Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/19/20231054

Review on multi-modality medical image fusion using deep learning

Ishika Gupta[1], Satyam Gupta[1], Manoj Diwakar[1], Prabhishek Singh[2], Achyut Shankar[3]and Sathishkumar V E[4,5]

[1]CSE Dept., Graphic Era Deemed to be University, Dehradun, India
[2]School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
[3]University of Warwick, UK
[4]Department of Software Engineering, Jeonbuk National University, Jeonju-si, Jeollabuk-do, Jeonju-si, Republic of Korea

[5]sathish@jbnu.ac.kr

Abstract. Multi-Modality medical image fusion is a method in which multiple images are merged having either single or multiple imaging modalities. This process is carried out to improve the quality of imaging while preserving all the essential and distinct features. Many areas such as Machine Learning, Artificial Intelligence, Image Processing, and Computer Vision are covered by Medical Image Fusion. This method has been adopted on a large scale by physicians to apprehend any damage or injury caused in organ tissues in clinical trials by performing a fusion of images with different modalities. In this review, Deep Learning methods carried out in the medical image fusion field have been discussed along with a comparison between their accuracies. The main objective of this paper is to list some of the most effective techniques in this domain and discuss their performance. At last, the paper concludes with the fact that although the development and growth in this area have increased over the years, many challenges have also come along the way.

Keywords: multimodal, image fusion, neural network, deep learning, diagnosis, medical image

1. Introduction

There has been a significant amount of progress in the technology field and the domains of image processing and fusion are no longer abstract. These areas find their application in many places such as medical imaging, astronomy, and the military. Thus, they have helped in collecting much useful information and data. It has been possible to extract valuable details from the images by performing the method called image fusion. The primary objective of conducting image fusion is to produce clearer and more informative images that provide us desired results with higher accuracy and are more appropriate [1-3]. As the name suggests, this process of image fusion is called so because it involves the merging of multiple images into one that has all the characteristic features contained in all the images separately [4-5]. The resultant fused image seems to be more reliable and informative when compared to the individual images studied. Moreover, the fusion increases accuracy and ensures to a high extent that there is no uncertainty in the result [6-7]. Image fusion, as mentioned before, finds a significant place in medical

DOI: 10.54254/2755-2721/19/20231054

diagnosis these days to understand the condition of an injury in the organ. Several Deep Learning algorithms have been adopted by researchers to study medical images and provide positive results.

Traditionally, the process of image fusion undergoes decomposition and then reconstruction (Fig. 1). There are several image decomposition algorithms that first break down the images received as input into a sequence of sub-images. After that, the second step is where the sub-images of varying resolutions having their specific features are merged by using some image fusion tools. Lastly, after getting the fused sub-images, they are reconstructed to get the final fused image with the help of the image reconstruction algorithm [8-10]. The quality of the image is hence improved and can play a larger role in the predictions required.

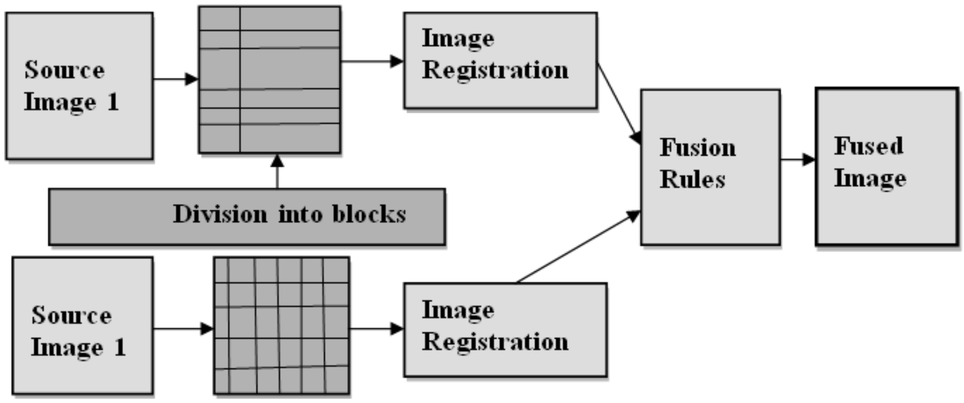

Figure.1. Image fusion model.

