Digital Engineering and Digital Twin

Tech Science Press

Doi:10.32604/dedt.2025.066344

ARTICLE

A Systematic Comparison of Discrete Cosine Transform-Based Approaches for Multi-Focus Image Fusion

Muhammad Osama[1], Sarwar Shah Khan[2,*], Sajid Khan[2], Muzammil Khan[3], Mian Muhammad Danyal[4] and Reshma Khan[1]

[1]Department of Computer Science, Iqra National University Swat Campus, Swat, 19200, Pakistan
[2]Department of Computer Science, University of Engineering and Technology, Mardan, 23200, Pakistan
[3]Department of Computer Science, University of Swat, Swat, 19200, Pakistan
[4]Center for Excellence in Information Technology, Institute of Management Sciences, Peshawar, 25000, Pakistan

*Corresponding Author: Sarwar Shah Khan. Email: sskha0092@gmail.com

Received: 06 April 2025; Accepted: 18 July 2025; Published: 19 August 2025

ABSTRACT: Image fusion is a technique used to combine essential information from two or more source images into a single, more informative output image. The resulting fused image contains more meaningful details than any individual source image. This study focuses on multi-focus image fusion, a crucial area in image processing. Due to the limited depth of field of optical lenses, it is often challenging to capture an image where all areas are in focus simultaneously. As a result, multi-focus image fusion plays a key role in integrating and extracting the necessary details from different focal regions. This research presents a comparative analysis of various Discrete Cosine Transform (DCT)-based methods for multi-focus image fusion. The primary objective is to provide a clear understanding of how these techniques differ based on mathematical formulations and to compare their visual and statistical performance. The analysed methods include: DCT + Variance and DCT + Variance + Consistency Verification (CV), DCT + Correlation Coefficient (CC) and DCT + CC + CV, DCT + Singular Value Decomposition (SVD) and DCT + SVD + CV, DCT + Sharpening and DCT + Sharpening + CV, DCT + Correlation Energy (Corr_Eng) and DCT + Corr_Eng + CV, Through experimental evaluations, the study finds that DCT + Variance and DCT + Variance + CV consistently deliver superior results across all tested image sets. The performance of these methods is evaluated using six different quantitative metrics, demonstrating their effectiveness in enhancing image quality through fusion.

KEYWORDS: Multi-focused image fusion; DCT based methods; variance; correlation coefficient; singular value decomposition

1 Introduction

In digital image processing, image fusion has emerged as a crucial research field, particularly in applications requiring high-quality visual data. The fundamental objective of image fusion is to integrate complementary information from multiple images into a single, more informative output. Among the various types of image fusion, multi-focus image fusion plays a significant role in scenarios where optical limitations prevent capturing an entirely focused image in a single shot. Due to the limited depth of field of optical lenses, objects at different distances from the camera may not appear sharp simultaneously. This results in some regions of the image being well-focused while others remain blurred. To solve this problem, multi-focus image fusion methods fuse several images captured at various focus levels to produce an image where all the details of interest are focused [1].

Multi-focus image fusion methods can generally be divided into spatial domain and frequency domain methods. Spatial domain methods are based on image pixels directly and encompass methods like the simple averaging method [2], principal component analysis (PCA) [3], and intensity-hue-saturation (IHS) [3] transformations. While these methods are computationally straightforward, they often suffer from spectral distortions and reduced image contrast, which limits their effectiveness [4,5]. On the other hand, frequency domain methods transform images into a frequency representation before fusion, offering advantages such as improved robustness and better feature preservation. Popular techniques in this domain include Discrete Wavelet Transform (DWT) [6], Stationary Wavelet Transform (SWT) [7], and Discrete Cosine Transform (DCT) [8]. Wavelet-based approaches like DWT offer multi-resolution analysis but cause problems like shift variance and ringing effects. To avoid these, SWT was developed as a shift-invariant alternative, with better localization and the capability to retain finer details better. DCT-based fusion approaches, frequently found in image compression standards like JPEG, enable direct fusion within the compressed domain, cutting down on computation overhead and enhancing efficiency [5].

The selection of an appropriate fusion technique relies on the particular application demands, i.e., real-time processing capability, computational efficiency, and noise robustness. Multi-focus image fusion remains a crucial research area, affecting a number of applications ranging from medical imaging [9], remote sensing, surveillance, computer vision [10], and robotics [11]. As research continues, new fusion methods are being investigated to obtain better image clarity, and multi-focus image fusion has become a critical tool in contemporary imaging applications.

This study focuses on methods based on the Discrete Cosine Transform (DCT) domain and explores various DCT-based approaches in detail. By thoroughly analyzing these variations, we aim to gain deeper insights into their functionality and effectiveness. This research also compares different DCT-based methods and presents a better understanding for readers to the point in brief. The techniques considered are DCT with Variance, DCT with Variance and Consistency Verification (CV), DCT with Correlation Coefficient, and DCT with Correlation Coefficient + CV. Next, we consider DCT with Singular Value Decomposition (SVD), DCT with SVD and CV, DCT with Sharpening, and its extended version with CV. Finally, the study also considers DCT with Correlation Energy (Corr_Eng) and DCT with Corr_Eng + CV. Through this comparative analysis, we intend to offer an overall picture of the effectiveness and dependability of various DCT-based fusion methods.

