Abstract—In image fusion tasks, an ideal image decomposition method can bring better performance. MDLatLRR has done a great job in this aspect, but there is still exist some space for improvement. Considering that MDLatLRR focuses solely on the detailed parts (salient features) extracted from input images via latent low-rank representation (LatLRR), the basic parts (principal features) extracted by LatLRR are not fully utilized. Therefore, we introduced an enhanced multi-level decomposition method named dual-decomposed MDLatLRR (D2-LRR) which effectively analyzes and utilizes all image features extracted through LatLRR. Specifically, color images are converted into YUV color space and grayscale images, and the Y-channel and grayscale images are input into the trained parameters of LatLRR to obtain the detailed parts containing four rounds of decomposition and the basic parts. Subsequently, the basic parts are fused using an average strategy, while the detail part is fused using kernel norm operation. The fused image is ultimately transformed back into an RGB image, resulting in the final fusion output.

I. INTRODUCTION

Image fusion is a critical application within the realm of medical imaging. The primary focus of medical image fusion lies in the integration of modalities such as single-photon emission computed tomography (SPECT), magnetic resonance imaging (MRI), computerized tomography (CT), and positron emission tomography (PET) modalities. By combining the high-resolution capabilities of CT scans for bones with MRI scans for soft tissue, the primary data is integrated to offer valuable assistance in clinical diagnosis, treatment planning, and ongoing observation. SPECT and PET can provide insights into the functional and metabolic aspects of the human body, albeit with limited spatial resolution. The fusion of SPECT or PET with MRI images effectively reveals metabolic activity in the thalamus, nucleus, and cortex, along with alterations in blood circulation.

Within the realm of image fusion, the central challenge revolves around extracting characteristics from the input images and applying an adaptive approach to generate the fused images.

It is common knowledge that one of the most typical fusion methods in the transform domain is the multi-scale transform (MST). Fusion methods based on MST typically involve three fundamental stages: decomposition, fusion, and reconstruction. Many MST methods are been proposed, such as wavelet transform [1], pyramid [6], curvelet [8], contourlet [3], and shearlet [4] etc.

Representation learning is also widely used [30]. For medical image fusion, sparse representation (SR) [5] combines nonsampled contourlet transform. Furthermore, Low Rank Representation (LRR) has been used in image fusion [2]. The presented approach introduces an innovative multi-focus image fusion technique that relies on dictionary learning and LRR, achieving improved performance in terms of both global and local structures.

The progress in deep learning has given rise to a multitude of fusion methods that leverage deep learning principles. In 2017, the convolutional neural network was employed to get fusion results by creating a weight map [7]. In 2019, Li et al. [9] introduced an innovative fusion framework utilizing deep learning techniques for the integration of infrared and visible images. In 2022, Zhao et al. [20] combined CNN and Transformer to decompose and fuse image features, achieving certain fusion results, but changing color information of source images. In 2023, Cheng et al. [19] adopted a deep learning method with memory units to generate fused images. While

This work was supported by the National Natural Science Foundation of China (62202205, 62020106012), the Fundamental Research Funds for the Central Universities (JUSRP123030).
deep learning-based techniques yield effective fusion results, they tend to give less consideration to the process of image decomposition. Hence, a fresh approach to multi-level image decomposition known as MDLatLRR was proposed [11], which is based on LatLRR principles.

However, in the MDLatLRR approach, the authors exclusively employed the detail components obtained through LatLRR, omitting the base components. Hence, this paper introduces an enhanced version of MDLatLRR, referred to as D2-LRR, that maximizes the utilization of features acquired through LatLRR.

The primary contributions can be outlined as follows:

1. D2-LRR explores features derived from the decomposition based on the latent low-rank representation and comprehensively utilizes them.
2. The fusion outcome is meticulously analyzed to assess the impact of sparse noise.
3. D2-LRR is utilized for medical image fusion and achieves improved fusion results both objectively and subjectively.