Some minimal requirements need to be satisfied for conducting this image fusion method, factors such as alignment, rotation, and change of scale are kept in mind to give the appropriate final integrated visual. These are the conditions:

- All the significant features of the individual input or source images should not be left behind and all of them should be transferred to the resultant fused image without any loss of information.
- The process of image fusion when carried out must not bring any inconsistency in the data and information that can mislead the study and results.
- The fused image must be reliable and be obtained by suppressing the parts of the image that were not required and irreverent.

Different types of details are provided by different multi-modality medical images: Magnetic Resonance Image (MRI) provides details about soft issues, X-Ray helps in the detection of fractures, Computed Tomography (CT) provides detailed internal images, and Single Photon Emission Computed Tomography (SPECT) gives metabolic details as well as the functional ones.

Several pairs of multi-modal medical images such as CT-MRI, X Rays-Ultrasounds, etc are merged to get some extra information clinically. It is not denied that single images cannot serve to give us all the essential information and thus medical image fusions are necessary to make a composed image that can be used by physicians for diagnosing purposes. The main benefit of the fused image is that it tends to reduce the problem and challenge the detection of multiple diseases.

In addition to all the qualities of the fused medical images, their visualization and predictive capability have also been increased over the years by moving two-dimensional fusion to higher dimensions such as 3-dimensional and 4-dimensional. Several studies are also being made by the researchers to present the concepts of multi-sensor and multi-focus image fusion for in-depth analysis of medical images. Therefore, the advantages and applications of multimodality medical image fusion cannot be denied, and several Deep Learning techniques and algorithms should be analyzed to carry out the same.

2. Literature review

Many methods have been used by researchers for medical image fusion based on domains such as the Spatial domain and the Transform domain in recent years, however, first has proven to give better effects and results of medical image fusion, since Deep Learning emerged in 2017 it has been used to carry out image fusion. So many Deep Learning concepts and algorithms such as CNN (Convolutional Neural Network), U-Net network, GAN, RNN (Recurrent Neural Network), and various other models have been used in recent years in performing medical image registration along with segmentation [1]. Out of these models in techniques, the U-Net network and CNN have been used to conduct medical image fusion. Deep Learning is a hot area for research in Medical Fusion recently, Convolutional Neural Network (CNN) is one of the most famous and widely used Deep Learning models. While the processes of spatial domain and transform domain are not so accurate being incapable of extracting all the essential features of the individual images and do not fully satisfy the required conditions of the fusion rules, deep learning has been able to overcome the problems. Deep Learning finds its application in the segmentation and registration of medical images. CNN was applied first in 2017 for image fusion and it proved to provide better accuracy and results than the spatial domain and transform domain-based methods. Similarly, another deep learning-based model called the U-Net network has also found its large application in the field of medical image fusion. Whether we consider the 2-Dimensional fusion or the 3-Dimensional one, it has proved to provide great results. Convolutional Neural Network (CNN) is a category of Deep Neural Networks that is widely used in analyzing visuals. The technique used in this method is known as Convolutional. It is a multistep feedforward ANN (Artificial Neural Network). It is a supervised form of Deep Learning where the first parameter is known as the input, the second one is known to be the kernel function and finally, the output is known to be the feature map [2]. The operation conducted is multi-dimensional. Consisting of subsequent layers of artificial neurons, the weighted sum of the numerous inputs provided is calculated and an activation value is given as the output. The method of matrix multiplication is used traditionally to understand the connection relationship among the layers. A considerable amount of storage space is required for the output U-Net but as the CNN is naturally represented to be sparse, the connection of neurons is between the previous stages and the neurons adjacent to it thus reducing the capacity or storage required and resulting in improvement of computational performance. The non-uniqueness of the weights in classical networks is eliminated via parameter sharing in CNN [3]. In the CNN stage, the weight being constant makes it superior in terms of storage requirements to others. Traditional automated encoders are completely linked [4].

The output and source image in the case of CNN is not mandatorily aligned, but in the case of U-Net, another deep learning-based model used for medical image fusion, a locally connected structure is used. So, because of this the vector output, as well as the input image, are aligned giving a better-fused image and better visual effects. U-Net is a fully convolutional network having paths of contraction and expansion [5]. It is not denied that in all the deep learning algorithms and models a dataset containing many data samples is required but U-Net proves to be exceptional being able to work and provide great results on small datasets as well as it uses data enhancement [6]. This is possible because of its full convolutional network model which has improved it making it advantageous in the medical image fusion scenario where often a shortcoming of a small dataset is faced frequently [7].

3. Methodology

In this section, the two deep learning models Convolutional Neural Network (CNN) and U-Net network are briefly discussed along with their functioning and process flow. The research and observations in these domains have been summarized along with the challenges faced along the way [8].