The remaining part of this paper is organized as follows: Section 2 presents an elaboration on Discrete Cosine Transform (DCT)-based Fusion Techniques. Section 3 describes the merit and de-merit of different DCT-based techniques. Section 4 discusses the performance metrics, experimental results, and results and analysis. Lastly, Section 5 presents the key findings and conclusions of the research.

2 Discrete Cosine Transform (DCT)-Based Fusion Approaches

In multi-focus image fusion, a variety of methods have been established in the spatial and frequency domains. However, research indicates that frequency-domain methods generally have more merits than spatial-domain methods, especially in maintaining critical image details and enhancing overall quality [5]. This research mainly investigates Discrete Cosine Transform (DCT)-based fusion schemes and seeks the best method to use by considering their performance using various image data sets. It takes into consideration both qualitative elements, including clarity of vision, and quantitative experimental results to define the best algorithm.

2.1 Discrete Cosine Transforms (DCT)

Discrete Cosine Transform (DCT) is a method that transforms an image from the spatial to the frequency domain. The process is especially suited to capturing fundamental image details like edges and textures by examining pixel frequency data. DCT is generally favored for frequency-based processing because it is efficient since it only depends on cosine functions for transformation and is therefore both computationally light and quick.

To recreate the image in its transformed form, the Inverse Discrete Cosine Transform (IDCT) is used, which returns the frequency components to their pixel values [12]. In short, DCT expresses a limited number of data points as a sum of cosine functions vibrating at different frequencies. Due to its efficiency, DCT is often used with other methods to improve the quality of multi-focus image fusion. The performance of DCT-based approaches is assessed as follows.

Two-dimensional Discrete Cosine Transform (DCT) is performed on an $N \times N$ block of an image, in the case of 8×8 images, to transform spatial data into frequency components. In the same vein, Inverse Discrete Cosine Transform (IDCT) is performed to revert to the original image from its frequency domain representation [13]. These two operations are mathematically expressed in Eqs. (1) and (3), respectively.

$$d(k,l) = \frac{2a(k)\,a(l)}{N} \times \sum_{m=0}^{N-1}\sum_{n=0}^{M-1} x(m,n)\left[\frac{(2m+1)\pi k}{2n}\right] \times \cos\left[\frac{(2n+1)\pi k}{2n}\right] \tag{1}$$

where, $k,\ l = 0, 1, \ldots,\ N-1$ and

$$a(k) = \begin{cases} \dfrac{1}{\sqrt{2}}, & \text{if } k = 0 \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

$$x(m,n) = \sum_{k=0}^{N-1}\sum_{j=0}^{M-1} \frac{2a(k)\,a(l)}{N} \times d(k,l) \times \cos\left[\frac{(2n+1)\,\pi k}{2n}\right] \tag{3}$$

where $m, n = 0, 1, \ldots, N-1$.

In Eq. (1), $d(0,\ 0)$ is the DC coefficient, which is the component that has zero frequency in both dimensions. The rest of the coefficients in the block are known as AC coefficients, which are the components of non-zero frequencies in either or both dimensions.

2.2 Consistency Verification

To improve the quality of the fused image and minimize errors due to improper block selection, Consistency Verification (CV) is used as a post-processing operation at the last stage of the image fusion process. The method presumes that if a middle block in the merged image is from image B, but most of its adjacent blocks are from image A, then the middle block probably belongs to image A, too. To guarantee consistency, Li et al. suggested employing a majority filter for the CV process [13]. This filter works based on the decision map M(x, y), substituting the central block with the corresponding block of image A when most surrounding blocks are from image A. By incorporating CV as a final step, the overall accuracy of the fused image is improved, reducing inconsistencies in block selection as illustrated in Fig. 1. This CV technique is widely used in DCT-based fusion methods, including those discussed in this study [14].

