relevant papers that have already been published in the subject are included in the systematic review. This thorough overview includes a detailed discussion as well as survey of important ideas, statistical analysis from many perspectives, novelties and key contributions, confiding challenges, future directions, key-enabling approaches, and prospective trends. For readers who are actively involved in the subject, researching state-of-the-art and hoping to present a contribution to future publications, the presented study offers a deep grasp and insight.

Key Words: Convolutional Neural Network (CNN), Medical Image Registration, Deep Reinforcement Learning, Deep Learning.

1. INTRODUCTION

For prognosis, diagnosis, follow-up, and treatment purposes, certain images must be taken in the majority of medical interventions. The spatial, temporal, modular, and dimensional aspects of such images could all differ. Physicians could benefit greatly from image fusion that creates information synergy in the process of decision-making, particularly when it occurs in real time and online. The accuracy and quality of the next analysis may be compromised by the lack of alignment that is unavoidable for such images obtained under various circumstances. Aligning at least two provided images using the same geometrical coordination system is known as image registration. The goal is to identify the optimal spatial transform which best registers structures of interest. This problem has several applications in the area od machine vision, such as satellite imaging, object tracing, remote sensing, and more [1]. Image registration is essential as well to the image-guided intervention, which includes precision medicine, image-guided radiotherapy (IGRT), and telesurgery [2]. To give an example, in IGRT, a pre-interventional image (which is generally a high-quality 3-D image) which is used for treatment planning must be registered on operational image (which is generally a noisy and low-quality 2D image); therefore linear accelerator (LINAC) machine could be calibrated and radiation fragment could be delivered with minimally invasive procedure as well as the least radiation risks to adjacent healthy tissues. The quality of the problem-solving process could be compromised by issues such as noise and low-quality in the inputted images, different inputted image modalities, thorax cavity's image deformation (which is a result of weight loss/gain throughout the treatment process), the abdominal cavity's image deformation (which results from the spontaneous contraction/inflation), and changing the size of the inputted image. The problem is extremely difficult and complex in practice, where particular considerations must be considered and various image processing algorithms must work together [3]. In essence, traditional image registration can be defined as iterative optimization process which calls for extraction of appropriate features, the choice of a metric of similarity (for the assessment of the quality of the registration), the transformation model, and, lastly, a method

to explore the search space [4]. Through the iterative sliding of the moving image over the fixed image, optimal alignment could be attained. The degree of the correspondence between the input images is initially determined by chosen measure of similarity. The parameters of the new transformation are determined with the use of an optimization algorithm that makes use of an updating mechanism. A new, purportedly better-aligned image is produced by applying such parameters to the moving image. The algorithm is terminated in the case where requirements of termination are met; if not, a new iteration has to begin. The moving image leads to the improvement of its correspondence with the fixed image with every iteration, and the process is repeated to the point where either no more registration is possible or certain predefined requirements have been met. The final interpolated fused image or the transformation parameters may be the system output [4,5].
This approach has the following two primary disadvantages:

- Even with an effective implementation on modern GPUs (NVIDIA TitanX), such iterative approach is extremely slow, with runtimes in tens of minutes typical for the widely used techniques of deformable image registration; however, the practical uses in the clinical operations is real-time, and this sort of prolonged waste of time isn't preferable.
- When dealing with images from various modalities (which are also known as the multi-modal image registration), most similarity measures decline in their effectiveness and cause premature convergence or stagnation, 2 common confining dilemmas in field of optimization. This is because the majority of the measures of similarity have many local optima that surround the global one.

Therefore, DNNs, which can be considered as one of the most potent methods that had ever been discovered by the community regarding ML, were used in a variety of applications in image processing. Naturally, medical image registration isn't an exception, and several DL-based methods have been put forth in literature; still, there is a promising potential for further research, and the number of works and employed methodologies is quite small [5]. This is how the remainder of the paper is structured. The DNNs utilized in literature are examined in Section 3 of the following section. In Section 4, the topic of constraining challenges and open problems in this subject is discussed. Lastly, the final section presents the conclusion and future trends.

