MULTIMODAL MEDICAL IMAGE FUSION USING HYBRID DOMAINS

A.RAJESH NAIDU, D.BHAVANA *

Abstract. In a variety of clinical applications, image fusion is critical for merging data from multiple sources into a single, more understandable outcome. The use of medical image fusion technologies to assist the physician in executing combination procedures can be advantageous. The diagnostic process includes preoperative planning, intra operative supervision, an intervensional treatment. In this thesis, a technique for image fusion was suggested that used a combination model of PCA and CNN. A method of real-time image fusion that employs pre-trained neural networks to synthesize a single image from several sources in real-time. A innovative technique for merging the images is created based on deep neural network feature maps and a convolution network. Picture fusion has become increasingly popular as a result of the large variety of capturing techniques available. The proposed design is implemented using deep learning technique. The accuracy of the proposed design is around 15% higher than the existing design. The proposed fusion algorithm is verified through a simulation experiment on different multimodality images. Experimental results are evaluated by the number of well-known performance evaluation metrics

Key words: Image fusion, PCA, CNN, VGG 16, VGG19

AMS subject classifications. 68U10

1. Introduction. Medical imaging may provide information in a number of ways. For clinical diagnosis, it is vital to keep an eye on the distinguishing aspects of different medical images. Image fusion is a method used to produce a single image by integrating the features of numerous image sensors. This information may come from a single source across many time periods or from a number of sensors during a single period. Fusion is the process of joining two or more things together to form a single entity[4],[10]. To improve the use of medical images and assist clinicians in interpreting image content, medical image fusion attempts to extract as much useful information as possible from source images[8]. Image fusion is the process of fusing two or more images into a series of images that incorporates the data from the individual images. The result is an image with higher information content than any of the input images. The purpose of image fusion is to create images that are more acceptable and understandable to both humans and machines.

MRI scanners employ magnetism, high magnetic fluctuations, and radio waves to generate images of the inside of the body’s tissues[5]. An MRI scan that detects structural issues before they become irreparable may help to avoid nerve damage. Neuropsychological testing is often used to identify neural injury [18]. Medical image fusion was used to merge images from several imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and proton emission tomography (PET) [6].

To eliminate superfluous data, source register images are joined with other important data in an image fusion process. To do quantitative analysis, entropy and other statistical approaches are applied. Liguo Zhang has developed a complete deep learning model for extracting features, merging features, and restoring pictures that does not involve the development of complex feature matching and fusion criteria [1]. According to the Associated Press, Jamesa provides a comprehensive list of possibilities as well as an explanation of the several scientific obstacles that medical image fusion faces [2].

Deep learning has been used to address some of the most difficult challenges in medical imaging. Simultaneous PET/MR imaging combines the soft-tissue contrast of MR imaging with the molecular sensitivity and specificity of PET. Obtaining a valid photon attenuation correction (AC) map, which is critical for accurate PET quantification, is still a difficulty [3] [19-22]. Because traditional MR imaging-based AC (MRAC) techniques lack direct bone measurement, a number of new methodologies have been used.[15] PET and SPECT imaging
applications of AI-based algorithms range from low-level electronic signal formation and processing to high-level internal dosimetry and diagnostic/prognostic modelling. Deep learning techniques have mostly been employed to improve incident photon timing resolution and localization accuracy with the objective of enhancing overall spatial and time-of-flight (TOF) resolutions in PET. The purpose of the research being done on quantitative SPECT and PET imaging right now is to get rid of the effects of noise, artifacts, and movement [16].

A wide variety of computer vision problems, including identification, fragmentation, and very low-level recognition, have been improved through deep learning. It’s called image fusion when two or more separate images are combined to create a new image. To enhance the original multi-spectral image’s resolution, it takes data from several sources while still conserving spectral information. Radiology uses computed tomography, or CT, to produce detailed images of the inside of the human body without the need for any incisions [9]. It’s impossible to view your bones, muscles, and other biological parts in the same way that this image does. Tumors may be identified on CT scans by their shape, size, and location.