In this paper, the organization of the subsequent sections is as follows: Concise overviews of latent low-rank representation and MDLatLRR are provided in Section III. A detailed presentation of D2-LRR and the fusion method are proposed in Section IV. Section V covers the experimental setup, results, and evaluations. Section VI concludes the paper.

II. RELATED WORK

A. LatLRR: Latent Low-Rank Representation

Latent Low-Rank Representation (LatLRR) [17] was introduced in 2011 for the purpose of feature extraction and subspace segmentation. LatLRR is formulated as the subsequent optimization challenge:

\[
\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda\|E\|_1 \quad \text{s.t., } X = XZ + LX + E \tag{1}
\]

In this context, \(X\) stands for the observed matrix, \(Z\) represents the low-rank coefficient matrix, \(L\) signifies the projection matrix, \(E\) denotes the sparse noise matrix, and \(\|\cdot\|_*\) refers to the nuclear norm, while \(\|\cdot\|_1\) represents the \(\ell_1\)-norm. We refer to \(E\) as the error term. The parameter \(\lambda > 0\) is a parameter.

In the context of visual content [17], the matrix product \(XZ\) encapsulates principal features and base information, whereas \(LX\) emphasizes salient features and detailed information.

For our methodology, dividing the input image into \(M\) patches is our assumption, each patch having a size of \(n \times n\). Consequently, the matrix \(X\) with \((n \times n) \times M\), the matrix \(L\) with \((n \times n) \times (n \times n)\), and the matrix \(Z\) with \(M \times M\). It is not difficult to understand from the perspective of matrix dot product that the size of \(L\) is associated with the image patch dimensions, allowing it to extract salient features from arbitrary-sized input images. However, the size of \(Z\) is related to the size of the input image. In addition, this approach also brings a problem, the noise \(E\) is difficult to separate.

B. MDLatLRR

Utilizing the principles of LatLRR, a groundbreaking multi-level image decomposition technique, MDLatLRR [11], was proposed. This work brilliantly applies LatLRR to image fusion tasks. In Fig. 1, the schematic representation of MDLatLRR is visually represented.

In Fig. 1, the input image \(I\) is decomposed into base parts \((I_b)\) and detailed parts \((I_d)\). Detailed parts are extracted using the projection matrix \(L\), while base parts are acquired by subtracting detailed parts from the input image. This decomposition process is iteratively applied to the base parts. Consequently, after \(n\) iterations, the result includes \(n\) detailed parts and one final base part derived from the input image. Among it, the acquisition of the basic part is too simple and does not make full use of the information that has been decomposed. The reason is that the projection matrix \(Z\) is closely related to the size of the training image, while the image size in the test phase is not fixed. The core improvement of this paper is precisely about this point.

III. METHODOLOGY

A. \(P\): A Low-Rank Projection Matrix

In accordance with [1], due to the close correlation between \(Z\) and the image size, it is not advisable to perform pre-training beforehand. Training for each individual image would inevitably result in a sharp decrease in the algorithm’s execution speed. Therefore, in order to extract principal features from LatLRR, it is necessary to learn a low-rank projection matrix, denoted as \(P\), which captures the principal features of the images. Motivated by sparse representation, the low-rank projection matrix \(P\) is defined by:

\[
\min_P \|P\|_* \quad \text{s.t. } XZ = PX \tag{2}
\]

In equation (2), the recovery result \(PX\) represents base parts (principal features) of \(X\), akin to \(XZ\) in LatLRR. Moreover, \(P\) is associated with the image patch size, similar to the role of \(L\). Consequently, we employ \(P\) to extract base parts, or principal features, from \(X\).

In the end, both \(L\) and \(P\) are utilized to extract detailed and base parts of the input image. In contrast to MDLatLRR, our approach maximizes the utilization of the salient and principal...
features acquired through LatLRR. Thus, our method is termed D2-LRR, and it has been applied to medical image fusion techniques.

Furthermore, we conduct a detailed discussion on the incorporation of the "error" term (E) for fusion. Notably, our experiments reveal that the fusion performance improves when excluding the error term. In Section IV-B, we will delve into the specific experimental results in detail.