3.1. Convolutional neural network (cnn) based image fusion

As the medical images, even at the same location have different intensities the method for the fusion of images is to be taken accordingly [9]. The first method based on CNN which was proposed dates to 2017 when a network known as the Siamese network was used. This network is used in the CNN model and is one of the three models used in drawing comparisons between image patch similarity. This model is highly advantageous when compared to other models such as 2-channel and pseudo-Siamese and is very easy to train making it favorable for conducting medical image fusion. The method of extracting essential features and measuring the activity level of different sources of individual images comes out to be the same because of its similar branches of two weights. So, this Siamese network provides the weight map after which a decomposition method called Gaussian Pyramid is applied along with Pyramid transform to conduct decomposition helping in making the image fusion process by the discernment of human visuals. After decomposition, a similarity waste fusion method is used to regulate it [10]. To provide a superior fusion approach, the algorithm integrates the standard pyramid based and the similarity-based fusion algorithm along with the CNN model [11]. In the medical field, CNN has become a major challenge because of the following reasons:

- A huge amount of explicit training dataset is needed to perform the algorithm.
- The training of the data takes much time because of being large.
- The problem of convergence is often faced which is quite complex and there is the constant requirement of adjustment of overfitting.

To overcome all these difficulties many researchers have proposed a fusion algorithm of CNN with other methods such as Shearlet, MatConvNet, and many similar algorithms. Although these have complex architecture and are difficult to deploy, they prove to be better than the previously used CNN fusion algorithm and provide a better effect of fusion [12].

3.2. U-Net-based image fusion

Almost all the methods of medical image fusion that have been generated do not take care of the semantics of the images, how should their conflicts be processed, or whether there is a chance of losing important information due to all this. Because of not keeping these things in mind the resultant fused image becomes blurry, making it more difficult for the physicians and medical workers to comprehend the fused image and detect the injury. To overcome this situation a fusion method was proposed to solve the issue of loss of data. This was a method based on the U-Net network. In this method, an FW-Net network was made by merging 2 U-Net using the automatic encoder [12]. There are two structures of FW-Net, namely left and right which serve as encoder and decoder respectively. While the encoder is used for the extraction of essential features from the source image, the decoder is used for the reconstruction of the same. FW-Net is capable of extracting source images having bright semantics mapping the various modalities of images and their brightness in the same space for fusion. This method provides images that are having clarity and smoothness if there is no semantic conflict. Further research is being done in this field but for now, this method is only applicable to CT and MRI. Increasing the boundaries of this method is a focus of the researchers so that its application is expanded, and it can be used in various other diagnoses. As we know this method is not yet mature but hence proven to be more efficient than the CNN model in many aspects and overcome some major challenges.

4. Comparative study

As the deep learning models of CNN and U-Net were discussed and how they provide fused images for medical purposes, it was learned how widely they are used for the segmentation and registration of medical images (Table 1). Many deep learning frameworks such as MatConvNet, TensorFlow, Caffe, and similar ones are used in the field of medical image fusion supporting the CNN model while the Pytorch framework helps in training the U-Net network. If CNN and U-Net are compared, there are several factors by which we can say that U-Net gives better results. While U-Net can handle 2-dimensional to 3-dimensional structures, CNN has not found much application in the segmentation of such medical images. Moreover, the output of CNN required a considerable amount of storage, and the source image along with the vector output is not aligned properly making it a bit difficult to predict the result. Due to a different structure called the local connection of U-Net, it allows the alignment of vector output and the source images. Being a fully convolutional network, it has paths of contraction and expansion. One of the major aspects which increases the scope of uses of the U-Net is that while CNN requires a large dataset to train to get clear results, U-Net can work with a smaller dataset, thus solving a major shortcoming of having a smaller sample of data.

Table 1. Deep learning models for medical image fusion.

Methods	Description	Merits	Demerits
Convolutional Neural Network (CNN) [11]	A weight map is generated which combines the information of the input images. The method of the Siamese network is used. Gaussian pyramid decomposition is applied after obtaining a weight map. Lastly, the CNN model is fused with the similarity-based and the pyramid-based fusion algorithm to form a superior fusion method.	The Siamese model can be trained easily making it favorable to perform medical image fusion.	Training the model takes a lot of time and a large dataset is required for prediction.
U-Net [12]	In this U-Net-based image fusion method, the FW-Net network is constructed using 2 U-Nets. An automatic encoder is used to combine the same following the structure of the individual units. The left structure of FW-Net is an encoder which is used for extraction while the right one is a decoder which is used for reconstruction.	This method provides a smooth and clear image that is easy to study. Moreover, image semantics are not neglected and there is no loss of useful information.	The research in the field of medical image fusion using the U-Net network is not so mature and is being continued to get results with accuracy.