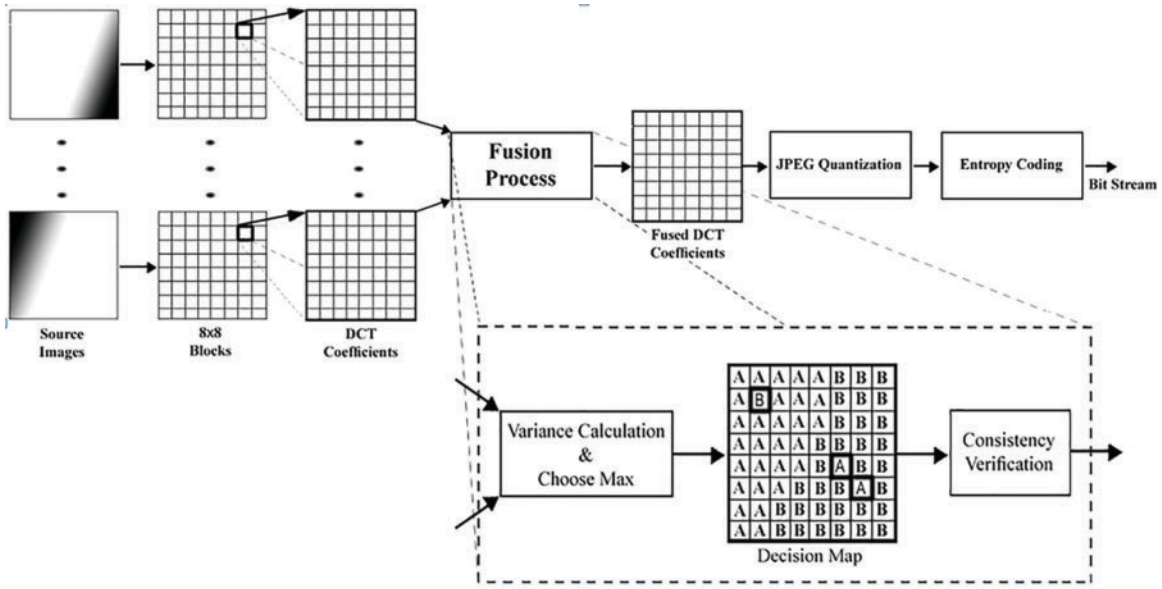

Figure 1: The architecture of a JPEG encoder merging with an image fusion approach

2.3 DCT with Variance and DCT with Variance + CV

In multi-focus images, the regions in focus contain additional information as they preserve finer details. This detailed information is directly related to high variance, which serves as an important indicator of focus. In image processing applications, variance is commonly used as a contrast criterion to differentiate between focused and blurred regions. The computation of variance in the DCT domain is straightforward, as explained in the previous section. Therefore, variance values are utilized to determine the activity level of 8 × 8 blocks within the source images. The fusion process begins by dividing the source images into 8 × 8 blocks and computing the DCT coefficients for each block. As illustrated in Fig. 1, the variance of corresponding blocks from both source images is computed using Eq. (9) as the activity measure. The block with the higher activity measure is selected as the most relevant for the resultant image. Consequently, the DCT representation of the resultant image is constructed using blocks with greater variance, which originate from either source image A or B [15].

The transform coefficients $d'(k,l)$ are normalized as follows:

$$d'(k,l) = \frac{d(k,l)}{N} \tag{4}$$

Compute the values of mean, μ, and variance, σ^2, of an $N \times N$ block in the spatial domain are as follows:

$$\mu = \frac{1}{N^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n) \tag{5}$$

$$\sigma^2 = \frac{1}{N^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n) - \mu^2 \tag{6}$$

As we already know, the mean value, μ, of the $N \times N$ block is given by $d(0,0)$. For providing the variance of the DCT coefficient for each block, a mathematical evaluation using (3). We have:

$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x^2(m,n) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x(m,n).x(m,n)$$
$$= \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x(m,n) \times$$
$$\left(\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{2a(k)\,a(l)}{N} \times d(k,l) \times \cos\left[\frac{(2m+1)\pi k}{2N}\right] \times \cos\left[\frac{(2n+1)\pi k}{2N}\right]\right) \tag{7}$$

By exchanging the places of the sums, we will have:

$$= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{2a(k)\,a(l)}{N} \times d(k,l) \times$$
$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x(m,n) \times \cos\left[\frac{(2m+1)\pi k}{2N}\right] \times \cos\left[\frac{(2n+1)\pi k}{2N}\right]$$
$$= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{2a(k)\,a(l)}{N} \times d(k,l) \times \frac{N \times d(k,l)}{2a(k)\,a(l)}$$
$$= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} d(k,l) \times d(k,l)$$
$$\Rightarrow \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x^2(m,n) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} d^2(k,l) \tag{8}$$

Then, from (7) and (8), we have:

$$\sigma^2 = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x^2(m,n) - \mu^2$$
$$\Rightarrow \sigma^2 = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{d^2(k,l)}{N^2} - d^2(0,0) \tag{9}$$

Finally, the variance of an $N \times N$ block of pixels can be correctly computed from its DCT coefficients by the sum of the squared normalized Eq. (4) AC coefficient of the DCT block [15]. Once the variance-based image fusion is completed in the DCT domain, Consistency Verification (CV) is employed as a post-processing step to further refine the fused image, ensuring improved accuracy and visual quality.

2.4 DCT with Correlation Coefficient and DCT with Correlation Coefficient + CV

The correlation coefficient (CC) is the statistical representation that examines the level of linear relationship among two variables. The correlation among the two $N \times N$ blocks of images A and B is defined as (10):

$$Corr(A,B) = \frac{\sum_{N-1}^{M-1}\sum_{N-1}^{M-1}(A(m,n)-a')(B(m,n)-b')}{\sqrt{\sum_{N-1}^{M-1}\sum_{N-1}^{M-1}(A(m,n)-a')}\sqrt{\sum_{N-1}^{M-1}\sum_{N-1}^{M-1}(B(m,n)-b')}} \tag{10}$$

whereas $A(m,n)$ is the intensity of the $(m,n)^{\text{th}}$ pixel in image A, $B(m,n)$ is the intensity of the $(m,n)^{\text{th}}$ pixel in image B, $\overline{b}$ is the mean intensity value of image A, and $\overline{b}$ is the mean intensity value of image B.

