

Deep Learning Approach for Fusion of Magnetic Resonance Imaging-Positron Emission Tomography Image Based on Extract Image Features using Pretrained Network (VGG19)

Abstract

Background: The fusion of images is an interesting way to display the information of some different images in one image together. In this paper, we present a deep learning network approach for fusion of magnetic resonance imaging (MRI) and positron emission tomography (PET) images. **Methods:** We fused two MRI and PET images automatically with a pretrained convolutional neural network (CNN, VGG19). First, the PET image was converted from red-green-blue space to hue-saturation-intensity space to save the hue and saturation information. We started with extracting features from images by using a pretrained CNN. Then, we used the weights extracted from two MRI and PET images to construct a fused image. Fused image was constructed with multiplied weights to images. For solving the problem of reduced contrast, we added the constant coefficient of the original image to the final result. Finally, quantitative criteria (entropy, mutual information, discrepancy, and overall performance [OP]) were applied to evaluate the results of fusion. We compared the results of our method with the most widely used methods in the spatial and transform domain. **Results:** The quantitative measurement values we used were entropy, mutual information, discrepancy, and OP that were 3.0319, 2.3993, 3.8187, and 0.9899, respectively. The final results showed that our method based on quantitative assessments was the best and easiest way to fused images, especially in the spatial domain. **Conclusion:** It concluded that our method used for MRI-PET image fusion was more accurate.

Keywords: Convolutional neural network, hue-saturation-intensity space, image fusion, VGG19

Submitted: 12-Dec-2020

Revised: 01-Jan-2021

Accepted: 14-Jun-2021

Published: 28-Dec-2021

Introduction

Fusion images are a combination of images from different modalities into one image for a different purpose. Medical images have different modalities, and each of them contains specific information and application for physicians. In clinical diagnosis, physicians usually need different image modalities such as X-ray based images, such as radiography and computed tomography (CT), or radiofrequency waves based, such as magnetic resonance imaging (MRI) that all these methods are used to capture high-resolution anatomical images within patient bodies, and also images such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT) that provide low spatial resolution functional images. Hence, with image fusion, physicians

can see two different modalities in one image. Many studies have been published in different areas to fused two different modalities.^[1,2] In recent years, deep learning networks have many applications in different fields such as computer vision and image processing problems such as classification,^[3] segmentation,^[4] registration,^[5] super-resolution.^[6] There are a lot of methods that fused medical images as PET, SPECT, MRI, and CT.^[7-10] Image fusion based on deep learning methods has also become a new common topic. These methods have also been used in digital photography, multi-focus image fusion, multi-modality imaging, medical image fusion, and infrared/visible image fusion.^[11-14] Artificial neural networks (ANNs) have many applications in different areas, including recognizing the good characteristics of data features. The most important advantage of the ANN is its high learning ability and used to

**Nasrin Amini,
Ahmad Mostaar**

*Department of Biomedical
Engineering and Medical
Physics, Faculty of Medicine,
Shahid Beheshti University of
Medical Sciences, Tehran, Iran*

Address for correspondence:

*Dr. Ahmad Mostaar;
Department of Biomedical
Engineering and Medical
Physics, School of Medicine,
Shahid Beheshti University of
Medical Sciences, Tehran, Iran.
E-mail: mostaar@sbmu.ac.ir*

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Amini N, Mostaar A. Deep learning approach for fusion of magnetic resonance imaging-positron emission tomography image based on extract image features using pretrained network (VGG19). J Med Sign Sens 2022;12:25-31.

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_80_20

Quick Response Code:



solve complex problems in various types of research. For classification problems, deep learning algorithm spatially convolutional neural network (CNN) helps to minimize preprocessing level. The input layer, assigns importance (learnable weights and biases) to various features in the input image, then hidden layer applied a convolutional action on the images then bring the result to the next layer, finally network be able to differentiate images to different classes. Machine learning algorithm tools have a key role in improving the automatic analysis of images. It is a new method in medical image analysis that has not been before.^[14] There are various CNN structures that have been used in building algorithms such as AlexNet, VGG16, VGG19, and GoogLe Net.^[15-17] Some articles used deep learning for fusion of different kinds of images.^[18-20]

In this paper, we used VGG19 that is trained on more than a million images from the ImageNet database. VGG19 with 19 layers has the ability to classify images into 1000 different classes, such as the keyboard, mouse, pencil, and many animals. In this paper, we have presented a diagnostic medical image fusion method (MRI and PET) by using VGG19 network. The rest of this paper is organized as follows: In Section 2, we briefly have explained deep learning technology and our methods for MRI-PET image fusion. The experiments and results are presented in Section 3. We have concluded the paper in Section 4.