5. Conclusion

In the field of biomedical research, multi-modality medical image fusion plays a significant role and strives to improve the quality of images by merging them into one having all the essential features of all singular source images. There are many approaches available for performing the fusion which has comparable performance and seems appealing when employed. While some or other challenges were faced while applying each of the techniques and models, the researchers have always come up with even better solutions to those problems. It was seen that while the CNN fusion algorithm worked almost perfectly but it took much time due to the large dataset and had some alignment issues. U-Net overcomes those and still seems to be a focus for future research for its advancement and increase in efficiency. To sum this up, this paper discussed the deep learning techniques for performing medical image fusion on images of different and multiple modalities which are in trend and have a scope of development in the future.

References

[1] Singh, P., Diwakar, M., Cheng, X., & Shankar, A. (2021). A new wavelet-based multi-focus image fusion technique using method noise and anisotropic diffusion for real-time surveillance application. Journal of Real-Time Image Processing, 18(4), 1051-1068. doi:10.1007/s11554-021-01125-8.

[2] Diwakar, M., Tripathi, A., Joshi, K., Sharma, A., Singh, P., Memoria, M., & Kumar, N. (2020). A comparative review: Medical image fusion using SWT and DWT. Materials Today: Proceedings, 37(Part 2), 3411-3416. doi:10.1016/j.matpr.2020.09.278.

[3] Chakraborty, A., Jindal, M., Khosravi, M. R., Singh, P., Shankar, A., & Diwakar, M. (2021). A secure IoT-based cloud platform selection using entropy distance approach and fuzzy set theory. Wireless Communications and Mobile Computing, 2021 doi:10.1155/2021/6697467.

[4] Dhaundiyal, R., Tripathi, A., Joshi, K., Diwakar, M., & Singh, P. (2020). Clustering based multi-

modality medical image fusion. Paper presented at the Journal of Physics: Conference Series, , 1478(1) doi:10.1088/1742-6596/1478/1/012024.

[5] Singh, P., & Diwakar, M. (2021). Wavelet-based multi-focus image fusion using average method noise diffusion (AMND). Recent Advances in Computer Science and Communications, 14(8), 2436-2448. doi:10.2174/2666255813999200720163938.

[6] Diwakar, M., Singh, P., & Shankar, A. (2021). Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain. Biomedical Signal Processing and Control, 68 doi:10.1016/j.bspc.2021.102788.

[7] Singh, P., Diwakar, M., Chakraborty, A., Jindal, M., Tripathi, A., & Bajal, E. (2022). A non-conventional review on image fusion techniques. Paper presented at the 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2021, doi:10.1109/UPCON52273.2021.9667653.

[8] Diwakar, M., Singh, P., Shankar, A., Nayak, S. R., Nayak, J., Vimal, S.,Sisodia, D. (2022). Directive clustering contrast-based multi-modality medical image fusion for smart healthcare system. Network Modeling Analysis in Health Informatics and Bioinformatics, 11(1) doi:10.1007/s13721-021-00342-2.

[9] Singh, P., Shankar, A., Diwakar, M., & Khosravi, M. R. (2022). MSPB: Intelligent SAR despeckling using wavelet thresholding and bilateral filter for big visual radar data restoration and provisioning quality of experience in real-time remote sensing. Environment, Development and Sustainability, doi:10.1007/s10668-022-02395-3.

[10] Singh, P., Shankar, A., & Diwakar, M. (2022). Review on nontraditional perspectives of synthetic aperture radar image despeckling. Journal of Electronic Imaging, 32(2), 021609.

[11] Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, Cheng Zhong, "A Review of Multimodal Medical Image Fusion Techniques", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 8279342, 16 pages, 2020.

[12] Chanumolu, Rahul, Likhita Alla, Pavankumar Chirala, Naveen Chand Chennampalli, and Bhanu Prakash Kolla. "Multimodal Medical Imaging Using Modern Deep Learning Approaches." In 2022 IEEE VLSI Device Circuit and System (VLSI DCS), pp.184-187. IEEE, 2022.