To derive the correlation of the $N \times N$ block of image A and B in the DCT domain, the $P_A(i,j)$ and $P_B(i,j)$ are defined as follows:

$$P_A(i,j) = d_A(i,j) - \overline{d_A} \tag{11}$$

$$P_B(i,j) = d_B(i,j) - \overline{d_B} \tag{12}$$

whereas $P_A(i,j)$ and $P_B(i,j)$ are the DCT coefficients of the $N \times N$ block of image A and B, respectively. Also $\overline{d}_A$ and $\overline{d}_B$ are mean values of the DCT coefficient of the $N \times N$ block of images A and B, respectively.

Then, the correlation coefficient among two $N \times N$ blocks of images A and B in the DCT domain is given as:

$$Corr_{DCT}(A,B) = \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} P_A(i,j)P_B(i,j)}{2\sqrt{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} P_A(i,j)}\; 2\sqrt{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} P_B(i,j)}} \tag{13}$$

hence, the correlation coefficient of two $N \times N$ blocks of images A and B can be obtained just from the DCT coefficients mathematically [16].

In multi-focus image fusion, the DCT-based correlation coefficient method follows a structured process. Initially, the input images are filtered using a low-pass filter, producing artificially blurred versions of the original images. These blurred images, along with their corresponding input images, are then divided into 8×8 blocks, and the DCT coefficients for each block are computed. Each block from the first input image and its corresponding blurred block, as well as each block from the second input image and its respective blurred block, are analyzed in this process. The difference between a sharp image and its blurred version becomes evident after low-pass filtering. Notably, blocks from the focused regions undergo greater changes when passed through the low-pass filter, whereas blocks from unfocused (blurred) regions remain relatively unchanged.

As a result, the correlation coefficient between a focused image block before and after filtering is lower than that of a blurred region, which remains largely similar. Consequently, blocks exhibiting greater variation after filtering—those with lower correlation coefficient values—are identified as belonging to the focused areas and are selected for the fused image. This correlation coefficient serves as a key decision-making criterion in constructing the decision map for the image fusion process [16].

$$\text{Decision}_{\text{Map}} = \left\{ \begin{array}{r} -1 \text{ if } \text{Corr}_{\text{DCT}}\left(\text{im}A, \overline{\text{im}A}\right) < \text{Corr}_{\text{DCT}}\left(\text{im}B, \overline{\text{im}B}\right) \\ +1 \text{ if } \text{Corr}_{\text{DCT}}\left(\text{im}A, \overline{\text{im}A}\right) > \text{Corr}_{\text{DCT}}\left(\text{im}B, \overline{\text{im}B}\right) \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{array} \right\} \tag{14}$$

The block selection for the fused output image (DCT + Corr) is defined as (10):

$$\text{Output Image(DCT + Corr)} = \left\{ \begin{array}{ccc} \text{im}A & \text{if} & \text{Decision}_{\text{map}} = -1 \\ \text{im}B & \text{if} & \text{Decision}_{\text{map}} = +1 \\ \dfrac{\text{im}A + \text{im}B}{2} & \text{if} & \text{Decision}_{\text{map}} = 0 \end{array} \right\} \tag{15}$$

Once the Correlation coefficient-based image fusion is completed in the DCT domain, Consistency Verification (CV) is employed as a post-processing step to further refine the fused image, ensuring improved accuracy and visual quality.

2.5 DCT with Singular Value Decomposition (SVD) and DCT with SVD + CV

Two-dimensional DCT (using vector processing) transform of $N \times N$ blocks of the image $a(m, n)$ are represented as Eq. (16):

$$A = C \cdot a \cdot C^{T} \tag{16}$$

where the orthogonal matrix C consists of coefficients of the discrete cosine transform and superscripts T stands for the transpose operator. On the other hand, A represents the DCT coefficients for the image's matrix of *a*. Furthermore, it is known as a Hermitian matrix, i.e.,

$$C^{-1} = C^{T} \tag{17}$$

The DCT inverse of A is therefore defined as Eq. (18).

$$a = C^{T} \cdot A \cdot C \tag{18}$$

The SVD Matrix applies to the spatial domain. But here transforming the discussion into the DCT domain and comparing the SVD matrix, it is concluded that the matrix $\sum$ has the exact data in both spatial and DCT domains, so

$$b_{8\times 8} = U \sum\nolimits_{8\times 8} V^{T} (\text{Spatial Domain}) \tag{19}$$

$$B_{8\times 8} = \acute{U} \sum\nolimits_{8\times 8} \dot{v}^{T} (\text{DCT Domain}) \tag{20}$$

All parameters are different, however, $\sum$ and $\sum'$ are identical.

In the equations shown above, it has been proven that $\sum$(DCT domain) and $\sum$ Spatial Domain matrices are exactly equal. To put it in other words, eigen values of bb^{T} and BB^{T} are the same [13].