B. Dual-decomposed MDLatLRR

We partition the source image (I) into M image patches using a sliding window approach with the n x n window and the step is 1. These patches are then reconstructed into a matrix X with (n x n) x M dimension, where each column represents an image patch. The basic and detailed data of X are calculated in the following manner:

\[ V_b = PX \]  \hspace{1cm} (3)
\[ V_d = LX \]  \hspace{1cm} (4)

where P and L are acquired through learning using (2) and (1), respectively. Here, \( V_d \) represents detail part vectors and \( V_b \) represents base part vectors. Utilizing \( V_d \) and \( V_b \), we reconstruct detail and base images, as demonstrated in Fig. 2.

The reconstruction operator is used to transform each column of \( V_d \) and \( V_b \) into an image patch. Pixels that overlap are handled by applying an averaging technique, which produces both the base image (I_b) and the detail image (I_d). From a subjective point of view, it can be intuitively seen that I_b as the basic part obviously contains more high-frequency information and texture information. The I_d details section contains more outline information and detailed information.

![Fig. 2. The reconstruction graph.](image)

Building upon the decomposition framework involving L and P, basic parts are broken down several times. Consequently, in Fig. 3 the architecture of the dual-decomposed MDLatLRR (D2-LRR) is depicted.

In Fig. 3 I means the input image, n signifies the decomposition level, and \( V_b^i \) and \( V_d^i \) represent the base and detail part vectors of the i-th layer. Hence, through decomposition repeated n times, \( V_b^{1:n} \) and \( V_d^{1:n} \) can be acquired.

C. Acquiring knowledge of the Projection Matrices: L and P

As the medical images for fusion are categorized into four types: CT, MRI, SPECT, and PET, ten images were chosen from each category for training purposes, amounting to a total of 40 training images.

The training approach corresponds to that of MDLatLRR. In order to have enough training samples in the training phase, we divided the existing samples into chunks. We apply sliding window technology with the 16 x 16 window and the step is 16 to process the training dataset, resulting in 5955 image patches. Subsequently, these patches are categorized into two groups: smooth image patches and detail image patches, determined by their Standard Deviation (SD) [10] using equations (5) and (6).

\[ SD(patch) = \sqrt{\frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} (patch(i,j) - \mu)^2} \]  \hspace{1cm} (5)
\[ C(patch) = \begin{cases} \text{detail} & \text{SD(patch)} > \text{Th} \\ \text{smooth} & \text{otherwise} \end{cases} \]  \hspace{1cm} (6)

Here, patch represents the image patch and the threshold value Th is utilized to verify whether an image patch falls into the smooth or detail category. In this paper, Th is set to 0.5.

After classification, our training dataset is divided into 4292 detail image patches and 1663 smooth image patches. Randomly select a certain number of batches as the training set. Subsequently, this finalized trained dataset is employed. It is aim to train the matrices L and P using equations (7) and (2). After training is completed, L and P can be directly used in the testing phase.

D. The Fusion Method for Medical Images Using D2-LRR

After achieving matrices L and P which has been trained, we introduce an innovative approach based on D2-LRR for fusion. General framework of this paper is illustrated in Fig. 4.

In Fig. 4 two sets of features \( V^{1:n}_{d} \) and \( V^{1:n}_{b} \) are obtained, where n signifies the number of decompositions and \( i \in 1, \ldots, k \), k represents the number of input images. Two sets of features have been acquired, \( V^{1:n}_{d} \) and \( V^{1:n}_{b} \). Here, n represents the number of decompositions, while \( i \in 1, \ldots, k \), k denotes the count of input images. We set \( k = 2 \) and \( n \in 1, 2, 3, 4 \). The features \( V^{1:n}_{d} \) are fused using a nuclear-norm-based fusion strategy [11], as outlined in (7). Subsequently, the detail image
is obtained through the reconstruction operator. The features $V_{bi}$ reconstruct the base image, which is then fused using an averaging strategy.