Materials and Methods

Data

Our dataset in this study consists of 30 images of color PET images and 30 images of high-resolution MRI images of the brain that register together. All images used in this article were obtained from the Harvard University site (<http://www.med.harvard.edu/AANLIB/home.html>). The dataset images of the brain exist in two groups: normal (coronal, sagittal, and transaxial) and Alzheimer's disease dataset images. PET images have three bands red-green-blue (RGB) based on metabolic processes in the body, and MRI images are high-resolution black and white images. Some sample of our dataset (MRI and PET images) is shown in Figure 1. To create the same conditions, as preprocessing, we resized all images to 256×256 .

Deep learning

In this part, we introduce the CNN and the structure of the pretrained classifier. CNN is a neural network with multilayer and learnable weights and biases that it can apply to a set of images, to be able to differentiate images from each other. In recent years, a lot of attention in the world of science has been toward to application of CNN in image classification. In this part, we tried to give a review of CNN in the area of image classification. CNN has different types of layers: image input layer, convolutional layer, ReLU layer, pooling layer, fully connected layer, and classification layer.^[3,21] The images are read in the

input layer after resizing the image size, and then the convolutional layer uses some filters (kernels) on the images. These filters move through all parts of the input image and convolve them into a single position (output image). The ReLU layer is the rectifier linear unit and normalization layer that is commonly used as an activation function. Pooling means downsampling of an image that reduces the image size and removes redundant information. There are different pooling techniques. The most-used pooling functions include max pooling, min pooling, and average pooling. A fully connected layer is the last layer with the same structure of the traditional neural network that each hidden layer with the number of neurons connects to all the previous layers. The neurons in each layer act independently and have no connection to each other. The last fully connected layer is known as the output layer that combines all the features and usually plays the role of representing the score of each class.^[21] Figure 2 shows the overall architecture of CNN (VGG19).^[22]

Recently, pretrained image classification networks are common and useable for image classification. Database of these networks contains more than a million images in 1000 different classes, such as the keyboard, coffee mug, pencil, and many animals. Because these networks are already trained, using these networks is much faster and easier. In 1998, LeCun *et al.* introduced the first pretraining to classify handwritten digits. Their CNN model is called LeNet-5.^[23] In 2012, Krizhevsky *et al.* designed a pretrained CNN (AlexNet) that used to classify ImageNet dataset. The AlexNet is very similar to LeNet-5 but much bigger.^[24] There are a lot of CNNs (like VGG19) that used ImageNet database (a million images) to trained.^[24]

Fusion rules

After selecting or extracting the appropriate features, the next step was to fuse the images using the rules. Fusion rules can be simple average, weighted average, maximum variance, minimum energy, etc. These rules are determined according to the features. Therefore, a part of the original images is placed as a result of fusion according to the fusion rule. For example, in simple averaging, the pixel values of two images are averaged. In this article, we used a kind of weighted average as fusion rule that the weighted coefficients were extracted from the VGG19 network, and also two constant coefficients were used.

Our method

In this study, we used VGG19, a pretrained network for image fusion. In our application, we extracted features from the learned image of the VGG19 network and used those features (weights from the first layer) to fusion two images in the spatial domain. Feature extraction is one of the application pretrained deep networks that these features could use for a different destination. Fusion rules must be applied to fused two images. Our fusion rules were the