2. DEEP LEARNING NETWORKS

In the year 1992, Lo became the first to apply the notion of DL process networks to medical image processing [6]. However, the first operational implementation is dated back to the year 1998, when a Convolutional Neural Network (CNN) has been utilized for the purpose of recognizing handwritten number characters for the post office applications. This was due to the lack of the necessary infrastructure for such a large calculation at the time [8]. The processing power offered through the hardware at the time made it possible to employ DL for a straightforward machine vision task like numerical character recognition. Unfortunately, until 2012, when [9] succeeded in training a deep CNN on a graphic card with a many-core GPU and had won the championship of grand world image processing dubbed ImageNet [9], the suggested method lost its way when applied to other extremely sophisticated and hard situations. Until it was revealed that the DL suggested methods exceed the human expert, the championship has essentially been over. Since then, numerous champions have been in the same family of DL methods, each with a contributing uniqueness [10]. As a result, DNNs of all types spread throughout machine vision and became the method of choice for many professionals working in the subject. Landmark localization, organ detection, lesion detection and classification, follow-up, and treatment planning are a few of the active areas in medical image analysis [5]. Naturally, this was also true for the task of image registration, which can be considered as one of the most significant and difficult problems in the field of image-guided intervention. Various topologies and designs of DNNs are available, each of which is appropriate for a particular set of applications. An Auto-Encoder (AE) is a very basic network that uses a single hidden layer to attempt to recreate input pattern as output. The hidden layer must obviously be smaller when compared to the input pattern in order to map to a more compact space regarding the hidden layer with greatest degree of discriminating capacity. Denoising AEs (DAEs) are networks that attempt to recreate the inputted patterns by applying noise. Adding some noise to the input improves the model's capacity for generalization. Staked Auto-Encoders (SAEs), a deep architecture of AEs, include additional hidden layers staked on top of each other. Since

DOI : 10.31695/IJERAT.2025.3.1

training such a network typically entails an unaffordable computational overhead, each layer is typically trained independently before the network is fine-tuned by a final, inexpensive integrated training. Rather than employing handcrafted features, such network has just been employed in the medical image registration literature to supply the most important and distinguishing features from images to feed to some other approach of registration. One of the most effective and potent DL methods is CNNs, which work by feeding the network the entire image or just selected portions of it. This contrasts with the CNN-based image processing methods, which initially retrieved a few manually created features and sent them into the network. To end up with a conventional fully-connected two or 3-layered network, a CNN often includes some pooling layers and interleaving kernel. Pooling layers reduce the dimensionality curse and result in making outputs invariant to various cases of geometrical transformation, while the kernels are trained for extracting the most important characteristics by the convolution with the input. When the number of layers is considerable, a hierarchical set of feature could be created, and the network is then referred to as deep CNN. The output of every one of the layers, which is known as feature map, is fed to the layer that comes after it. In order to feed a fully-connected 2- or 3-layered network for final classification, the final layer's feature maps are concatenated as well as vectorized. The so-called U-Net [11] is used in many situations, such as deformable image registration, where a direct end-to-end field of registration could be obtained by dropping out the last fully-connected layer (FCL). Additionally, CNN can receive diverse patterns as input from a variety of representations. A network with this functionality is referred to as multi-channel. Each representation is considered a channel. Let's use the classification problem as an example and look at small, inputted patches from images. One could think about a few larger patches around chosen patch, compress them, and send them to the network as separate channel if the context is instructive. Naturally, the network is unable to process these compressed, larger patches over the same channel as original-size patches. A different scenario is when we're dealing with color images and can utilize three RGB channels rather of a single intensity channel for every point in the image. The network could delay the channel to the final levels, which are known as multi-steam networks, or fuse it in the early stages.