$$V_{d} = \sum_{i=1}^{k} W_j^i \times V_{d_i}^j$$  \hspace{1cm} (7)

$$W_i^j = \frac{u_i^j}{\sum_{i=1}^{k} u_i^j}$$  \hspace{1cm} (8)

$$u_i^j = ||re(V_{d_i}^j)||,$$  \hspace{1cm} (9)

Here, $re(\cdot)$ represents the operator utilized for reconstructing image patches from $V_{d_i}^j$, where $j$ corresponds to the column of $V_{d_i}$. Additionally, $||\cdot||$ stands for the nuclear norm, which accumulate all singular values of the matrix. Once the fused basic images ($f_b$) obtained and detailed images ($f_d^{1:n}$), we calculate the fusion results using (10). This paper performs an accumulation operation on them to obtain fused images.

$$f = f_b + \sum_{i=1}^{n} f_d^i$$  \hspace{1cm} (10)

The source images, including those that appear black and white, are all 3-channel color images. This paper employs the technique outlined in [12] to convert color images from RGB channels into the YUV color space. Next, we employ the proposed fusion technique to merge the grayscale image and the Y channels. Specifically, in the case of the CT and MRI image set, the CT image is converted to grayscale and then fused with the Y channels of the MRI image. Likewise, for the MRI and PET image set, the MRI image is transformed into grayscale and combined with the Y channels of the PET image using the same approach. The same procedure is applied to the MRI and SPECT image set.

IV. Experiments and Assessment

A. Experiment Configuration

During this experimentation, we have three distinct fusion scenarios, namely, the fusion of MRI with CT, MRI with SPECT and MRI with PET [29], as illustrated in Fig. 5.

Next, this paper evaluates this method by comparing it with various fusion methods in recent years, including: Image fusion based on full Convolutional neural network (IFCNN) [13], Unified Fusion Network Based on Similarity Constraints (U2Fusion) [16], Medical Image Fusion Network Using Unsupervised Enhancement (EMFusion) [18], Multimodal Image Fusion Network (MFUFusion) with Memory Units [19]. The concept of feature separation is incorporated into the
Regarding the experimental configuration, 60 images were used for fusion, with 20 images in each group. These images are obtained from the same source as cited in [29]. The fusion algorithms were executed using Matlab R2016a.

B. Ablation Analysis

We will assess how \( E \) affects fusion performance. Therefore, there are two sets of ablation experiments within the D2-LRR framework. The initial approach involves breaking down the image \( I \) into three components (detail part \( D \), base part \( B \), and error \( E \), \( E = I - B - D \)) using the LatLRR method, referred to as D2-LRR_\( B \_D \_E \). The second approach involves using only \( B \) and \( D \) components without including \( E \), known as D2-LRR_\( B \_D \). We employ two techniques for the fusion process and assess the fusion outcomes using specific metrics across three sets of medical image pairs (comprising 20 pairs in each group), as presented in Table II.

It’s evident that the fusion quality of D2-LRR_\( B \_D \) surpasses that of D2-LRR_\( B \_D \_E \). Due to the separation of sparse noise by LatLRR, the remaining features are devoid of artifacts and better suited for the fusion process. Hence, we opt for D2-LRR_\( B \_D \) as the ultimate iteration of our proposed technique. In the subsequent experiments, D2-LRR decomposes the image into a base component and a detail component, while disregarding the error (sparse noise).

C. Baseline

Since the approach builds upon the enhancements made to MDLatLRR, we conduct a comparison between D2-LRR and MDLatLRR using three sets of image pairs (with 20 pairs in

