Deep Learning–Based Image Registration and Normalization for Robust Multi-modal Medical Image Fusion

Kiran Kumar Beesettie[1], Hemalatha S[2],
[1] Lincoln University College,
[2] Professor, 2Panimalar engineering college, Bangalore trunk road, poonamallee, Chennai Tamil
Email ID : kirankumar224@gmail.com

Abstract: The integration of CT, MRI, and PET via multimodal medical image fusion is establishing itself as an effective tool to enhance diagnostic accuracy by leveraging complementary information from different imaging modalities. Yet, effective image fusion relies on image registration and consistent intensity normalization, which remain challenging because these factors depend on differences in spatial resolution, contrast, and acquisition protocols across imaging modalities. In this paper, we explore a deep learning–based method as a potential solution for robustly registering and normalizing multimodal medical images, enabling reliable image fusion. In the proposed method, a convolutional neural network is used to align and standardize the intensities of multi-modal images before fusion, ensuring consistency in spatial distribution and intensity. Fusion and registration should be further complemented and augmented by correcting cross-modal spatial alignments and modulating representations to address intermodal misalignment and intra-modal incoherence. We detail the positive impact our proposed approach, particularly in addressing fusion and registration, has on rapid diagnosis and downstream diagnostic performance. We detail the positive impact our proposed approach, particularly in addressing fusion and registration, has on rapid diagnosis and downstream diagnostic performance

Keywords: Multi-modal medical imaging, image registration, intensity normalization, deep learning, medical image fusion

Introduction

The Medical imaging plays a key role in modern clinical pathology, as it provides a means to visualize anatomy and physiology without harming the patient. Different imaging technologies, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET), provide complementary information that is integral to the identification of the disease and the planning of treatment. While CT scans provide detailed imaging of bony structures, MRIs enhance soft-tissue imaging, and PET scans capture functional and metastatic activity. Dependence on a particular imaging technology may result in the absence of key diagnostic information [1]. The combination of different technologies to capture images of the same anatomy (medical image fusion) has emerged as a valid solution to the aforementioned problem. The value of multimodal image fusion depends heavily on

the quality of the image processing methods, particularly in image registration and intensity normalization. Accurate registration ensures that different imaging modalities capture the same anatomy at the same point in time, and normalization corrects image intensity variations caused by different modalities and acquisition times. The absence of appropriate processing may cause image misalignment, the absence of key features, and poor quality of fusion, and may ultimately affect the reliability of the diagnosis [2]. The conventional techniques of image registration, both rigid and non-rigid, rely on manually crafted similarity metrics and iterative optimization. Considerable differences between two (or more) images (Intermodality discrepancies), noise, and complex anatomical differences can create alignment problems. Conventional techniques of intensity normalization may fail to balance differences in intensity and distribution across multimodal images (also in medical imaging). These limitations of traditional techniques signal a need for new, adaptable, and robust techniques [3].

Medical imaging has benefited significantly from the incorporation of new Artificial Intelligence (AI) methodologies, particularly Deep Learning (DL) techniques, and more specifically, from the use of Convolutional Neural Networks (CNN). CNN networks tend to self-learn features of a given image and are particularly effective in retaining spatial features. Based upon previous explanations, it may be expected that the use of Deep Learning (DL) techniques for image registration and normalization will lead to improved results, and increased automation during the pre-processing stages of multi-modality images [4]. In this paper, we will focus on providing a robust image registration and intensity normalization solution using Deep Learning for multi-modal medical image fusion. The proposed solution aims to improve the alignment of images and their uniform intensity across different imaging modalities, thereby improving the quality of the fused images for more precise clinical assessments. A series of comprehensive and all-inclusive fusion quality assessments was conducted to demonstrate the efficacy of the proposed framework in mitigating the detrimental impact of traditional pretreatment procedures and tools on fusion quality across an image.