The discriminator and generator are two competing subnetworks that make up a Generative Adversarial Network (GAN), which has first been suggested through Goodfellow *etal.* in 2014 [12]. The discriminator must distinguish between the fake (synthesised) and real data as a binary output, whereas the generator had been trained on a ground-truth data-set in order to create fake samples. Similar to game theory, a network could be trained on some small dataset depending on survival struggle between the discriminator and the generator. This way, the network aims for equilibrium and generated samples can't be distinguished. The network gets its name from adversarial training of the generator using discriminator's feedback. Although the original GAN was used to remove noise from images, it has become more and more popular recently and may be used to solve nearly any medical imaging issue [13].

With regard to image registration, the generator takes input fixed as well as moving images and attempts to generate transform parameters which would prevent the discriminator from distinguishing the warped image—a transformed moving image—from the ground-truth, as would be expected from the expert registration agent. An additional data loop (i.e., feedback) from hidden-layer nodes to themselves distinguishes Recurrent Neural Network (RNN) from AEs or the conventional Multi-Layer Perceptron (MLP). Because of such feedback loops, RNNs are the most effective networks for temporal analysis, even while other networks, such as AEs or CNNs, are appropriate for temporal analyses. The next state could be calculated with the use of the present state entered into the network and the previously stored states since previous states are routinely stored across hidden layers. The network is regarded as deep when there are more hidden levels to store farther states. Similar to SAEs, computational training load such a densely connected network is not affordable in the conventional way. As a result, the community has been following research continuously, and fortunately, simplified memory models like Gated Recurrent Units (GRD)[15] as well as Long Short Term Memory (LSTM)[14] were introduced and widely used recently. Even worse is the fact that RNNs are majorly utilized for optical flow in image registration, when one of the modalities is linked to temporal dimensionality, such as X-ray fluoroscopy or TRUS. Finally, Deep Reinforcement Learning (DRL) is a final step. Its foundation is the theory of Markov chains and stochastic processes.

An agent has a reward/penalty rate, transition probabilities, and a few internal states. Iteratively, it learns how to

DOI : 10.31695/IJERAT.2025.3.1

engage with its environment. Using a probabilistic decision-making process, DRL machine selects some action from its action-list at each iteration depending on feedback from the environment, its internal states at the time, and transition probabilities. The chosen action is done to the environment, and the machine receives a reward or a penalty depending on how desirable the feedback is. Thus, DRL machine learns to choose the optimal course of action in every case, in which the optimal course of action is the one that has the highest likelihood of receiving a reward from the environment. Such DRL agents have been used in image registration for affine or affine transformations in which the number of the states is limited and it is feasible for the agent to converge. For instance, the agent could choose to rotate in a clockwise or counterclockwise direction by one degree or translate in all directions by one millimeter. Those chosen actions are applied to moving image, and the agent receives a reward or a penalty depending on how desirable the actions are, such as a similarity measure. To optimize its efficiency, it uses learning methods like Q-Learning to alter its internal transition probabilities.

3. LITERATURE REVIEW

Iterative optimization algorithms are used in conventional image registration. Depending on a predefined similarity metric, a better alignment is intended to be attained in each iteration. Until certain predefined conditions are met or a better registration cannot be obtained, the activities continue. The most difficult problems to overcome in order to take advantage of such paradigm are the lengthy running time and flawed nature of introduced measures of similarity, particularly for the multi-modal registration, which leads to becoming stuck in local minima. Recently, DL-based approaches gained more popularity as a solution to the aforementioned issues. Thosemethods' basic philosophical underpinnings fall into one of two categories:

- To assist other registration techniques, DNNs serve as similarity approximator between input images as comprehensive and no-faulty metric of similarity.
- To optimize runtime speed, a DNN serves as regressor, directly estimating parameters of transformation in a single shot.