| Images       | Methods          | Qabl | MI | PSNR\(_R_o\) | EN | AG | FMI\(_{pixel}\) | SSIM\(_M_a\) | MS SSIM | SCD |
|--------------|------------------|------|----|-------------|----|----|----------------|-------------|----------|-----|
| CT_MRI       | MDLatLRR_\( level\_1 \) | 0.516 | 8.853 | 15.722 | 4.427 | 0.084 | 0.869 | **0.753** | 0.882 | 1.036 |
|              | MDLatLRR_\( level\_2 \) | 0.565 | 8.664 | 15.457 | 4.332 | 0.122 | **0.870** | 0.752 | 0.945 | 1.161 |
|              | MDLatLRR_\( level\_3 \) | 0.546 | 8.261 | 14.672 | 4.131 | 0.153 | 0.866 | 0.729 | 0.949 | 1.257 |
|              | D2-LRR_\( level\_1 \) | 0.463 | 7.891 | 13.701 | 3.795 | 0.171 | 0.854 | 0.702 | 0.920 | 1.290 |
|              | D2-LRR_\( level\_2 \) | 0.564 | 8.666 | 15.459 | 4.333 | 0.121 | 0.869 | 0.752 | 0.882 | 1.035 |
|              | D2-LRR_\( level\_3 \) | 0.546 | 8.266 | 14.675 | 4.133 | 0.152 | 0.865 | 0.729 | **0.950** | 1.257 |
|              | D2-LRR_\( level\_4 \) | 0.463 | 7.891 | 13.701 | 3.795 | 0.171 | 0.854 | 0.702 | 0.921 | 1.291 |
| MRI_PET      | MDLatLRR_\( level\_1 \) | 0.561 | 7.010 | **19.829** | 3.550 | 0.067 | 0.879 | 0.710 | 0.937 | 1.101 |
|              | MDLatLRR_\( level\_2 \) | **0.645** | 6.894 | 19.409 | 3.447 | 0.106 | 0.872 | **0.803** | 0.965 | 1.365 |
|              | MDLatLRR_\( level\_3 \) | 0.577 | 6.119 | 17.545 | 3.309 | 0.140 | 0.862 | 0.783 | 0.941 | 1.410 |
|              | MDLatLRR_\( level\_4 \) | 0.646 | 6.309 | 15.819 | 3.155 | **0.166** | 0.849 | 0.758 | 0.906 | 1.400 |
|              | D2-LRR_\( level\_1 \) | 0.560 | **7.102** | 19.828 | **3.551** | 0.067 | **0.880** | 0.800 | 0.937 | 1.100 |
|              | D2-LRR_\( level\_2 \) | 0.644 | 6.895 | 19.410 | 3.449 | 0.104 | 0.872 | 0.802 | **0.966** | 1.364 |
|              | D2-LRR_\( level\_3 \) | 0.577 | 6.622 | 17.549 | 3.311 | 0.139 | 0.862 | 0.873 | 0.941 | **1.411** |
|              | D2-LRR_\( level\_4 \) | 0.464 | 6.313 | 15.823 | 3.156 | 0.164 | 0.849 | 0.758 | 0.906 | 1.402 |
| MRI_SPECT    | MDLatLRR_\( level\_1 \) | 0.612 | **8.595** | 19.987 | **4.297** | 0.051 | 0.874 | 0.752 | 0.937 | 0.888 |
|              | MDLatLRR_\( level\_2 \) | **0.687** | 8.497 | 19.661 | 4.249 | 0.082 | 0.864 | 0.753 | 0.968 | 1.013 |
|              | MDLatLRR_\( level\_3 \) | 0.573 | 8.442 | 17.767 | 4.221 | 0.119 | 0.853 | 0.723 | 0.924 | 1.093 |
|              | MDLatLRR_\( level\_4 \) | 0.405 | 8.212 | 15.516 | 4.106 | **0.150** | 0.839 | 0.687 | 0.861 | 1.116 |
|              | D2-LRR_\( level\_1 \) | 0.611 | 8.594 | **19.988** | 4.297 | 0.051 | **0.875** | 0.752 | 0.937 | 0.888 |
|              | D2-LRR_\( level\_2 \) | 0.686 | 8.497 | 19.663 | 4.249 | 0.082 | 0.864 | **0.754** | **0.969** | 1.013 |
|              | D2-LRR_\( level\_3 \) | 0.574 | 8.443 | 17.771 | 4.221 | 0.118 | 0.853 | 0.724 | 0.924 | 1.093 |
|              | D2-LRR_\( level\_4 \) | 0.406 | 8.214 | 15.520 | 4.107 | 0.149 | 0.839 | 0.688 | 0.861 | **1.117** |

Fig. 5. There are three sets: the first row comprises CT, PET, and SPECT, while the second consists of MRI images.
each group). In addition, certain indicators were selected for evaluating, as illustrated in Table II.

