An Unsupervised Image Fusion Approach Using Convolution Neural Network for Feature Enhancement of Medical Brain Images

SUMIT RAJ[1], B. K. SINGH[1], AMITAVA CHOUDHURY[2]

[1]Computer Science and Engineering
National Institute of Technology,Jamshedpur
Jamshedpur, INDIA

[2]Computer Science and engineering
Pandit Deendayal Energy University
Gandhinagar, INDIA

Abstract: Image fusion is the process of amalgamation of two or more images into a single image that contains composite, enriched with diverse details from the original images. In the medical field, image fusion serves as an indispensable tool for elevating the precision of medical imagery and facilitating diagnostic processes. With the advent of deep learning, there has been a significant increase in the accuracy and effectiveness of image fusion techniques. This paper presents a deep-learningbased approach for medical image fusion that combines the advantages of deep-learning techniques with traditional image fusion methods. The proposed method is evaluated on medical data from different modalities, and the experimental results show that the proposed method outperforms existing state-of-the-art image fusion techniques.

Key-words: —supervised, unsupervised, image fusion, spatial frequency

Received: March 19, 2024. Revised: August 23, 2024. Accepted: September 21, 2024. Published: October 30, 2024.

1. Introduction

Medical imaging is one of the crucial tools in modern medicine for the diagnosis and treatment of various diseases. Medical imaging techniques such as CT, MRI, and PET capture images of the body's internal structures. However, the images obtained from these modalities have different characteristics such as resolution, contrast, and noise. To get a complete and accurate diagnosis, multiple modalities are often used. Image fusion is an effective technique for combining information from different modalities to improve the accuracy of diagnosis.

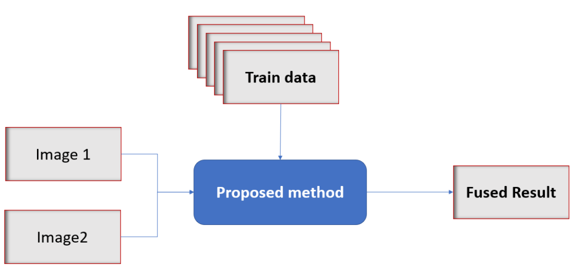

Fig. 1. The Framework of the proposed method.

Traditional image fusion techniques use various methods such as multi-resolution analysis, transform-based, and pixel-based methods. However, these methods have some limitations such as the requirement of prior knowledge of the data, and the resulting images often have artifacts and loss of information. In recent years, deep learning-based techniques have significantly improved various computer vision tasks, including image fusion.

In this paper, we propose a deep learning-based approach for medical image fusion that combines the advantages of deep learning techniques with traditional image fusion methods. The proposed method is evaluated on medical data from different modalities, and the experimental results show that the proposed method outperforms existing state-of-the-art image fusion techniques.

2. Related Work

Various deep learning-based image fusion techniques have been proposed in recent years. Li et al. [1] explain deep learning-based image fusion. Weighted-averaging basic parts and detailed content is the method. Using the l1-norm and weighted-average approach, a deep learning network extracts multi-layer features for the detailed material. The fused base component and detail content are combined using the max selection approach to recreate the fused image. According to experiments, the proposed technique performs well in objective assessment and visual quality.

Xu et al [2] provide a medical image fusion method that resolves blurring edge structure, texture loss, distortion, and slow running speed. The gradient operator optimises weighted guided image filtering's weight factor to preserve edges. A soft threshold fusion rule merges the smooth layer of the MRI image with the PET/SPECT image. The MRI detail layer is selectively updated to improve texture information and merged with the fusion sub-image to create the fused image. The method yields a pleasing image with basic spatial structure, rich texture, and colour information.

Multi-focus image fusion (MFIf) combines images to increase information. This paper introduces Luo et al [3] morphology-based pixel-wise voting multi-focus image fusion. The sliding window gray-scale variance on the multi-focus image is calculated next. Voting matrices are calculated by comparing sliding window gray-scale variance. Voting generates multi-focus decision maps (IFM). Image fusion improves with morphological operations.