Supervised End-to-End Registration (SE2ER), Unsupervised End-to-End Registration (UE2ER), Deep Reinforcement Learning (or Agent-Based Registration) (DRL), Deep Similarity Metrics (DSM), and Weakly/Semi-Supervised End-to-End Registration (WSE2ER) are the five generations of a taxonomy that could be derived from the literature's advancements. The initial generation of efforts, which have been influenced by [16] and [9], focused on using various types of DNNs to learn visual metrics of similarity from a sizable collection of the paired annotated ground-truths. They have been dubbed metrics or measures of profound similarity. Following training, the learned model should be capable of accurately and significantly representing structural variations between input image/patching pairs. The two most significant examples for this primary generation are [17] and [18]. For generating final transformation parameters, deep similarity measurements are essentially supplied to the traditional iterative deformable registration techniques. Since many comparable methods have already been developed, we could say that this model, in the most basic form, could be a strong competitor to conventional multi-modal measures of similarity, like Mutual Information (MI), provided that there are a sufficient number of the available clearly annotated ground-truths. This has been considered as one of the major barriers to the development of this kind of approaches. Furthermore, it was discovered that there is no compelling reason to use deep similarity measures for the unimodal registration in the case where the measure of similarity can be appropriately chosen depending on modality and context.

The second generation falls under the category of end-to-end supervised registration, in which several DNN types are trained on ground truth in order to build models of regression that generate parameters of transformation in a single shot. CNN as well as U-Net (also known as fully CNN) are the most common methods for affine as well as deformable transformation models, respectively [19] and [20]. Initially, a Dense Displacement Field (DDR) can be defined as a grid of control points. The underlying deformation could be captured by freely translating each control point in both vertical and horizontal directions. The accuracy regarding the model in capturing the deformation is determined by quantity of control points and the distance between them. In order to reduce computing overhead, B-Spline techniques just take into account nearby control points (local transformation), whereas Thin-Plate Spline (TPS) broadcasts every movement in control point to all ones (i.e., global

DOI : 10.31695/IJERAT.2025.3.1

transformation). The regularizer, which penalizes unacceptable transformations, regulates deformations. Similar to traditional methods, there is a great deal of debate and dispute on the term of regularization that should be specified for DNN. This debate has led to numerous advances in the field [21]. The launch of the Spatial Transformer Network (STN) by Jadderberag *etal*. in 2015 [22] is another source of advances in this generation. It is an explicit module which could be added to various DNN types to make the data flow between hidden layers transformation invariant. In the case of being combined with the pooling layers which are implicitly translation as well as scaling invariant, they could work in tandem to introduce a full spatial invariance set, which might significantly improve the performance of CNNs utilized in a variety of image processing applications, which include medical image registration. Three successive components make up the STN. A localization network, such as a standard MLP, uses a predetermined measure of similarity as loss function to learn how to regress parameters of transformation for input feature map. This network has a very flexible topology. The second is a grid generator that applies the localization network's estimated transformation parameters to input feature map. Lastly, the sampler that creates the final outputted twisted image by acting as an interpolator. The STN is a subject of contention in the community since it is completely differentiable, meaning it can be placed anywhere in the network, depending on the context. Large transformations could result in significant output distortion, which the sampler cannot tolerate, and boundary interpolation is particularly challenging for the sampler because a part of the output must be brought from outside the input, which is non-existent. STN is not perfect.