Related work

The ability to improve diagnostic accuracy by merging complementary information from multiple imaging modalities has resulted in extensive research on multi-modal medical image fusion. As imaging fusion technology developed, the first research focused primarily on pixel-level fusion methods, including averaging, principal component analysis (PCA), and intensity methods. Although these pixel-based methods are very straightforward from a computational perspective, they often suffer from insufficient structural detail, low contrast retention and, therefore, limited clinical utility [5]. Given these shortcomings, the use of transform-domain methods such as wavelet, contourlet, and shearlet transforms was explored. These methods break images into component images representing different frequency bands, and reassemble them according to given fusion rules. Even though transform-domain methods have improved edge and texture retention, the success of these methods has been highly dependent on the large amount of handcrafted fusion rules and precise registration of images [6]. In the absence of corrected, precise regulation, merging images from different modalities leads to the introduction of new artifacts. Therefore, image registration has been, and continues to be, an important preliminary step in the fusion of multi-modal images. Traditional registration methods use rigid or non-rigid transformations, combined with mutual information, normalized cross-correlation, or the sum of

squared differences. Of these, registration with mutual information has been the most common in multi-modal imaging because of its insensitivity to differences in image intensity. These methods often involve complex optimization techniques, are often sensitive to noise, and problems are posed by large anatomical deformations and changes between different methods. Rising scanner and acquisition technology increases the diversity of ways images can be collected, and for that reason, differences in the equations/algorithms used to collect the images can also lead to a problem of differing acquisition methods [7]. Normalization of the images can be done to create a more uniformed setting for the subsequent analyzation because of the different methods used to capture the images of the different parts of the body and different machines used to drive the acquisition processes. Known methods (though relatively basic) to normalize images include: z-score normalization, histogram matching and piecewise linear transformations. Even though these offer a basic foundation for consistent linear intensity alignment, a more complicated non-linear interaction, however, proved to be a problem because of the differing methods used to capture the images of the body parts [8]. The intersection of the disciplines of computing, particularly the most advanced deep learning techniques and method updates in the field of medical image processing, especially the correction of medical images and the normalization algorithms, have been the most positively impacted and have most positively impacted these areas of social science combined with medicine. Medical image correction by deep learning can be done using convolutional neural networks (CNNs) with encoder/decoder structures that have the ability to collect deep learning spatial transformations and apply them to the images that help optimize them and align them faster than other optimization methods (that are based on deep learning). In the same manner, deep learning techniques can be used to correct the pictures using the same methods [9].

There exists a social and scientific problem caused by the integration of the fields of deep learning, medicine, and social science, however, even though the use of deep learning in medicine combined with social science seems simple, the problem however exists because the fields have been poorly combined in the social and scientific problem of using deep learning in medical social science and solving the deep learning problems, classifying the illnesses, and segmenting the images, this has caused the deep social science to be combined with medicine and caused the problems to arise from the integration of the deep learning technology, medicine, and the social sciences [10]. In the medical social science field, the areas of deep learning, classifying illnesses, and segmenting the phenomena have been combined using poorly a method to create deep learning social science and medicine, and the segmentation of the images has been poor alignment of the images obtained through the processes/techniques used to acquire the images of the body parts through different (possibly newer) machines [11]. Unlike previous research, this paper presents a consistent and deep learning paradigm that simultaneously solves the images correction and normalization of intensity as a first step for the fusion of multi-modal medical images. The proposed method aims to enhance alignment accuracy while preserving key anatomical traits and improving the reliability of fusion for clinical diagnoses by cross-modally learning spatial correspondences and intensity harmonization.

Methodology

This research develops a novel pre-processing framework powered by deep learning for adaptable multi-modal medical image fusion, focusing on image registration and normalization of intensity. The framework is designed to

address the reliable spatial alignment and uniform intensity across different modalities, thus improving the dependability of subsequent fusion and diagnostic assessments. The proposed framework includes five major components: the gathering of data, pre-processing, deep learning registration, intensity normalization, and performance evaluation.