Ma et al [4] proposed a new mechanism, Multi-focus image fusion combines images with focused regions which are Unsupervised deep learning for multi-focus image fusion. Unsupervised encoder–decoder networks learn deep image characteristics. Spatial frequency Li et al. [5] and Bo et al. [6] a gradient-based method, measures sudden variation from deep characteristics to estimate activity levels. Updating the decision map and rendering the fused result involves consistency checking. Instead of original images, our method analyses sharp appearances in deep features.

3. Methodology

The proposed deep learning-based image fusion approach consists of three main steps: feature extraction, feature fusion, and image reconstruction.CNN architecture Figure 2 consists of several layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers use filters to extract features from the input image, while the pooling layers downsample the output of the convolutional layers. The fully connected layers at the end of the network use the extracted features to make predictions The feature extraction step uses a pre-trained CNN to extract features from the input images. The feature fusion step fuses the extracted features using a fusion module that combines the features based on their importance. The image reconstruction step generates the fused image by using the fused features to reconstruct the image.

Overall algorithm approach followed

3.1 Training Phase

1) Convolutional neural network and utility were initialized.
2) Image was grayscale,passed through Gaussian blur.
3) Image was passed through CNN to initiate feature extraction.
4) Spatial frequency was used in 5 propagated convolutional neural networks to extract deep features in the inference phase.
5) The two deep features are validated and errors are modified.
6) The features are fused and final image is discarded and training is done again on different images.

3.2 Testing Phase

Testing images were fed and fused images were outputs that were not discarded and were used for calculations, CNN in this phase does not learn.

3.3 Fusion

Here we calculate fused result F using the weight of pixel rule,

$$F_{(y,z)} = DM_{(y,z)} * I1_{(y,z)} + (1 - DM_(y,z)) * I2_{(y,z)} \quad (1)$$

where DM is the decision map, and images 1 and 2 are the input images. **??** displays all of the results and, furthermore, a visual version of the focus results.

$$SD = \sqrt{\frac{1}{P.Q}\sum_{i=1}^{P}\sum_{j=1}^{Q}(f_{i,j} - \xi)^2}. \quad (2)$$

The structure similarity [8] between two input images a and b is measured

$$\text{SSIM}(O, I) = \frac{(2\xi_a\xi_b + c_1)(2\rho_{ab} + c_2)}{(\xi_a^2 + \xi_b^2 + c_1)(\rho_a^2 + \rho_b^2 + c_2)} \quad (3)$$

where, the mean is ξ and the standard deviation is ρ. where C_1 and C_2 are stabilisation constants, ξ_a and ξ_b are the average values of a and b, ρ_a^2 and ρ_b^2 are the variances of a and b, and ρ_{ab} is the covariance between the two regions.

The term ”Peak signal to noise ratio” (PSNR) [8]refers to a measure that indicates distortions by calculating the ratio of peak value-capacity-to-noise-capacity. PSNR is an abbreviation for ”Peak signal-to-noise ratio.”

$$PSNR = 10 \cdot \log_{10}\frac{R^2}{MSE} \quad (4)$$

where ”R” refers to the fused image’s highest value and ”256” is the number that was utilised for the calculation in this study. The ”MSE” or ”mean square error” is a metric that was developed to quantify how dissimilar the fused image is to the source images. The ”MSE” [8] was defined as follows:

$$MSE = \frac{1}{P.Q}\sum_{i=1}^{P}\sum_{j=1}^{Q}(Si - fi)^2 \quad (5)$$

where P indicates the width of the image and Q indicates the height of the image, respectively. S_i and f_i represent, respectively, the source image and the fused images. If the PSNR is higher, then there is less distortion between both the source images and the fused image.

In the feature extraction step, we use a pre-trained CNN and extract features from the input images. The extracted features are then passed through a feature fusion module that combines the features based on their importance. The feature fusion module consists of a set of fully connected layers that learn the importance of each feature and combine them accordingly.