Utilize contourlet transforms and adaptive weighted PCA in the frequency domain to create a multimodal fusion. Describe the fusion mechanism for the DTCWT and SOFM modalities. Describe the use of NSCT-based image fusion for clinical research. For the purpose of improving input photos, the hybrid fusion method is introduced. Review and elucidate the fusion algorithms overview. developed a hybrid approach for accurate treatment analysis based on a blend of curvelet and wavelet transforms. To combine structural and functional information from the source images, introduce the fusion method. A-PCNN has been developed, and NSST has been suggested for multimodality medical pictures [8]. Describe the NSCT for multimodality medical image fusion. a curvelet-based fusing algorithm was invented [10]. An algorithm for fusing multimodal medical images has been created [11]. The DFRWT technique is described here for effective image fusion [12]. Create and modify the surface of the fusion system [13]. Describe the fusion technique for medical pictures based on SWT and NSCT [14].

2. Proposed Methodology.
In this paper, the proposed method is a combination of CNN and PCA. Radiological images were collected, followed by normalization, which enhances both computation efficiency and execution. The normalised images were then fed into the CNN algorithm for fusion, and the fused images were then gathered and submitted to the PCA method for dimensionality reduction. the whole process of proposed model is shown in figure 2.1. As a result, the technique is now known as a CNN + PCA combine model, which produces superior results than other techniques. The normalized image is submitted to VGG16, VGG19, SqueezeNet, and PCA to verify the accuracy of alternative algorithms for image fusion, although the results are less accurate than the combination model of CNN+PCA. For image fusion, several models of radiological imaging such as MRI, CT, PET, and SPECT were collected. Two different format photographs are used in this image fusion.

2.1. Convolutional Neural Network. Convolutional networks, also known as convolutional neural networks, are a type of neural network that uses a grid-like design to process input (LeCun, 1989). In real-world applications, convolutional networks have had a lot of success. A "convolutional neural network" is a network that uses the convolutional mathematical approach as shown in figure 2.2. It’s a very specialized linear procedure. Simple neural networks with a minimum of one layer that simply use convolution instead of generic matrix multiplication are used in these networks. Convolutional networks were among the first neural networks to solve key business problems, and they continue to be at the core of commercial deep learning applications today.[14] A convolutional neural network consists of several layers which are classified as the Input layer, Convolutional layer, Max Pooling layer, flatten layer, Output layer which helps us to extract features for the fusion of images. A tiny portion of the visual field reacts to stimuli in the Receptive Field which are known as neurons. A ring of comparable fields forms around the entire field of vision.

2.2. VGG16. It is a deep Convolutional Neural (CNN) architecture with several layers. VGG means Visual Geometry Group. VGG-16 is a deep convolutional neural network with 16 layers. It has a very appealing architecture due to its uniformity. It features only 3x3 convolutions, but a lot of filters, similar to AlexNet [19]. It may be taught for two to three weeks on 4 GPUs. It is now the most popular method for extracting characteristics from images in the communityIn ImageNet, The top-5 precision of the VGG16 model was 92.7 percent. The VGG16 Architecture was invented and presented by Karen Simonyan with Andrew Zisserman of
Collecting Radiology Images

Normalizing those Images

Keeping the images in a list

Applying CNN Algorithm

Applying Inception Learning Technique

Fused Image

Applying PCA to the Fused Image

By comparing with different metrics
finalizes the best Algorithm for fusion

Fig. 2.1: Proposed Model

Fig. 2.2: CNN sequence to classify handwritten digits

Oxford University in their article "Very Fully Convolutional Networks for Large Object Recognition" in 2014.