![Fig. 6. CT and MRI fusion results.](image1)

![Fig. 7. MRI and PET fusion results.](image2)

It's evident that D2-LRR outperforms MDLatLRR in terms of numerical superiority across most of the evaluation criteria. This demonstrates that D2-LRR more comprehensively and efficiently makes use of the image information acquired through LatLRR while eliminating sparse noise. In addition, it can be seen from the indicators that the results of this paper have higher structural similarity than the source images and better retain the structural information of the source images. Moving forward, we proceed with the comparison of the fusion technique using D2-LRR against other fusion methods. We exclusively get the fusion results following four decompositions in MDLatLRR as the benchmark for comparison.

**D. Assessment from Subjective and Objective Perspectives**

This paper selects a pair of images from each series of experiments as examples and analyzes their visual effects as depicted in Fig. 6 - Fig. 8.

![Fig. 8. MRI and SPECT fusion results.](image3)

As evident from the three tables above, where red bold indicates the maximum values and blue highlights the second-highest, it is easy to find that our work have more number one objective metrics, which signify that our method retains a greater amount of prominent details from source images, ultimately improving visual perception. The second-highest value (VIF, AG) suggests that the fusion outcomes exhibit noise that impacts visual perception. Furthermore, both EMFusion and CDDFuse have altered original color in their results. The results of TL-SR, MDLatLRR and D2-LRR are more consistent with the perception of the human eye. Moreover, we assess all methods from an objective standpoint, and the findings are presented in Table III - Table V.

**TABLE III**

| CT_MRI | VIF   | AG    | SD | FMU_dct | FMU_w | MS_SSIM |
|--------|-------|-------|----|---------|-------|---------|
| IFCNN  | 0.6003| 8.3779| 76.6345 | 0.3779 | 0.4171 | 0.9424  |
| U2Fusion | 0.3148| 4.6460| 57.6950 | 0.3095 | 0.2420 | 0.8511  |
| EMFusion | 0.4782| 5.6179| 70.8714 | 0.3049 | 0.3483 | 0.8914  |
| MUFusion | 0.4897| 6.4662| 68.1673 | 0.3255 | 0.3111 | 0.8523  |
| CDDFuse | 0.6581| 8.9587| 83.1489 | 0.3394 | 0.3977 | 0.9286  |
| TL-SR  | 0.6809| 7.5238| 86.4074 | 0.4112 | 0.4561 | 0.9402  |
| MDLatLRR | 1.0571| 10.5751| 87.2797 | 0.3211 | 0.3989 | 0.9371  |
| D2-LRR_level1 | 0.3506| 5.0780| 61.1988 | 0.4243 | 0.4611 | 0.8815  |
| D2-LRR_level2 | 0.5695| 7.2384| 70.0968 | 0.4044 | 0.4454 | 0.9448  |
| D2-LRR_level3 | 0.8379| 9.2837| 80.9316 | 0.3646 | 0.4242 | 0.9490  |
| D2-LRR_level4 | 1.0364| 10.4513| 88.7297 | 0.3216 | 0.4047 | 0.9205  |

As evident from the three tables above, where red bold indicates the maximum values and blue highlights the second-highest, it is easy to find that our work have more number one objective metrics, which signify that our method retains a greater amount of prominent details from source images, ultimately improving visual perception. The second-highest value (VIF, AG) suggests that the fusion outcomes exhibit reduced noise, resulting in a more natural and clear appearance.
TABLE IV  
THE MEAN QUALITY METRICS VALUES FOR 20 MRI AND PET FUSED IMAGES.