In the image reconstruction step, we use the fused features to reconstruct the image. The reconstruction is performed using a decoder network that takes the fused features as input and generates the fused image. The decoder network can be a simple fully connected network or a deep CNN, depending on the complexity of the images.

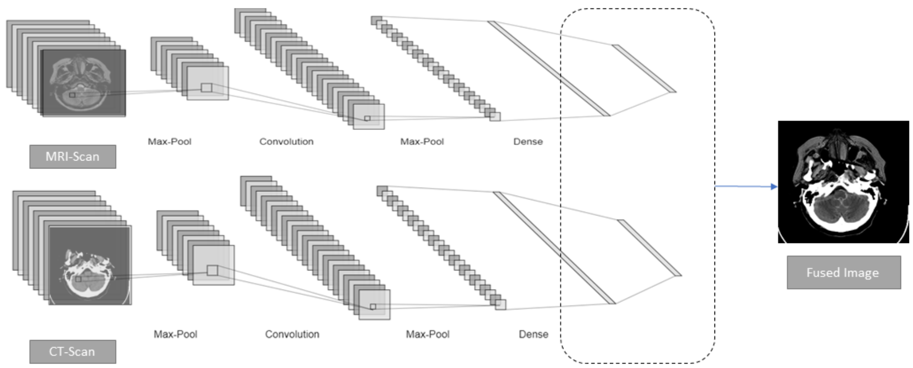

Fig. 2. overall Architecture of the proposed model.

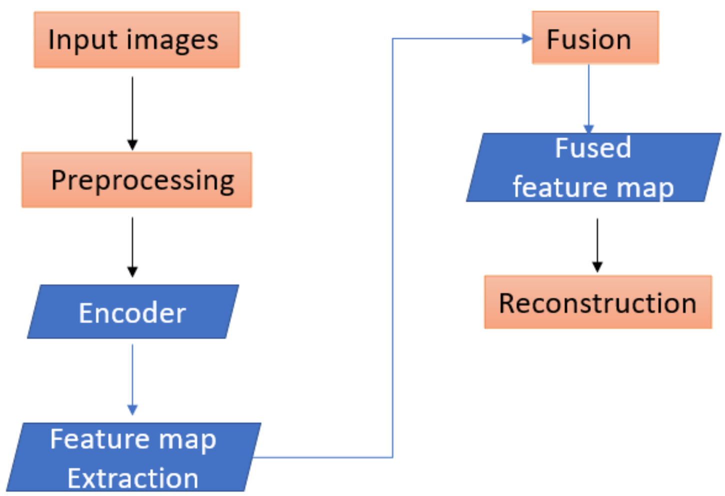

Fig. 3. overall proposed model.

4. Experiment

In this study, multi-modal data sets were used, which included images from the Whole Brain Atlas [7]. The training set consisted of 82784 images, while 40505 images were used for verification of the image reconstruction ability. Among these, 2502 images were selected and cropped to 256×256 to serve as tagged images. To ensure consistency across all images, grey-scale conversion and resizing to 256×256 were applied to all of them.

During the training phase, images were grayscale; however, during testing, images could be either grayscale or coloured using RGB channels. To achieve the required fusion of colour pictures, the researchers converted them to grayscale and then computed a decision map. These steps were taken to ensure that the results obtained from testing were as accurate as possible and consistent with the training phase.

The learning figure was scaled down from its initial value of 0.0001 by a factor of 0.8 per two epochs. Adam optimizer [16] was used to find the optimal value for the objective function with the parameter set to 3. The batch size is 12, and there will be 50 epochs in total. We ran picture fusion on the test set with these settings once training was complete and the trainable variable was fixed. Our code is based on the publicly available PyTorch framework [17]. The network was trained and tested on a computer equipped with an AMD RYZEN-7 CPU, a 4 GB NVIDIA 3050 GPU, and 16 GB of RAM.