By calculating SVD and eigenvalues of BB^{T} for 8 × 8 blocks in DCT, singular values $\sigma_i(\sigma_1, \ldots, \sigma_8)$ are extracted for each block. The lower indexes are for low frequencies and the higher indexes are for high frequencies. Singular values σ_6, σ_7 and σ_8 have smaller values, and is almost zero, which block noise has a great impact on changing their values. Therefore, to keep choosing unsuitable blocks, it is better to use $\sigma_1, \ldots, \sigma_5$. The geometric mean of 5 largest singular values is selected as a standard for the focused blocks:

$$SVD^{5}_{DCT} = \sqrt{\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5} \tag{21}$$

Accordingly, the block with the most value of SVD^{5}_{DCT} is selected as the focused block. The output resultant image is constructed by the decision map (M) of selected suitable focused blocks. The decision map of two multi-focused images (im$A$ and im$B$) is as follows:

$$M(i, j) = \left\{ \begin{array}{cc} 1 \text{ if } SVD^{5}_{DCT}(\text{im}A) > SVD^{5}_{DCT}(\text{im}B) \\ -1 \qquad\qquad \text{if else} \end{array} \right\} \tag{22}$$

where i and $j = 1, 2, \ldots, \dfrac{\text{size of the input image}}{8}$

Accordingly, for the $(i, j)^{\text{th}}$ block of the output image, if $M(i, j) = 1$, im$A$ is used and if $M(i, j) = -1$, im$B$ is used [13]. Once the SVD based image fusion is completed in the DCT domain, Consistency Verification (CV) is employed as a post-processing step to further refine the fused image, ensuring improved accuracy and visual quality.

2.6 DCT with Sharp and DCT with Sharp + CV

The focused image is highly informative and has high contrast. The image quality and contrast are specified by variance, and the variance of an image is used as a criterion. Thus, the block of the focused area in image h has a higher variance value than the block of the unfocused area. But sometimes the variance criterion finds error in selecting some similar blocks. To avoid selecting the unsuitable block for making the resultant image, and because of the small difference in variance among some corresponding blocks of input multi-focus images, the introduced method increases and enhances the amount of difference among the variances by sharpening the input images [17].

The sharpened input image is obtained using the unsharp mask, as shown in (23).

$$\text{unsharp mask} = \begin{pmatrix} -0.1667 & 0.6667 & -0.1667 \\ 0.6667 & 4.3333 & 0.6667 \\ -0.1667 & 0.6667 & -0.1667 \end{pmatrix} \tag{23}$$

The DCT-based multi-focus image fusion technique divides the input images and the sharpened input images into 8 × 8 blocks and the DCT coefficients of the sharpened in put images blocks are calculated. So, the variance of the sharpened input image's blocks is calculated. This algorithm compares the variance value of the corresponding blocks of sharpened input images and gives discrete values defined as (24) to them for making a decision map $M(m, n)$.

$$M(m,n) = \begin{Bmatrix} 1 \text{ if var}(\text{im}A\text{sharp}) > var(\text{im}B\text{sharp}) + \text{T} \\ -1 \text{ if var}(\text{im}A\text{sharp}) < var(\text{im}B\text{sharp}) - \text{T} \\ 0 \qquad\qquad \text{otherwise} \end{Bmatrix} \tag{24}$$

Once the sharpness-based image fusion is completed in the DCT domain, Consistency Verification (CV) is employed as a post-processing step to further refine the fused image, ensuring improved accuracy and visual quality.

2.7 DCT with DCT + Corr_eng and DCT with DCT + Corr_eng + CV

The correlation coefficient detailed mathematical explanation is given in the above Section 2.4. However, we will continue to explain the energy-correlation coefficient. So coalescing the image energy of A and B in the correlation coefficient relation could better the focus measurement performance. The input image is represented by symbol A, and the artificially blurred input image is represented by symbol B. The energies of the input images, A and B, are presented as Eqs. (25) and (26), respectively (A comprehensive detail is presented in [18]):

$$\text{Energy}_{\text{DCT}}(\text{im}A) = \sum_{i=0}^{\text{n}-1}\sum_{i=0}^{\text{n}-1} \text{d}_{\text{im}A}(i,j)^2 \tag{25}$$

$$\text{Energy}_{\text{DCT}}\left(\overline{\text{im}A}\right) = \sum_{i=0}^{\text{n}-1}\sum_{i=0}^{\text{n}-1} \text{d}_{\overline{\text{im}A}}(i,j)^2 \tag{26}$$

The calculation of (Eng_Corr) is using (27) by combining of correlation coefficients Eq. (13) with Eqs. (25) and (26), accordingly [18]:

$$\text{Eng\_corr}_{\text{DCT}}\left(\text{im}A, \overline{\text{im}A}\right) = \text{Energy}_{\text{DCT}}(\text{im}A) \times 1_{-\text{corr}\,\text{DCT}}\left(\text{im}A, \overline{\text{im}A}\right) = \text{Energy}_{\text{DCT}}\left(\overline{\text{im}A}\right) \tag{27}$$

3 Comparison

Comparison of the advantages and disadvantages of DCT-based multi-focus image fusion methods in Table 1.