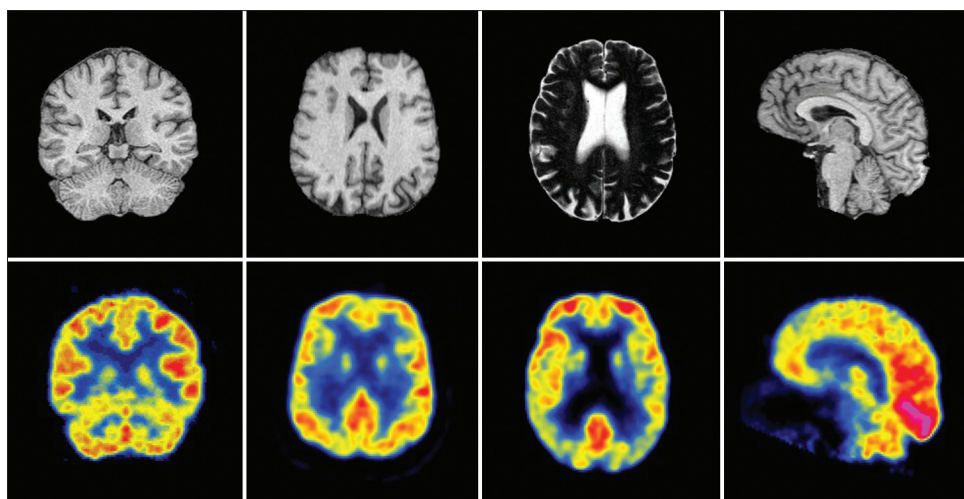


Figure 1: Some sample of data set of magnetic resonance imaging (top row) and positron emission tomography (bottom row) images from the Harvard University site (<http://www.med.harvard.edu/AANLIB/home.html>)

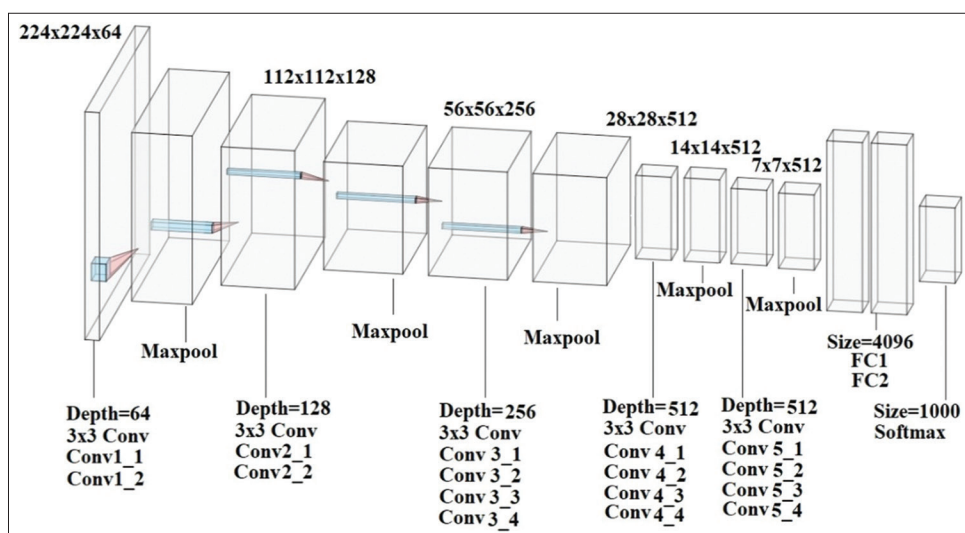


Figure 2: The overall architecture of the convolutional neural network VGG19^[22]

inner product of MRI and PET weights (from the VGG19 network) to original images. The formula of our fusion rules is given by Eq. 1:

$$\text{Fused image } (i, j) = (A \times \text{MRI } (i, j)) + (B \times \text{PET } (i, j)) + (W_{\text{MRI}}(i, j) \cdot \text{MRI } (i, j) + W_{\text{PET}}(i, j) \cdot \text{PET } (i, j)) \quad (1)$$

Where MRI (i, j) and PET (i, j) are original images, A and B are two constants that used for adding an original image with coefficient to fusion result, and $W_{\text{MRI}}(i, j)$ and $W_{\text{PET}}(i, j)$ are the weight matrixes that extracted from VGG19 network. The features extracted from the first layer of the VGG19 network were [256, 256, 64] which we just used the first three features [256, 256, 3] as weights. The purpose of using inner product was to perform fusion in a very simple way so that the appropriate weights of each image are used.

Therefore, details of the steps of PET and MRI image fusion based on the VGG19 CNN can be summarized as follows.

Step 1: Registration is the first step of the fusion process of two medical images with different modalities.

Step 2: PET image needs to be converted from RGB space to hue-saturation-intensity (HSI) space and just we used intensity for fusion.

Step 3: Two images separately are introduced to the VGG19 network and weights are extracted from the first layer of the network (VGG19 architecture has 47 layers). The extracted weights from the first layer of the VGG19 network were [256, 256, 64] which we just used the first three of them [256, 256, 3] as weights.

Step 4: Weight matrixes were normalized between zero and one.