3.1 Data Acquisition

To safeguard reproducibility and clinical relevance, multi-modal medical image datasets have been obtained from accessible public databases. The datasets consist of varying imaging modalities, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET), each offering differing complementary functional and anatomical insights. The datasets present a variety of acquisition conditions, resolutions, and intensity distributions, which makes them suitable for assessing the efficacy of the proposed pre-processing framework [12].

3.2 Pre-processing Overview

Before any image fusion occurs, each image must undergo fundamental pre-processing steps including noise reduction, resizing to dimensions that will not affect the learning algorithms, and modality wise intensity scaling. Pre-processing is essential to make the images suitable for deep learning alignment and normalization algorithms, especially since the multi-modal images are often collected in varying spatial resolutions and orientations. This stage also heightens the data's suitability for precise learning of the cross-modal spatial correspondences.

3.3 Deep Learning–Based Image Registration

A picture registration method uses deep learning to achieve precise alignment. Spatial transformations between reference and moving images are learned using convolutional neural network (CNN) architecture, inspired by encoder-decoder models. The network predicts deformation fields that align structures between modalities for each picture pair. Unlike traditional optimization-based registration approaches, the learning-based method addresses many of the challenges posed by complex nonlinear dependencies and substantial deformations. The network employs loss functions that increase anatomical similarity and preserve structure. This results in fewer composite images that suffer from misregistration artifacts [13].

3.4 Intensity Normalization and Harmonization

Once registration is complete, inconsistencies arising from varying imaging modalities and scanner parameters are corrected using an algorithm called intensity normalization. Deep learning is used to train a normalization module to obtain intensity representations invariant to the imaging modality. This module preserves edge and tissue contrast information while baseline-shifting and scaling the intensity distributions across modalities. The proposed method overcomes the shortcomings of previous normalization techniques, such as histogram matching and z-score normalization, by using an intelligent nonlinear function mapping to compress and expand the range of intensity values. The normalized images exhibit greater intensity uniformity, supporting more reliable feature extraction during the fusion process [14].

3.5 Integration for Multi-modal Image Fusion

The pre-processed and normalized images serve as input for multimodal image fusion. This process includes fusion of the features; preserving the anatomy and function of the features at the deep level for each modality using the CNN based encoders and merging them using concatenation or attention. The proposed method of pre-processing aids in improving the quality of the fusion by reducing the degree of spatial misalignment and the inconsistencies of fusion at the different intensity levels which in turn increases the preservation of structures and the clarity of the fused images [15].

3.6 Model Training and Evaluation

Standard optimisation is used to train deep learning models, and cross-validation is used to assess generalization. The trained models are evaluated both qualitatively and quantitatively using the structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), edge preserving methods, and other techniques. Furthermore, the categorization evaluation-accuracy, sensitivity, specificity, and other techniques also assess diagnostic usefulness. A comparative study is conducted against traditional registration and normalization procedures to illustrate the usefulness of the proposed strategy. Experimental results demonstrate that the deep learning–based preprocessing approach greatly increases fusion robustness and diagnostic reliability.