Table 1: Advantages and disadvantages of DCT-based methods

Sr. no.	Fusion method	Advantages	Disadvantages
1	DCT + Variance	➢ Enhances the quality of the resultant image. ➢ Reduces complexity, making it suitable for real-time applications. ➢ Simple to compute and implement. ➢ Works well in high-contrast regions.	➢ Variance for a single image may not always capture all important details. ➢ Sensitive to noise, which can lead to undesired distortions.
2	DCT + Correlation Coefficient	➢ Simple and efficient approach. ➢ Offers flexibility in fusing images with different depths of focus. ➢ Helps determine the strength of relationships between image blocks. ➢ Works well in statistical image analysis. ➢ Balanced fusion: This technique balances the energy information and the correlation between images, leading to a fused image that retains both energy features and structural details. ➢ Flexibility: The method can adapt to different types of input images, making it versatile for various fusion tasks [19].	➢ Cannot determine cause and effect relationships between image features. ➢ No control over external factors that may influence correlation. ➢ Less effective when dealing with complex textures and noisy images. ➢ Increased complexity: The combination of energy and correlation measures with DCT increases the overall complexity, requiring careful tuning of parameters. ➢ Resource-intensive: This method may require significant computational resources, particularly for high-resolution images or large datasets.

(Continued)

Table 1 (continued)

Sr. no.	Fusion method	Advantages	Disadvantages
3	DCT + Singular Value Decomposition (SVD)	➢ Enhances the quality of the fused image while reducing processing time. ➢ More stable in noisy conditions compared to other methods. ➢ Can be applied in both DCT and spatial domains. ➢ Retains important structural information of the image. ➢ Enhanced detail preservation: The combination of DCT and SVD enhances detail preservation by leveraging the strengths of both techniques—DCT's energy compaction and SVD's ability to extract significant features. ➢ Robustness: This method is robust against noise and minor image distortions, making it suitable for various practical applications [20].	➢ Computationally expensive, making it less suitable for real-time applications. ➢ Sensitive to missing data. ➢ Requires careful tuning for optimal performance. ➢ Increased computational load: Adding SVD to DCT increases the computational complexity, which may not be ideal for real-time processing. ➢ Dependency on singular values: The performance heavily depends on the selection of singular values, which can be challenging to optimize for different types of images.
4	Sharpening Method (Unsharp Masking)	➢ Improves the quality of the resultant image. ➢ Simple and widely used in many applications. ➢ Enhances edges and fine details effectively. ➢ Works well for improving contrast [21].	➢ Highly sensitive to noise, leading to undesirable distortions in the image. ➢ Over-enhancement of high-contrast areas can introduce unnatural artifacts. ➢ May cause halo effects around the edges.
5	DCT + Energy-correlation Coefficient	➢ Enhanced the overall quality of the resultant image. ➢ Importantly reduces errors caused by incorrect block selection. ➢ Provides better structure preservation. ➢ Works well in complex image fusion scenarios [8].	➢ Computationally complex, requiring higher processing power. ➢ Difficult to implement due to complex mathematical operations. ➢ Requires careful parameter tuning for optimal results.

4 Experiments

4.1 Performance Metrics

This Comparative analysis is based on DCT-based methods in multi-focused image fusion. We are trying to demonstrate the best results in all of them and broaden the knowledge about each technique of DCT based on multi-focused. For demonstrating the best results using six quantitative measures in Table 2.

Table 2: Summary of quantitative measures

Ref.	Metrics	Formula	Description	Best value
[22]	Root Mean Square Error (RMSE)	$RMSE = \frac{1}{MN}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(G(i,j)-F(i,j))^2}$	RMSE is utilized in calculating the difference between the original image and resultant image by explicitly calculating pixel value variations. RMSE is very denoting the resultant image spectral quality.	Lower (close to zero)
[23]	Entropy	$E = -\sum_{k=0}^{G-1} S_k \mathrm{Log} S_k$	Entropy (E) is an important quantitative measure, which can be utilized to differentiate the texture, appearance, or information contents in the image.	Higher value
[24]	Percentage Fit Error (PFE)	$\mathrm{PFE} = \frac{\mathrm{norm}\ (\mathrm{Ir}-\mathrm{If})}{\mathrm{norm}\ (\mathrm{Ir})} * 100$	It is computed as the norm of the difference between the corresponding pixels of the true image and the resulting image to the norm of the true image.	Lower (equal to zero)
[19]	Correlation Coefficient (Corr)	$Corr = \frac{2C_{zp}}{C_z + C_P}$ $C_{zp} = \sum_{a=1}^{M}\sum_{b=1}^{N} I_z(a,b) * I_p(a,b)$ $C_z = \sum_{a=1}^{M}\sum_{b=1}^{N} I_z(a,b)^2$ $C_p = \sum_{a=1}^{M}\sum_{b=1}^{N} I_p(a,b)^2$	The CORR is a quantitative metric that demonstrates the correlation between the true image and the resultant image. When the true and resultant images look the same, the value will be near to one. If the true and resultant images are dissimilar, then the value will be near zero.	Higher value
[25]	Signal to Noise Ratio (SNR)	$SNR = 10\log_{10}\left(\frac{\sum_{a=1}^{M}\sum_{b=1}^{N}(I_z(a,b))^2}{\sum_{a=1}^{M}\sum_{b=1}^{N}(I_z(a,b)-I_z(a,b))^2}\right)$	SNR is the measure of performance used to obtain the ratio between the information and the noise of the output image.	Higher value
[26]	Peak Signal to Noise Ratio (PSNR)	$PSNR = 20\log\left[\frac{G^2}{\frac{1}{M\times N}\sum_{a=1}^{M}\sum_{b=1}^{N}(I_z(a,b)-I_p(a,b))^2}\right]$	PSNR is one of the major metrics and is widely applied in fusion. PSNR is employed for the measurement of the spatial quality of the image. The calculation is done by the number of grey levels divided by the same pixels in the actual and the outcome images.	Higher value