Step 5: For fusion rules, these normalized weights multiplied to original images with an inner product. Then, we added original images with a constant coefficient to the fusion result (Eq. 1).

Step 6: The final fused image is considered intensity (I) and returned to RGB space.

The block diagram structure of our proposed fusion method is shown in Figure 3.

Results

As we mentioned, our dataset consisted of 30 images, and all images have been downloaded from the Harvard University site (<http://www.med.harvard.edu/AANLIB/home.html>). At first, all images in our dataset were registered. Then, all images of MRI images and PET images resized to 256×256 . To extract the weights from the network, we used the “VGG19” toolbox in MATLAB software (version 2019a, The MathWorks Inc., Natick, Massachusetts, United States). Extracted weights were from the first layer of the VGG19 network. Our constant coefficients that are considered in the fusion formula were $A = 0.3$ and $B = 0.1$.

We compared some other methods with our method in the spatial domain such as pixel averaging, HIS based, CNN, and the proposed method (VGG19), as well as in transform domain such as Laplacian pyramid, wavelet transform, curvelet transform (CVT), contourlet transform, and nonsubsampling contourlet transform (NSCT).^[9] In our method and CNN, we used features from the first layer of the VGG19 network. All images in the transform domain are decomposed into four levels by DWT (db2), CVT, NSCT, and our method. The MRI and PET images and fusion results are shown in Figures 4 and 5 for normal and diseased brains, respectively. For comparison, we used entropy, mutual information, discrepancy, and overall performance (OP) as quantitative evaluation.

The average of the quantitative measurements for 30 images is listed in Table 1. To show we have a high quality of fusion, we should have the lowest amount of discrepancy and OP, as well as the highest amount of entropy and mutual information.

Table 1 shows methods in the spatial domain have fewer OP and discrepancy and higher mutual information in comparison with methods in the transform domain, but the situation is opposite about entropy and the transforms method has better entropy. Another problem with transform domain methods is more complex and time-consuming. Mutual information of our method increased 17% in comparison with NSCT (best amount in transform domain methods). In spatial methods, the mutual information of pixel averaging is better, but the problem of this method is reducing contrast and spectral distortion. Discrepancy and OP of our method were the smallest. We decreased them 24% and 11% in comparison with transform domain methods, respectively.

Discussion

MRI-PET fusion is the process of integrating information from two images into a single image that is more appropriate for a physician to have visual perception or computer analysis. The purpose of this study was to find the best features from both PET and MRI images by using a pretrained CNN (VGG19) to fuse them.

Compared to the results of the study carried out by Haddadpour *et al.* that they fused the PET and MRI images by using Hilbert transform, our method had fewer discrepancy and overall performance.^[25] Further, compared