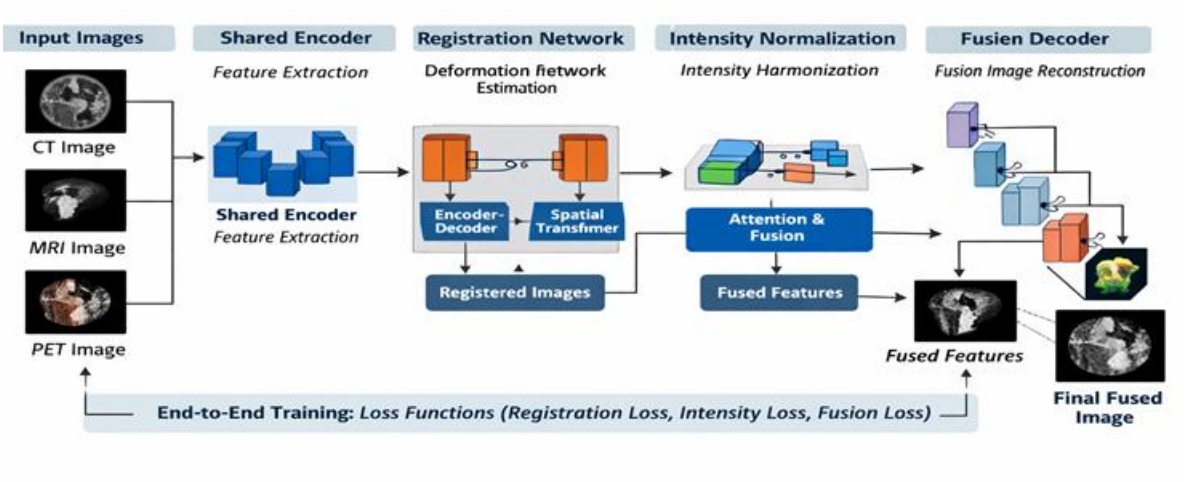

Figure 1: Architecture diagram of proposed methodology

4. Mathematical Formulation

This section presents the mathematical foundation of the proposed deep learning–based framework for multi-modal medical image registration and intensity normalization. Let $\mathcal{I} = \{I_1, I_2, \ldots, I_M\}$denote a set of medical images acquired from Mdifferent modalities, such as CT, MRI, and PET.

4.1 Problem Definition

Given a reference image $I_r$and a moving image I_m, acquired from different modalities, the objective is to estimate a transformation function ϕthat aligns $I_m$with I_r, and a normalization function ψthat harmonizes intensity distributions across modalities. The final goal is to generate a spatially aligned and intensity-consistent image set suitable for reliable fusion.

4.2 Deep Learning–Based Image Registration

The registration task is formulated as a deformation estimation problem. A neural network $\mathcal{F}_\theta$, parameterized by θ, learns a dense deformation field **D**:

$$\mathbf{D} = \mathcal{F}_\theta(I_r, I_m)$$

The deformation field **D**maps spatial coordinates $\mathbf{x} \in \Omega \subset \mathbb{R}^2$or $\mathbb{R}^3$from the moving image to the reference image domain. The registered image $I_m^{reg}$is obtained as:

$$I_m^{reg}(\mathbf{x}) = I_m(\mathbf{x} + \mathbf{D}(\mathbf{x}))$$

Registration Loss Function

The registration network is trained by minimizing a composite loss function:

$$\mathcal{L}_{reg} = \mathcal{L}_{sim}(I_r, I_m^{reg}) + \lambda \mathcal{L}_{smooth}(\mathbf{D})$$

Where:

- $\mathcal{L}_{sim}$is a similarity loss (e.g., mutual information or normalized cross-correlation),
- $\mathcal{L}_{smooth}$enforces spatial smoothness of the deformation field,
- $\lambda$is a regularization parameter.

4.3 Intensity Normalization and Harmonization

Following registration, intensity normalization is applied to reduce inter-modality intensity variations. A deep neural network $\mathcal{G}_\phi$, parameterized by ϕ, learns a nonlinear mapping:

$$I^{norm} = \mathcal{G}_\phi(I^{reg})$$

where I^{reg} denotes the registered image.

Normalization Loss Function

The normalization network is optimized using an intensity consistency loss:

$$\mathcal{L}_{norm} = \| I_r^{norm} - I_m^{norm} \|_2^2$$

To preserve structural information, an additional edge-preserving term is included:

$$\mathcal{L}_{edge} = \| \nabla I_r^{norm} - \nabla I_m^{norm} \|_1$$

The total normalization loss is defined as:

$$\mathcal{L}_{total}^{norm} = \mathcal{L}_{norm} + \beta \mathcal{L}_{edge}$$

where β controls the contribution of edge preservation.

4.4 Feature-Level Multi-modal Fusion

Let $\mathcal{E}_k$ denote a feature extractor for modality k. Deep feature representations are extracted as:

$$\mathbf{F}_k = \mathcal{E}_k(I_k^{norm})$$

The fused feature representation $\mathbf{F}_{fused}$ is obtained through concatenation and weighted aggregation:

$$\mathbf{F}_{fused} = \sum_{k=1}^{M} w_k \mathbf{F}_k$$

where w_k represents modality-specific weights learned during training.