| Image Set | VIF | AG | SD | FMI_dct | FMI_w | MS_SSIM |
|-----------|-----|----|----|---------|-------|---------|
| IFCNN     | 0.8816 | 6.1111 | 52.0373 | 0.3711 | 0.3987 | 0.9072 |
| U2Fusion  | 0.1078 | 1.0746 | 38.7405 | 0.2777 | 0.2158 | 0.8761 |
| EMFusion  | 0.7651 | 5.6699 | 50.9259 | 0.3651 | 0.3145 | 0.9032 |
| MUFusion  | 0.9277 | 3.6009 | 58.7297 | 0.2627 | 0.3212 | 0.9279 |
| CDDFuse   | 0.9723 | 6.0223 | 60.2942 | 0.4264 | 0.4040 | 0.9541 |
| TL-SR     | 0.8140 | 5.8393 | 54.6295 | 0.2437 | 0.1751 | 0.9224 |
| MDLatLRR  | 2.3711 | 10.5380 | 68.2742 | 0.2990 | 0.2690 | 0.9531 |
| D2LRR_level1 | 1.0083 | 5.9393 | 59.4244 | 0.3671 | 0.1608 | 0.9681 |
| D2LRR_level2 | 1.6816 | 8.5310 | 59.4864 | 0.3258 | 0.3837 | 0.9237 |
| D2LRR_level4 | 2.3331 | 10.6559 | 69.6247 | 0.2864 | 0.3625 | 0.8806 |

TABLE V  
THE MEAN QUALITY METRICS VALUES FOR 20 MRI AND SPECT FUSED IMAGES.

| Image Set | VIF | AG | SD | FMI_dct | FMI_w | MS_SSIM |
|-----------|-----|----|----|---------|-------|---------|
| IFCNN     | 0.8816 | 6.1111 | 52.0373 | 0.3711 | 0.3987 | 0.9072 |
| U2Fusion  | 0.1078 | 1.0746 | 38.7405 | 0.2777 | 0.2158 | 0.8761 |
| EMFusion  | 0.7651 | 5.6699 | 50.9259 | 0.3651 | 0.3145 | 0.9032 |
| MUFusion  | 0.9277 | 3.6009 | 58.7297 | 0.2627 | 0.3212 | 0.9279 |
| CDDFuse   | 0.9723 | 6.0223 | 60.2942 | 0.4264 | 0.4040 | 0.9541 |
| TL-SR     | 0.8140 | 5.8393 | 54.6295 | 0.2437 | 0.1751 | 0.9224 |
| MDLatLRR  | 2.3711 | 10.5380 | 68.2742 | 0.2990 | 0.2690 | 0.9531 |
| D2LRR_level1 | 1.0083 | 5.9393 | 59.4244 | 0.3671 | 0.1608 | 0.9681 |
| D2LRR_level2 | 1.6816 | 8.5310 | 59.4864 | 0.3258 | 0.3837 | 0.9237 |
| D2LRR_level4 | 2.3331 | 10.6559 | 69.6247 | 0.2864 | 0.3625 | 0.8806 |

V. Conclusion

Improved decomposition framework has been proposed, named D2-LRR, built upon MDLatLRR. D2-LRR primarily leverages image features obtained by LatLRR, both from the base level and the detail level, effectively. Additionally, we incorporate D2-LRR into the fusion process, exploring the impact of sparse noise on the fusion task. As a result, the proposed fusion method attains best performance. Like other methods, our method is also limited by already registered images. How to register and merge is our future direction. We will attempt to combine potential rank representations with deep neural networks to further improve performance by utilizing the learning capabilities of the network.

REFERENCES

[1] Aghamaleki, J.A., Ghorbani, A. Image fusion using dual tree discrete wavelet transform and weights optimization. Vis Comput 39, 1181–1191 (2023).
[2] Li, H., Wu X.J. Multi-focus image fusion using dictionary learning with deep neural networks to further improve performance by leveraging the learning capabilities of the network.