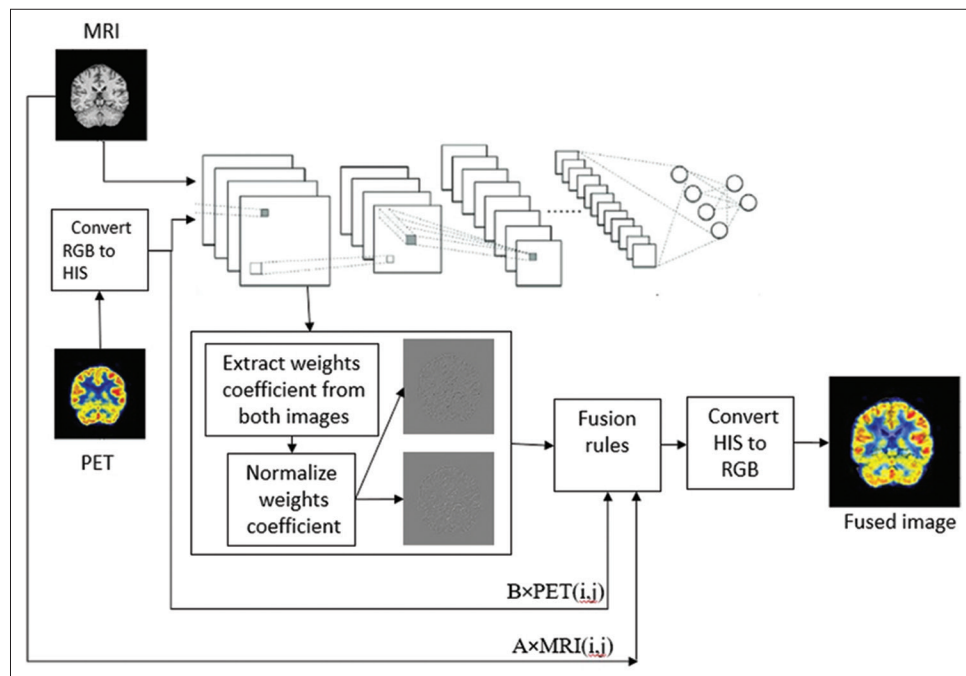


Figure 3: Block diagram of our method based on VGG19 network

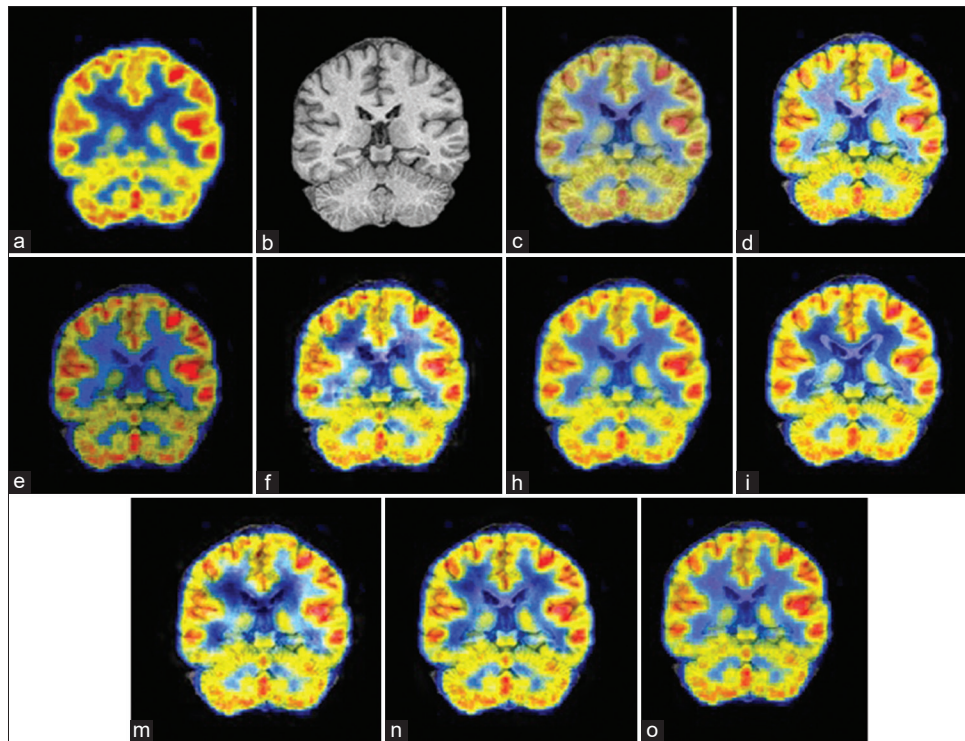


Figure 4: Normal positron emission tomography and magnetic resonance imaging images (a and b), averaging (c), hue, saturation, intensity model (d), convolutional neural network method (e), Laplacian pyramid (f), wavelet transform (h), curvelet transform (i), contourlet transform (m), nonsubsampling contourlet transform (n), and the proposed method (VGG19) (o). (c-n)^[9]