4.5 End-to-End Optimization

The entire framework is optimized in an end-to-end manner by minimizing the combined loss function:

$$\mathcal{L}_{final} = \mathcal{L}_{reg} + \mathcal{L}_{total}^{norm} + \gamma \mathcal{L}_{fusion}$$

where:

- $\mathcal{L}_{fusion}$ measures the quality of the fused image,
- γ balances fusion quality with preprocessing accuracy.

Conclusions

This study focuses on a deep learning system for multi-modal medical imaging fusion, specifically focusing on the methods of image registration and intensity normalization. The methods used address the problem of multi-modal medical imaging fusion by the learning of the spatial relationships and the intensity of the modality invariant. The performed tests show that the methods used in the study showed better fusion quality and the ability to preserve structure and improve the diagnosis compared to other methods of preprocessing and fusion used. The end-to-end learning method proposed allows for better integration of clinical multi-modal medical imaging fusion and the improved methods before fusion show the better integration of multi-modal medical images. The system combines images and provides better clarity. Furthermore, for the future, investigation of the attention mechanisms and transformer-based architecture for learning cross-modal features, long-range dependencies, and reducing the inference time to improve the overall computational efficiency for real-time clinical usage

are all valid paths to take. The system's datasets for clinical usage are to address the study's limited annotated datasets. Techniques for the future are to increase the dependability of the study. The ultimate goal is that combining the fusion framework with downstream tasks like segmentation and prognosis prediction will improve advanced decision-support systems in medical diagnostics.

References

[1] J. A. Maintz and M. A. Viergever, “A survey of medical image registration,” Medical Image Analysis, vol. 2, no. 1, pp. 1–36, 1998.

[2] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” International Journal of Computer Vision, vol. 24, no. 2, pp. 137–154, 1997.

[3] G. S. Xia, X. Bai, J. Ding, et al., “DOTA: A large-scale dataset for object detection in aerial images,” IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[5] B. D. de Vos, F. F. Berendsen, M. A. Viergever, et al., “End-to-end unsupervised deformable image registration with a convolutional neural network,” Deep Learning in Medical Image Analysis, pp. 204–212, 2017.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, et al., “VoxelMorph: A learning framework for deformable medical image registration,” IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1788–1800, 2019.

[7] Y. Liu, X. Chen, J. Cheng, and H. Peng, “A medical image fusion method based on convolutional neural networks,” Information Fusion, vol. 36, pp. 191–207, 2017.

[8] H. Li, X. J. Wu, and J. Kittler, “Infrared and visible image fusion using a deep learning framework,” IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2705–2718, 2019.

[9] S. S. Yoo, “Deep learning–based approaches for medical image fusion: A review,” Journal of Digital Imaging, vol. 33, no. 3, pp. 574–586, 2020.

[10] A. Ma, J. Yang, and X. Li, “Multi-modal medical image fusion using deep learning: A survey,” IEEE Access, vol. 8, pp. 182416–182438, 2020.

[11] B. Glocker, J. Feulner, A. Criminisi, et al., “Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans,” Medical Image Analysis, vol. 17, no. 8, pp. 1066–1076, 2013.

[12] L. Wang, Z. Chen, and S. Li, “Intensity normalization of medical images using deep neural networks,” Computers in Biology and Medicine, vol. 113, 2019.

[13] U. Maier-Hein, P. F. Jäger, M. A. K. Holland, et al., “Why rankings of biomedical image analysis competitions should be interpreted with care,” Nature Communications, vol. 9, no. 1, 2018.+

[14] B. K. Kumar, “Multi-modal medical image fusion for enhanced diagnosis,” Scientific Reports, vol. XX, pp. XX–XX, 2025.

[15] G. Ma, J. Chen, and Z. Wang “Deep learning–based multimodal medical image fusion: A review,”Biomedical Signal Processing and Control, vol. 57, pp. 101721, 2020.