4.2 Results and Analysis

This study presents a comparative analysis of various DCT-based methods for multi-focus image fusion. The approaches under evaluation include DCT + Variance, DCT + Variance + CV (Consistency Verification), DCT + Correlation Coefficient (CC), DCT + CC + CV, DCT + Singular Value Decomposition (SVD), DCT + SVD + CV, DCT + Sharpening, DCT + Sharpening + CV, DCT + Correlation Energy (Corr_Eng), and DCT + Corr_Eng + CV. The performance of these methods is assessed using three evaluation criteria: Quantitative measures, Qualitative Error Image (QEI) measures, and Qualitative visual assessment. The quantitative evaluation is conducted using six different measures, namely RMSE, PEF, Entropy, SNR, PSNR, and Correlation. For experimentation, well-known standard multi-focus images, including lab, and doll images, are used. Each image in the dataset has a resolution of 520 × 520 pixels. The fusion process involves two images, but the proposed methods can also be extended to handle more than two multi-focus images [18].

Additionally, the Qualitative Error Image (QEI)—also known as the difference image—is generated by subtracting the fused image from the reference image. A less visible QEI indicates a higher similarity between the fused and reference images, signifying better fusion quality [27].

The qualitative results of all the DCT-based approaches are illustrated in Fig. 2. This figure presents the fused image alongside its corresponding error (difference) image for the DCT + Correlation (DCT + Corr) method. Following that, additional images showcase the fused and error images generated by the extended DCT + Corr method incorporating Consistency Verification (CV). From these visual comparisons, it is evident that the difference image of DCT + Corr + CV appears to be more informative than that of DCT + Corr alone. This observation leads us to explore the reasoning behind the improvements brought by the Consistency Verification (CV) process. To illustrate, consider a scene composed of multiple blocks, all of which fall entirely within the depth of focus of image A. Ideally, all blocks in this region should be selected from image A. However, due to the presence of noise or unintended artifacts, some blocks may erroneously be chosen from image B, leading to inconsistencies in the fused image. The CV process effectively corrects such errors by ensuring that the final selection aligns with the majority of correctly chosen blocks, thereby enhancing the overall quality of the fused image [28,29].