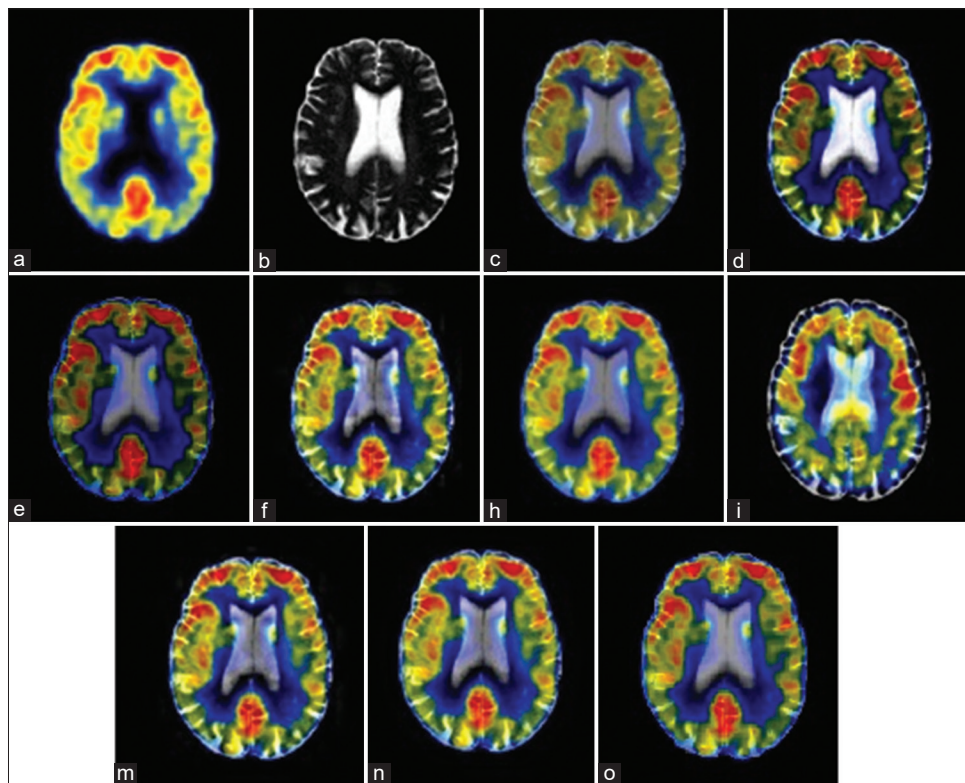


Figure 5: Alzheimer's disease positron emission tomography and magnetic resonance imaging images (a and b), Averaging (c), hue, saturation, intensity model (d), convolutional neural network method (e), Laplacian pyramid (f), wavelet transform (h), curvelet transform (i), contourlet transform (m), nonsubsampling contourlet transform (n), and the proposed method (VGG19) (o). (c-n)^[9]

to results of Javed *et al.* that they fused the PET and MRI images by using fuzzy logic and image local features, our

method had more amount of mutual information and fewer entropy.^[26] The advantages of our method are using the

Table 1: Quantitative results of transform and spatial methods on previous study^[9] in comparison with our method

Domain	Method	Entropy	Mutual information	Discrepancy	Overall performance
Transform	Wavelet	3.3285	1.8502	6.0827	1.7598
	Curvelet	3.3090	1.9159	5.7767	1.1166
	Contourlet	4.1828	1.7949	7.4307	2.5320
	NSCT	3.3794	1.9901	5.0510	0.8707
Spatial	Pixel averaging	3.1141	2.8881	6.1358	3.2292
	HSI based	2.7465	2.3796	7.8387	3.3772
	Our method (deep learning based method)	3.0319	2.3993	3.8187	0.9899

Blue color showing the best result of transform-based methods, and red color showing the best result of spatial-based methods. NSCT – Nonsubsampled contourlet transform; HIS – Hue, saturation, intensity

pretrained models that have been trained on a large dataset, so it can be a good way for the features extraction method of original images. Further, it is easy and fast to use them for image fusion because it does not have complicated calculations.

Our challenge in this paper was that the extracted coefficients from the network were not sufficient for image fusion and the final image had a low brightness; hence, in addition to those coefficients, we add two constant coefficients to our method which were obtained experimentally and by experimenting with different numbers. The disadvantage of our method is that these constant coefficients cannot be generalized to all images, and for other images with different natures, these constant coefficients must be obtained.

Based on the quantitative comparison, the proposed methods had the best results in discrepancy and have fewer OP in the spatial domain methods. For our dataset, spatial methods had better mutual information rather than transform methods, and for entropy, transform methods were better. These results enable us to affirm the effectiveness and robustness of deep learning methods for image fusion purpose.

Conclusion

We presented a spatial domain method for PET and MRI image fusion based on the CNN. Our novelty is extracting weights from a pretrained network and uses these weights as features of our dataset. Hence, our process was so fast and easy with an acceptable result. We used VGG19, a pretrained network for image fusion. We extracted features from the learned image of the VGG19 network and used those weights from the first layer to fusion two images. Fusion rules must be applied to fused two images. Our fusion rules were an inner product of MRI and PET weights extracted from the VGG19 network to original images. MRI-PET fusion is the process of integrating information from two images into a single image that is more appropriate for the physician to have visual perception or computer analysis. We employed that the spatial domain methods included averaging, HSI-based, CNN, and our method and transform domain methods included wavelet transform, CVT, contourlet transform, and NSCT to fuse

the MRI and PET images. Fusion-based spatial domains lead to reduced contrast, and we solved this problem as we mentioned. The problem of transform domain methods is long running time, complexity, and higher discrepancy. The values obtained from mutual information, discrepancy, and OP show that deep learning method is a very good method. A future study may be using other deep learning methods into SPECT, PET, and MRI images.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