Inverse Compositional STN (IC-STN) was lately introduced through Lucey and Lin in the year 2017 [23]. They contended that we could delay reconstruction through the sampler and communicate parameters of transformation together with the output, where the actual CNN determines the way of handling the transformation. In fact, the problem is still unresolved and the issue remains yet in its infancy. In order to optimize the positive feedback from environment (in this case, the measure of similarity), the 3rd generation of deep reinforcement learning (DRL) agents (or several agents) learn to achieve the final transformation in a gradual manner. The similarity measures are frequently given in a traditional manner, such as Local Cross Correlation (LCC) and Normalized MI (NMI), in place of the first deep similarity measure concept. The incapacity of agents to interact with the vast state space that had been created by deformable registration field is the main limiting aspect that has caused the production of this paradigm to go extinct. The concept is doomed to construction if it cannot capture the deformation necessary for successful registration of elastic organs and a comparatively long registration period. The fourth generation falls under the category of the unsupervised end-to-end registration, in which various DNN types are trained with no ground-truth for constructing models of regression for producing parameters of transformation in a single shot. This is because the preceding generations relied on ground-truth for constructing the model, and typically, annotated data-sets in the medical field, and particularly for image registration, are small-sized and unsuitable for exhaustive DL. They employ methods of data augmentation on a small number of input samples as seeds rather than some large grand-truth set, and they utilize a conventinal measure of similarity (or a mix of them) as the loss function to direct learning process. Whereas multi-modal registration is much more difficult due to the fact that multi-modal measures of similarity remain insufficient, and networks that are trained on them would inherit such insufficiency, the majority of this generation's techniques have proven effective with unimodal registration. When a CNN performs the final estimation of transformation and an SAE is trained for extracting the features, [24] could serve as a decent example. The use of SAEs in place of more traditional multimodal similarity metrics, such as MI, supports our previously stated claim. The regularization term rule is essential for controlling applied deformations for creating realistic fake samples using data augmentation techniques. We anticipate a convoluted path ahead with a major research focus in foreseeable future, although experts and practitioners are skeptical in such regard.

Lastly, the 5th generation falls within weakly/semi-supervised methods because both unsupervised as well as supervised end-to-end image registration have their own set of disadvantages. This category has two distinct key paradigms. A few methods rely on the grand-truth data that has been completely annotated with as many landmarks as feasible. Those landmarks are typically legions, contours, lines, corners, turning points, and so forth, and every one of them is assigned a unique class designation. Those properly labeled data were used for training the network, but there were several exceptions. It learns to identify landmarks in any pair of input images in addition to its primary function, which is image registration. Finding these landmarks is essential to

DOI : 10.31695/IJERAT.2025.3.1

building effective models and improving system accuracy. Furthermore, the network could be trained using the non-trivial loss function of Target Registration Error (TRE), the most valuable metric of structural similarity. Another paradigm has been based upon the use of GAN [12], in which the generator takes inputs of both fixed as well as moving images and attempts to generate parameters of transformation, which prevent the discriminator from distinguishing transformed moving images from ground-truth, as would be expected from an expert registration agent. Similar to the game theory, the network could be trained on a small dataset depending on the survival competetion between the discriminator and the generator. This way, the network aims for equilibrium and the generated samples cannot be distinguished.

4. DISCUSSION ON THE CONFIDING CHALLENGES , OPEN PROBLEMS AND PROSPECTIVE DIRECTIONS

A review of the literature on application of DL techniques to medical imaging shows us several common problems. However, the collection of innovative solutions put forth by relevant writers, specialists, and researchers in their papers can serve as a valuable resource and manual for addressing future issues in DNN-based medical image registration field.

Here is a list of some of these issues and their fixes:

• **Challenge:** Medical data-sets are sometimes of a small-size.

Solution: Increasing the number of samples in an artificial manner through the use of augmentation methods. Training the network

using different datasets and after that refining it using the dataset under consideration through the use of transfer learning.

Using semi-annotated data for training the network using weakly-supervised learning. Lastly, the overfitting effect can be

reduced by using dropout, a technique that randomly removes some inputs from each layer.

• **Challenge:** Because experts and doctors frequently lack consent, medical data-sets' allocated labels are noisy to a significant degree.

Solution: the problem can be solved by modeling the distribution of noise as well as feeding it into the network, or by applying

techniques like fuzzy logic.

• **Challenge:** Despite the system's great accuracy and precision, DL techniques do not provide the rule-chain for its conclusions, which might be unsatisfactory to doctors. This is in contrast to Decision Support Systems (DSSs).

Solution: Although a few encouraging research has already been done to illustrate how networks infer information by visualizing their internal (hidden) layers, the issue remains unresolved and the progress is modest.

• **Challenge:** As doctors ask the patients various relevant questions and review their various records and test findings, background information and context could be very instructive.