2.3. VGG19. The VGG19 model is a VGG version with nineteen layers which is a combination of sixteen convolution layers, three fully linked layers, five max pool layers, and one Soft Max layer. The VGG19 (also known as VGGNet-19) is a model similar to VGG16 only the difference is it contains 19 layers [11,13]. The weight layers of the model are represented by the digits "16" and "19". In comparison to VGG16, VGG19 has three more convolution layers. VGG19 is a complex CNN having pre-trained phases and a deep understanding of how a picture is described in terms of appearance, color, and structure. VGG19 is a neural network that was learned on millions of images to solve difficult classification issues. For a wide range of images, the system has learned rich feature representations.
2.4. **SqueezeNet.** SqueezeNet is a deep convolutional neural network with eighteen layers. SqueezeNet is a form of neural network that uses design tactics to minimise dimensionality, most notably the usage of flame components, which use $1x1$ CNNs to "squeeze" variables. SqueezeNet was created with the goal of creating a compact neural network with fewer elements that could fit into memory storage and be conveyed more easily over a computer network. SqueezeNet is a small network designed to be a more compact version of AlexNet. It has more than 50 times less parameters than AlexNet yet is three times faster [12]. The SqueezeNet design is made up of the "squeeze" and "expand" layers. In a squeeze convolutional layer, only one filter is used. In an extended layer, these are routed via a mix of $1x1$ and $3x3$ convolution filters. SqueezeNet was trained with a learning method of 0.04, which lowers linearly over time. SqueezeNet makes the setup process easier because of its small size. This network was created in Caffe, but it has since gained popularity and has been adopted by a number of platforms.

2.5. **Principal Component Analysis.** Principal Component Analysis is an unsupervised learning approach for reducing dimensionality in machine learning. It is a mathematical method that uses orthogonal information to turn correlated attribute outputs into a collection of uncorrelated, linear qualities. The Principal Components are the newly amended features.

It is necessary to limit the number of variables and interpret linear combinations of data in order to understand data with a large number of variables in a meaningful way. PCA is a method for reducing a large number of potentially associated variables to a smaller number of variables known as principal components [7]. A method for condensing a large number of samples into a smaller set that preserves the majority of the data from the larger set is principal component analysis. We can utilize the principal component analysis technique to create and use a smaller set of variables called primary factors. It’s much easier to research and interpret a smaller set of data.

3. **Performance Evaluation.** The source images are obtained using several radiological techniques from a normal brain. The fuse output is formed again by integrating its inverted output of the resultant connection with actual U and V segments, using real brain MRI and PET input images as a source for testing. For source images, the process is simple to implement; nevertheless, the fusion result is not flawless. The fusion image is unstable because the information in the fused image changes from one algorithm method to the next, while the data image has many modalities depending on the algorithm utilized. The hue of the PCA image changes when
compared to other fused images. Because we are using PCA, the generated fused image will be in grayscale mode. In this section, we will discuss the comparative analysis of different algorithm methods using various metric indexes. The comparative analysis is based on source images and fused image with help of comparing algorithm methods including the add-weighted average of them. The fused images were created utilizing several methods such as VGG-16, VGG-19, Squeeze Net, PCA, and CNN+PCA.

The following metrics can be used to assess the performance of image fusion algorithms:

**Mean Square Error** [17] (MSE):

$$MSE = \frac{1}{X \times Y} \sum_{m=0}^{X} \sum_{n=0}^{Y} (S_1(m,n) - S_f(m,n))^2$$

(3.1)

where $X \times Y$ image size. $S_1(m,n)$ is the input image and the fused image.

**Structural Similarity Index Measure** [17] (SSIM)

$$SSIM = \frac{(2\mu_{S_r}\mu_{S_f} + K_1)(\sigma_{S_r}\sigma_{S_f} + K_2)}{\mu_{S_r}^2 + \mu_{S_f}^2 + K_1}$$

(3.2)

where $\mu(S_r)^2 + \mu(S_f)^2$ is close to 0, the constant $K_1$ is used to prevent instability and when $\sigma(S_r)^2 + \sigma(S_f)^2$ is close to 0, the constant $K_2$ is used to prevent instability.

$$\mu_{S_r} = \frac{1}{XY} \sum_{m=0}^{X} \sum_{n=0}^{Y} S_r(m,n)$$

$$\mu_{S_f} = \frac{1}{XY} \sum_{m=0}^{X} \sum_{n=0}^{Y} S_f(m,n)$$

$$\sigma_{S_r}^2 = \frac{1}{XY-1} \sum_{m=1}^{X} \sum_{n=1}^{Y} (S_r(m,n) - \mu_{S_r})^2$$

$$\sigma_{S_f}^2 = \frac{1}{XY-1} \sum_{m=1}^{X} \sum_{n=1}^{Y} (S_f(m,n) - \mu_{S_f})^2$$

$$\sigma_{S_r}S_f = \frac{1}{XY-1} \sum_{m=0}^{X} \sum_{n=0}^{Y} (S_r(m,n) - \mu_{S_r})(S_f(m,n) - \mu_{S_f})$$

**Visual Information Fidelity** [17] (VIF): It is the parameter for evaluating image quality that is based on information aspects of natural scenes statistical (NSS).