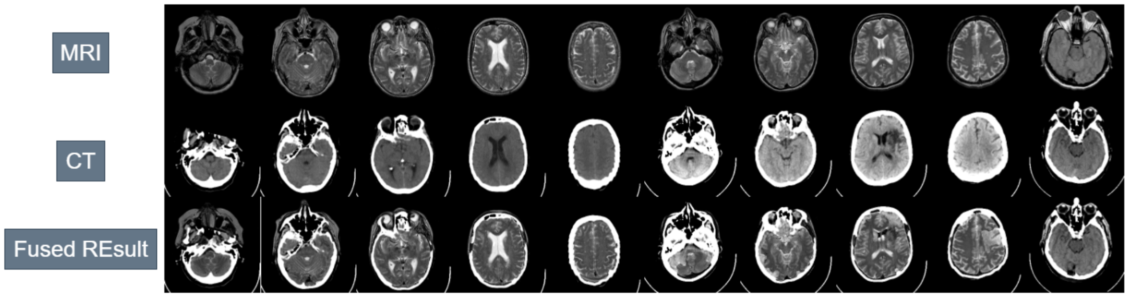

Fig. 4. Final result

5. Results

In Figure 4, the uppermost row showcases MRI images, while the lower row displays CT scan images, and the third row illustrates the outcomes of the image fusion process. Initially, we initiate the feature extraction process to obtain essential feature maps. Subsequently, we compute spatial frequencies, ensuring precision in our analysis. Following this, we engage in meticulous consistency verification, culminating in the creation of final decision maps. Once these decision maps are secured, they are thoughtfully integrated with the deep features extracted from both original images, ultimately yielding the fused images that are presented below.

TABLE I
VALUES OF STANDARD DEVIATION OF DIFFERENT MODELS

Metrics	SSIM	PSNR
CNN [15]	0.702	15.017
CSMCA [9]	0.718	15.593
NSST-PAPCNN [10]	0.694	14.872
U2fusion [11]	0.319	15.012
RCGAN [12]	0.389	14.900
IFCNN [13]	0.748	14.991
EMFusion [14]	0.778	15.026
Present Study	0.826	16.981

We evaluate the proposed method on medical data from different modalities, including CT and MRI. We compare the proposed method with state-of-the-art image fusion techniques, including traditional methods and deep learning-based methods. The evaluation metrics used are peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM)

6. Conclusion

A novel approach to address the issue of obtaining clear brain scan images by combining MRI and CT images. The model employs a deep learning method to estimate the depth of field by merging the deep features extracted from the images. The approach involves training a CNN model in an unsupervised manner to obtain the deep features, which are then fine-tuned in a semi-supervised way using labelled data. The spatial frequency anchor technique is employed to measure the activity levels in the images, which facilitates the effective fusion of the deep features. The experimental results demonstrate promising improvements compared to existing models, highlighting the efficacy of the proposed unsupervised deep learning-based image fusion technique.