[6] J. Yao, Y. Zhao, Y. Bu, S. G. Kong and J. C. -W. Chan, “Laplacian Pyramid Fusion Network With Hierarchical Guidance for Infrared and Visible Image Fusion,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 9, pp. 4630-4644, Sept. 2023.
[7] Liu Y, Chen X, Cheng J, et al. A medical image fusion method based on convolutional neural networks[C]//2017 20th International Conference on Information Fusion (Fusion). IEEE, 2017: 1-7.
[8] Nagaraja Kumar N, Jayachandra Prasad T, Prasad K S. An intelligent multimodal medical image fusion model based on improved fast discrete curvelet transform and Type-2 fuzzy entropy[J]. International Journal of Fuzzy Systems, 2023, 25(1): 96-117.
[9] Li H, Wu X J. DenseFuse: A Fusion Approach to Infrared and Visible Images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.
[10] Rao Y J. “In-fibre Bragg grating sensors.” Measurement science and technology 8.4 (1997): 555.
[11] Li H, Wu X J, Kittler J. MDLatLRR: A Novel Decomposition Method for Infrared and Visible Image Fusion[J]. IEEE Transactions on Image Processing, 2020-4733-4746.
[12] Yin M, Liu X, Liu Y, et al. Medical Image Fusion With Parameter-Adaptive Pulse Coupled Neural Network in Nonsubsampled Shearlet Transform Domain[J]. IEEE Transactions on Instrumentation and Measurement, 2018 (99): 1-16.
[13] Zhang Y Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion,2020,54(C).
[14] Li H, Wu X J, Kittler J. Infrared and Visible Image Fusion using a Deep Learning Framework[C]//2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 2705-2710.
[15] Xiaosong Li, Fuqiang Zhou, Haishu Tan, et al. Joint image fusion and denoising via three-layer decomposition and sparse representation[J]. Knowledge-Based Systems, 2021, 107087, ISSN 0950-7051.
[16] Han X.Jiayi M,Junjun J, et al. U2Fusion: A Unified Unsupervised Image Fusion Network[J]. IEEE transactions on pattern analysis and machine intelligence,2020,PP.
[17] Liu G, Yan S. Latent low-rank representation for subspace segmentation and feature extraction[C]//2011 international conference on computer vision. IEEE, 2011: 1615-1622.
[18] Xiaosong L,Fuqiang Z,Haishu T, et al. Multimodal medical image fusion based on joint bilateral filter and local gradient energy[J]. Information Sciences,2021,569.
[19] Chunyang C,Tianyang X,Xiao-Jun W. MUFusion: A general unsupervised image fusion network based on memory unit[J]. Information Fusion,2023,92.
[20] Zhao, Zixiang et al. “CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion.” ArXiv abs/2211.14461 (2022): n. pag.
[21] Qu G, Zhang D, Yan P. Information measure for performance of image fusion[J]. Electronics letters, 2002, 38(7): 313-315.
[22] Xydeas C S, Petrovic V. Objective image fusion performance measure[J]. Electronics letters, 2000, 36(4): 308-309.
[23] Haghhighat M, Razian M A. Fast-FMI: non-reference image fusion metric[C]//2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2014: 1-3.
[24] Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity, Inf. Fusion 14 (2) (2013) 127–135.
[25] Ma K , Zeng K , Wang Z . Perceptual Quality Assessment for Multi-Exposure Image Fusion[J]. Image Processing, IEEE Transactions on, 2015, 24(11):3345-3356.
[26] Aslantas V, Bendes E. A new image quality metric for image fusion: the sum of the correlations of differences[J]. Aeu-international Journal of electronics and communications, 2015, 69(12): 1890-1896.
[27] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
[28] Dwyer S P. The mean and standard deviation of the distribution of group assembly sums[J]. Psychometrika,1964,29(4).
[29] http://www.esience.cn/people/xiaomi/index.html.
[30] Li H, Xu T, Wu X J, et al. LRRNet: A Novel Representation Learning Guided Fusion Network for Infrared and Visible Images[J]. IEEE transactions on pattern analysis and machine intelligence,2023.