Solution: the clinical records of the patients, as well as their genomic information, biopsies, and the results of other experiments could be collected

and fed into the network through various channels to improve performance. However, there aren't enough integrated datasets to examine the impact, which makes matters worse.

• **Challenge:** Although medical imaging is 3D by nature, the majority of DNNs now in use analyze it in 2D or 2.5D. The

reason is because in various situations, 3D processing with 3D DNNs is computationally prohibitive.

Solution: Nothing has to be done!

We should take a seat and observe if the infrastructure's advancements will eventually allow us to draw in the processing power needed to accomplish it.

According to a survey of literature on the DNN-based medical image registration, the suggested methods aim to improve the two parameters listed below:

DOI : 10.31695/IJERAT.2025.3.1

1. **Registration Runtime:** the authors didn't want to over-engineer a network by adding more connections, layers, and parameters in order to significantly increase performance; instead, they designed the suggested methods to reduce registration runtime at the same time as maintaining performance. The suggested method can register a usual pair of input images in less than 50ms, according to de Vos et al. (2019), which is very appropriate and valued for the real-time clinical usage.

2. **Network Receptive Field:** Usually, a small sliding window with a degree of overlap is used in order to choose extracted patches from the two input images. The receptive field of the network is confined but background context might be completely informative. Patches are typically small, ranging from 13×13×13 to 30×30×30 voxels, in order to reduce the computational intensity to be tractable. In order to solve the problem, some studies separate the larger patches from the usual patches, compress and shrink them (to lessen the computational load), and then feed them into network through a new channel.

5. CONCLUSIONS AND FUTURE TRENDS

SE2ER, DSM, UE2ER, DRL, and WSE2ER are the five categories of the taxonomy on DL-based methods for medical image registration that was developed in the presnted work. Each category's methods include some of the same requirements, underlying drawbacks, models, advantages, and concepts. In general, deep reinforcement learning hasn't been promising because the learning agents cannot withstand the large state-space related to the deformable registration, which prevents them from properly convergent. We anticipate that, despite the obstacles and limiting issues they confront, unsupervised as well as weakly-supervised techniques will receive more attention and research focus because, among other things, they rely less on ground-truth, which is quite expensive to gather for the applications of medical imaging. Unless we come across the release of large, publicly accessible annotated the data-sets for many organs of interest with several kinds of modality. One other different option for the supervised methods is transfer learning, which has been well used for a few organs and modalities, yet requires considerably more proof to be generalized. While dual-supervision, which has been based upon learning from a measure of similarity, such as MI (just like the unsupervised methods) as well as some ground-truth samples for the purpose of fine-tuning the network, is highly important and deserves more consideration, the results become questionable as we move farther away from the real-world ground-truth. Adverbial learning, such as the use of GANs, is the main contribution amongst weakly-supervised end-to-end registration. In order to dispel this concern, we anticipate more research on the registration validity.

This study leads to the next conclusions:

- Numerous publications have indicated that the multistage policy, which defines a rigid registration before moving on to a
- deformable one, has a favorable effect.
- Multiresolution policy has been identified as a powerful concept to improve the precision of the registration, whereby the registration procedure is progressively carried out from low-resolution to the highest resolution.
- Transfer learning from other modalities or body organs is completely feasible for the registration of medical images, according to literature, provided that a small number of ground-truth samples are available.
- Including geometry theory in the methods is a new concept that requires further research.
- One of the main contributions is the Spatial Transformer Network (STN), which could significantly enhance performance when combined with a CNN, for example.
- Since the program is real-time, over-engineering—that is, adding more connections, layers, and parameters to significantly boost performance—is applicable in this instance.
- The increase of network's receptive field is a beneficial influencing component that various authors follow because the background context could be totally instructive.