$A(p_{i,k}, q_{i,k})$ = actual image details

$A(p_{i,k}, r_{i,k})$ = unbalanced image details

$p_{i,k}, q_{i,k}, and r_{i,k}, k = p, q, and r$ in images frames in a variety of sizes

$M_k$ = scale’s amount of image pixels $k$

$s$ = the variety of categories in which the image is split

**Mutual Information** [17] (MI) Mutual information is used to measure the quality of the fused image, it is given as:

$$MI = I(S_X; S_f) + I(S_Y; S_f)$$

(3.3)

where $(S_X; S_f) = \sum_{m=1}^{M} \sum_{n=1}^{M} h_{R,F}(m,n) \log_2 \left( \frac{h_{R,F}(m,n)}{h_{R}(m)h_{F}(n)} \right)$, where $h_{R}(m), h_{F}(n)$, are $S_R$ and normalised grey level histograms, which correspond to the reference and fused images, respectively. When the $MI$ value of a fused image is higher, it means the image has more features and texture information.

**Entropy** The expected value along with the random variable is measured using entropy [17]. The image’s color histogram can be thought of as a probability density function. The detailed entropy of a picture is represented with $h(i)$ where $h(i)$ reflects the quantized color’s % of pixel $i$ in the overall image.

$$Entropy = \sum_{i=1}^{n} h(i) \log_2 h(i)$$

(3.4)

4. Results. The comparative analysis of different algorithm methods using various metric indexes. The obtained results indicates the proposed algorithm is the best algorithm method among the tested algorithms. How to accurately evaluate the efficiency of fusion methods is an essential topic in picture fusion. In recent years, several other assessment indexes have been proposed. There are two types of assessment methods now in use,
Fig. 4.1: Normal brain MRI, PET medical images fusion results

Fig. 4.2: Normal brain MRI and CT medical images fusion results

Fig. 4.3: Normal brain MRI and SPECT medical images fusion results
Table 4.1: MRI-PET Metric Score [19]

<table>
<thead>
<tr>
<th></th>
<th>VGG-16</th>
<th>VGG-19</th>
<th>SQUEEZE NET</th>
<th>PCA</th>
<th>CNN+PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>38.8189668</td>
<td>43.2626860</td>
<td>52.4530215</td>
<td>38.4114794</td>
<td>37.9574948</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.634070678</td>
<td>0.599964187</td>
<td>0.576918209</td>
<td>0.593477189</td>
<td>0.631366731</td>
</tr>
<tr>
<td>VIF</td>
<td>0.020115451</td>
<td>0.026662346</td>
<td>0.017607955</td>
<td>0.022900916</td>
<td>0.082619463</td>
</tr>
<tr>
<td>Entropy</td>
<td>3.77039411</td>
<td>4.070330749</td>
<td>4.112055998</td>
<td>3.990689894</td>
<td>3.001083163</td>
</tr>
<tr>
<td>MI</td>
<td>0.436537877</td>
<td>0.435251605</td>
<td>0.371655424</td>
<td>0.467361777</td>
<td>1.14531026</td>
</tr>
</tbody>
</table>

Table 4.2: MRI-CT Metric Score

<table>
<thead>
<tr>
<th></th>
<th>VGG-16</th>
<th>VGG-19</th>
<th>SQUEEZE NET</th>
<th>PCA</th>
<th>CNN+PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>52.56273155</td>
<td>46.16454714</td>
<td>53.09027707</td>
<td>22.22134702</td>
<td>20.8591113</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.338170304</td>
<td>0.373294043</td>
<td>0.353296692</td>
<td>0.428951084</td>
<td>0.60058943</td>
</tr>
<tr>
<td>VIF</td>
<td>0.03505622</td>
<td>0.053515519</td>
<td>0.045902515</td>
<td>0.078706307</td>
<td>0.01524352</td>
</tr>
<tr>
<td>Entropy</td>
<td>6.42061688</td>
<td>6.279529215</td>
<td>6.17356001</td>
<td>5.98936538</td>
<td>4.957725674</td>
</tr>
<tr>
<td>MI</td>
<td>0.658167886</td>
<td>0.724539243</td>
<td>0.718558411</td>
<td>0.87839076</td>
<td>1.26598822</td>
</tr>
</tbody>
</table>