References

[1] Li, Hui, Xiao-Jun Wu, and Josef Kittler. "Infrared and visible image fusion using a deep learning framework." 2018 24th international conference on pattern recognition (ICPR). IEEE, 2018.
[2] Xu, Yujie, Huabin Wang, Xiumei Yin, and Liang Tao. "Mri and pet/spect image fusion based on adaptive weighted guided image filtering." In 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), pp. 275-280. IEEE, 2020.
[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] Ma, B., Zhu, Y., Yin, X., Ban, X., Huang, H. and Mukeshimana, M., 2021. Sesf-fuse: An unsupervised deep model for multi-focus image fusion. Neural Computing and Applications, 33, pp.5793-5804.
[5] Li, S. and Yang, B., 2008. Multifocus image fusion using region segmentation and spatial frequency. Image and vision computing, 26(7), pp.971-979.
[6] Xiao-Bo, Q., Jing-Wen, Y., Hong-Zhi, X.I.A.O. and Zi-Qian, Z., 2008. Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain. Acta Automatica Sinica, 34(12), pp.1508-1514.
[7] Keith, A., 1999. The whole brain atlas. http://www. med. harvard. edu/AANLIB/home. html.
[8] Sara, U., Akter, M. and Uddin, M.S., 2019. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications, 7(3), pp.8-18.
[9] Liu, Yu and Chen, Xun and Ward, Rabab K. and Wang, Z. Jane, journal=IEEE Signal Processing Letters, title=Medical Image Fusion via Convolutional Sparsity Based Morphological Component Analysis, year=2019, volume=26, number=3, pages=485-489, doi=10.1109/LSP.2019.2895749
[10] Yin, Ming and Liu, Xiaoning and Liu, Yu and Chen, Xun, journal=IEEE Transactions on Instrumentation and Measurement, title=Medical Image Fusion With Parameter-Adaptive Pulse Coupled Neural Network in Nonsubsampled Shearlet Transform Domain, year=2019, volume=68, number=1, pages=49-64, doi=10.1109/TIM.2018.2838778
[11] Xu, Han and Ma, Jiayi and Jiang, Junjun and Guo, Xiaojie and Ling, Haibin, journal=IEEE Transactions on Pattern Analysis and Machine Intelligence, title=U2Fusion: A Unified Unsupervised Image Fusion Network, year=2022, volume=44, number=1, pages=502-518, doi=10.1109/TPAMI.2020.3012548
[12] Li, Qilei and Lu, Lu and Li, Zhen and Wu, Wei and Liu, Zheng and Jeon, Gwanggil and Yang, Xiaomin, journal=IEEE Sensors Journal, title=Coupled GAN With Relativistic Discriminators for Infrared and Visible Images Fusion, year=2021, volume=21, number=6, pages=7458-7467, doi=10.1109/JSEN.2019.2921803
[13] IFCNN: A general image fusion framework based on convolutional neural network, journal = Information Fusion, volume = 54, pages = 99-118, year = 2020, issn = 1566-2535, doi = https://doi.org/10.1016/j.inffus.2019.07.011, url = https://www.sciencedirect.com/science/article/pii/S1566253518305505, author = Yu Zhang and Yu Liu and Peng Sun and Han Yan and Xiaolin Zhao and Li Zhang,

[14] EMFusion: An unsupervised enhanced medical image fusion network, journal = Information Fusion, volume = 76, pages = 177-186, year = 2021, issn = 1566-2535, doi = https://doi.org/10.1016/j.inffus.2021.06.001, url = https://www.sciencedirect.com/science/article/pii/S1566253521001275, author = Han Xu and Jiayi Ma,

[15] Multi-focus image fusion with a deep convolutional neural network, journal = Information Fusion, volume = 36, pages = 191-207, year = 2017, issn = 1566-2535, doi = https://doi.org/10.1016/j.inffus.2016.12.001, url = https://www.sciencedirect.com/science/article/pii/S1566253516302081, author = Yu Liu and Xun Chen and Hu Peng and Zengfu Wang,

[16] Adam: A method for stochastic optimization, author=Kingma, Diederik P and Ba, Jimmy, journal=arXiv preprint arXiv:1412.6980, year=2014

[17] Paszke, Adam and Gross, Sam and Massa, Francisco and Lerer, Adam and Bradbury, James and Chanan, Gregory and Killeen, Trevor and Lin, Zeming and Gimelshein, Natalia and Antiga, Luca and Desmaison, Alban and Kopf, Andreas and Yang, Edward and DeVito, Zachary and Raison, Martin and Tejani, Alykhan and Chilamkurthy, Sasank and Steiner, Benoit and Fang, Lu and Bai, Junjie and Chintala, Soumith, booktitle = Advances in Neural Information Processing Systems, editor = H. Wallach and H. Larochelle and A. Beygelzimer and F. d'Alché-Buc and E. Fox and R. Garnett, pages = , publisher = Curran Associates, Inc., title = PyTorch: An Imperative Style, High-Performance Deep Learning Library,

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.