We think that the majority of upcoming developments and contributions will come from other issues of the field of medical imaging or perhaps from domains a little further away, such as ML and computer vision, rather than being intrinsic in medical image registration itself. While additional models have a great potential for

DOI : 10.31695/IJERAT.2025.3.1

contribution, DL approaches applied to the field of medical image registration are limited to SAEs, CNNs, DRL, GANs, and deep RNNs from an application standpoint. For example, the recurrent DL model known as GRD has a great potential for use in situations where time is the fourth dimension, such as in discrete fluoroscopy or continuous US images. However, from the perspective of method, the domains of computer vision and ML are constantly evolving, with new and promising methods being introduced on a regular basis. For instance, Spiking Neural Networks (SNNs), the third generation of NNs, use biologically accurate neuron models for computation in an effort for bridging the gap between neuroscience as well as ML. In essence, an SNN differs from the traditional NNs that the ML community is familiar with. Rather than using continuous values, SNNs utilize spikes, which can be described as discrete events that happen at time instances. Differential equations representing multiple biological processes—of which membrane potential is the most crucial—determine where a spike should occur. Generally speaking, a neuron spikes and resets its potential in the case where it reaches a particular potential. Leaky Integrate-and-Fire (LIF) model is the most often utilized one for that. Additionally, SNNs frequently utilize sparse network topologies and are sparsely connected.

REFERENCES

[1] A. A. Goshtasby, *Theory and Applications of Image Registration*. John Wiley & Sons, 2017.

[2]T. Peters, and K. Cleary, *Image-guided interventions: technology and applications*. Springer Science & Business Media, 2008.

[3]J. Hajnal, D. Hawkes, and D. Hill, *Medical Image Registration*. Biomedical Engineering, Jun. 2001.

[4]F. P. Oliveira, and J. M. R. Tavares, “Medical image registration: a review” *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73-93, 2014.

[5]G. Ronnebergerro, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.

[6] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[7] S.-C. B. Lo, M. T. Freedman, J.-S. Lin, B. Krasner, and S. K. Mun, “Computer-assisted diagnosis for lung nodule detection using aneural network technique,” in *Medical Imaging VI: Image Processing*, Jun. 1992.

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the *IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] G. Huang, M. Mattar, H. Lee, and E.G. Learned-Miller, “Learning to align from scratch,” in *2012 Advances in Neural Information Processing Systems*, 2012, pp. 764-772.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Apr. 2015.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. *MedicalImage Computing and Computer-Assisted Intervention* (MICCAI 2015), 2015, pp. 234-241.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y.Bengio, “Generative adversarial nets,” in *Advances in neural information processing systems*, 2014, pp. 2672-2680.

[13] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” arXiv preprint arXiv:1809.07294, v2, pp. 1-19, 2018.

DOI : 10.31695/IJERAT.2025.3.1

[14] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[15] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation,” in proc. *2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2014.

[16] E. Nowak and F. Jurie, “Learning Visual Similarity Measures for Comparing Never Seen Objects,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007.

[17] G. Wu, M. Kim, Q. Wang, Y. Gao, S. ao, and D. Shen, “Unsupervised Deep Feature Learning for Deformable Registration of MR Brain Images,” *Lecture Notes in Computer Science*, pp. 649–656, 2013.

[18] X. Cheng, L. Zhang, and Y. Zheng, “Deep similarity learning for multimodal medical images,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 248–252, Apr. 2018. (Accepted & Online from 2016)

[19] S. Miao, Z. J. Wang, and R. Liao, “A CNN Regression Approach for Real-Time 2D/3D Registration,” *IEEE Transactions on MedicalImaging*, vol. 35, no. 5, pp. 1352–1363, May 2016.

[20] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, “Nonrigid Image Registration Using Multi-scale 3D Convolutional Neural Networks,” *Lecture Notes in Computer Science*, pp. 232–239, 2017.

[21] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable Medical Image Registration: A Survey,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[22] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *2015 Advances in neural information processing systems*, 2015, pp. 2017-2025.

[23] C.-H. Lin and S. Lucey, “Inverse Compositional Spatial Transformer Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Jul. 2017.

[24] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, “Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.

DOI : 10.31695/IJERAT.2025.3.1