Table 4.3: MRI-SPECT Metric Score

<table>
<thead>
<tr>
<th></th>
<th>VGG-16</th>
<th>VGG-19</th>
<th>SQUEEZE NET</th>
<th>PCA</th>
<th>CNN+PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>42.7168407</td>
<td>42.27611387</td>
<td>45.56879791</td>
<td>57.7901058</td>
<td>34.5578519</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.486359572</td>
<td>0.473912872</td>
<td>0.470555961</td>
<td>0.59149644</td>
<td>0.4803565</td>
</tr>
<tr>
<td>VIF</td>
<td>0.047010167</td>
<td>0.044480608</td>
<td>0.037290363</td>
<td>0.107424537</td>
<td>0.04830219</td>
</tr>
<tr>
<td>Entropy</td>
<td>4.72863707</td>
<td>4.815929541</td>
<td>4.693647639</td>
<td>4.38023584</td>
<td>3.20291852</td>
</tr>
<tr>
<td>MI</td>
<td>0.362361713</td>
<td>0.324649206</td>
<td>0.308850748</td>
<td>0.586975232</td>
<td>1.57374802</td>
</tr>
</tbody>
</table>

Table 4.4: MRT1- MRT2 Metric Score

<table>
<thead>
<tr>
<th></th>
<th>VGG-16</th>
<th>VGG-19</th>
<th>SQUEEZE NET</th>
<th>PCA</th>
<th>CNN+PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>51.78371531</td>
<td>54.1766435</td>
<td>30.95919525</td>
<td>47.351909752</td>
<td>26.6870736</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.178852148</td>
<td>0.167316343</td>
<td>0.478195945</td>
<td>0.51509571</td>
<td>0.409438284</td>
</tr>
<tr>
<td>VIF</td>
<td>0.031350142</td>
<td>0.024649206</td>
<td>0.062285785</td>
<td>0.075168443</td>
<td>0.061285748</td>
</tr>
<tr>
<td>Entropy</td>
<td>5.695356778</td>
<td>5.763748050</td>
<td>5.299079636</td>
<td>5.27351219</td>
<td>3.963502677</td>
</tr>
<tr>
<td>MI</td>
<td>0.493450079</td>
<td>0.459140614</td>
<td>0.55102598</td>
<td>0.562701992</td>
<td>1.119491017</td>
</tr>
</tbody>
</table>

the quality evaluation using a source image and a non-source image. Evaluation of image quality, we choose five metrics, for the accurate understanding of their outcomes Mean Square Error (MSE) be a part of the objective evaluation index that must be consulted the Structural Similarity Index (SSIM), Information Entropy (EN), Normalized Mutual Information (Qmi) and visual information fidelity (VIF) pixel, all of which are objective evaluation indicators. a few pairs of radiology medical images taken as the test data. For test data, applied filters for better understanding. The comparing algorithms are based on VGG-16, VGG-19, SQUEEZENET, Principal Component Analysis (PCA), and CNN+PCA. Among them, we randomly choose two source images with diverse modalities. The input image is transformed from RGB to HSI space at the time of testing, and various another format like grayscale, YCbCr etc.

5. Conclusion. The method extracts further information from source images using weights that are not shared. CNN is used with PCA in the proposed method for image reconstruction, and it also aids during
lowering the computational complexity and dimensionality of the original images. MSE loss is the basis for the proposed method loss function. Minimizing MSE loss raises the objective assessment index, however, this does not guarantee acceptable image fidelity. Its focus on the future discussion will be refining the loss function as well as enhancing its effectiveness. The proposed method outperforms other comparison methods in terms of evaluating image fusion metrics. The proposed method is stable in terms of durability and may be applied to image fusion.

REFERENCES


Edited by: Vinoth Kumar
Received: Jun 30, 2022
Accepted: Nov 11, 2022