1. Masood S, Sharif M, Yasmin M, Shahid MA, Rehman A. Image fusion methods: A survey. *J Eng Sci Tech Rev* 2017;10:186-95.
2. Mozaffarilegha M, Yaghobi Joybari A, Mostaar A. Medical Image Fusion using bi-dimensional empirical mode decomposition (BEMD) and an Efficient Fusion Scheme. *J Biomed Phys Eng* 2020;10:727-36.
3. Sultana F, Sufian A, Dutta P. Advancements in image classification using convolutional neural network. In: Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). Kolkata, India: IEEE; 2018.
4. Zhou T, Ruan S, Canu S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* 2019;3-4:100004.
5. Kuppalu K, Banda S, Barige TR. An overview of deep learning methods for image registration with focus on feature-based approaches. *Int J Image Data Fusion* 2020;11:113-35. [doi: 10.1080/19479832.2019.1707720].
6. Hayat K. Multimedia super-resolution via deep learning: A survey. *Digit Signal Process* 2018;81:198-217.
7. Alipour SH, Houshyari M, Mostaar A. A novel algorithm for PET and MRI fusion based on digital curvelet transform via extracting lesions on both images. *Electron Physician* 2017;9:4872-9.
8. Amini N, Fatemizadeh E, Behnam H. MRI and PET image fusion by using curvelet transform. *J Adv Comput Res* 2014;5:23-30.
9. Amini N, Fatemizadeh E, Behnam H. MRI-PET image fusion based on NSCT transform using local energy and local variance fusion rules. *J Med Eng Technol* 2014;38:211-9.
10. Saboori A, Birjandtalab J. PET–MRI image fusion using adaptive filter based on spectral and spatial discrepancy. *Signal Image Video Process* 2019;13:135-43.

11. Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey. *Inf Fusion* 2019;45:153-78.
12. Li S, Kang X, Fang L, Hu J, Yin H. Pixel-level image fusion: A survey of the state of the art. *Inf Fusion* 2017;33:100-12.
13. Patel H, Upla K. Survey on image fusion: Hand designed to deep learning algorithms. *Asian J Convergent Technol (AJCT)* 2019;5:1-9.
14. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. In: *Classification in BioApps. Automation of Decision Making*: Springer; 2018. p. 323-50.
15. Khan S, Rahmani H, Shah SA, Bennamoun M, Medioni G, Dickinson S. A guide to convolutional neural networks for computer vision. *Synth Lect Comput Vis* 2018;8:1-207.
16. Shridhar K, Laumann F, Liwicki M. A comprehensive guide to bayesian convolutional neural network with variational inference. *arXiv* 2019;13:3-14.
17. George A, Routray A. Real-time eye gaze direction classification using convolutional neural network. In: *International Conference on Signal Processing and Communications (SPCOM)*. IEEE, Bangalore, India; 2016. p. 1-5.
18. Liu Y, Chen X, Wang Z, Wang ZJ, Ward RK, Wang X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf Fusion* 2018;42:158-73.
19. Rajalingam B, Priya R. Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. *Int J ChemTech Res* 2018;11:160-76.
20. Piao J, Chen Y, Shin H. A new deep learning based multi-spectral image fusion method. *Entropy (Basel)* 2019;21:570.
21. Véstias MP. A Survey of Convolutional Neural Networks on Edge with Reconfigurable Computing. *Algorithms* 2019;12:154.
22. Sudha V, Ganeshbabu TR. A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Comput Mater Con.* 2021;66:827-842.
23. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278-324.
24. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *The Proceedings of the 25th International Conference on Advances in Neural Information Processing Systems*. Lake Tahoe, NV: USA; 2012. p. 1097-105.
25. Haddadpour M, Daneshvar S, Seyedarabi H. PET and MRI image fusion based on combination of 2-D Hilbert transform and IHS method. *Biomed J* 2017;40:219-25.
26. Javed U, Riaz MM, Ghafoor A, Ali SS, Cheema TA. MRI and PET image fusion using fuzzy logic and image local features. *Scientific World Journal* 2014;2014:Article ID: 708075, 1-8.