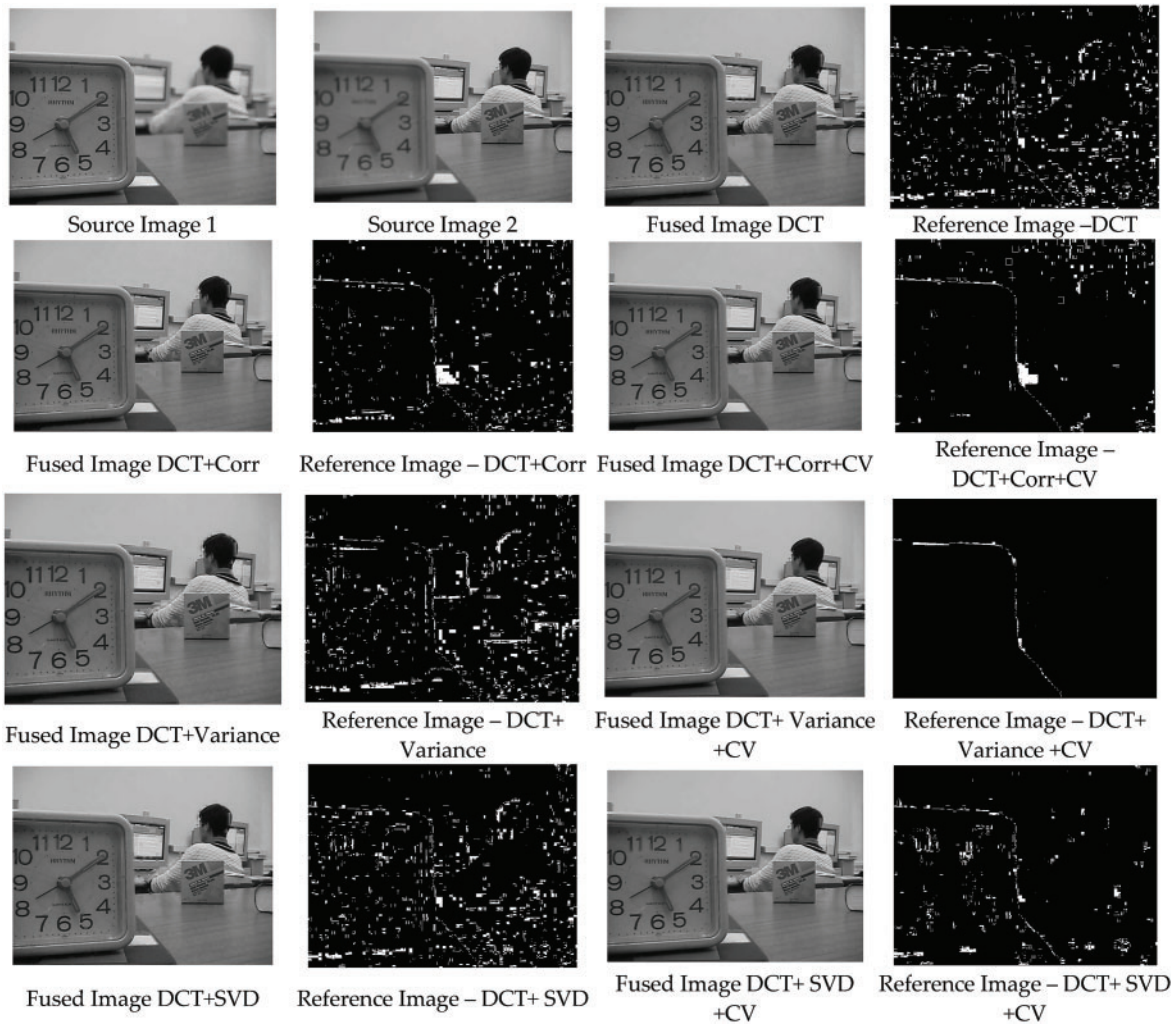

Figure 2: (Continued)

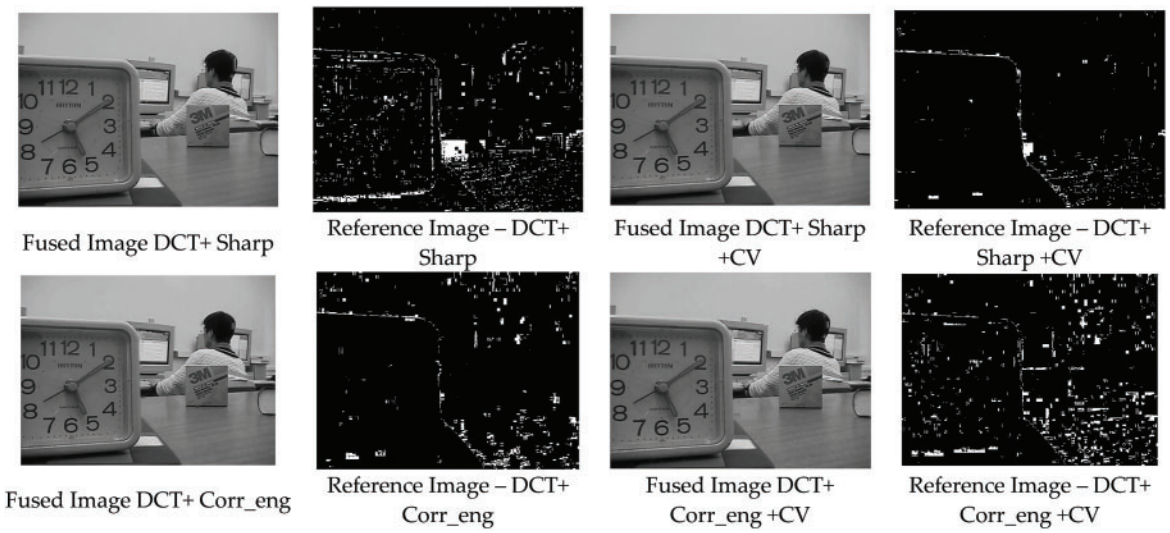

Figure 2: The fused images and corresponding error images produced by different DCT-based methods for test Image 1

Burger and Burge [13] introduced a majority filter that plays an important part in the consistency verification (CV) process. Based on this method, if the center block of a region is chosen from source image B, while most of its surrounding blocks are drawn from source image A, the center block is replaced with its corresponding block from image A.

This method ensures that the fused image is constructed with greater accuracy and consistency. By following this straightforward procedure, all DCT-based methods integrate consistency verification (CV) to enhance fusion results. The proposed method, when extended with CV, further refines the output, leading to improved image quality and reduced errors. Based on our observations, a comparison of the error images in Figs. 2 and 3 for all DCT-based methods reveals a clear pattern. Observing these images, it is clear that the DCT + Variance + CV approach consistently yields better qualitative results for all the three datasets. This indicates that the use of Consistency Verification (CV) improves the fusion process and results in a more accurate and refined final image.

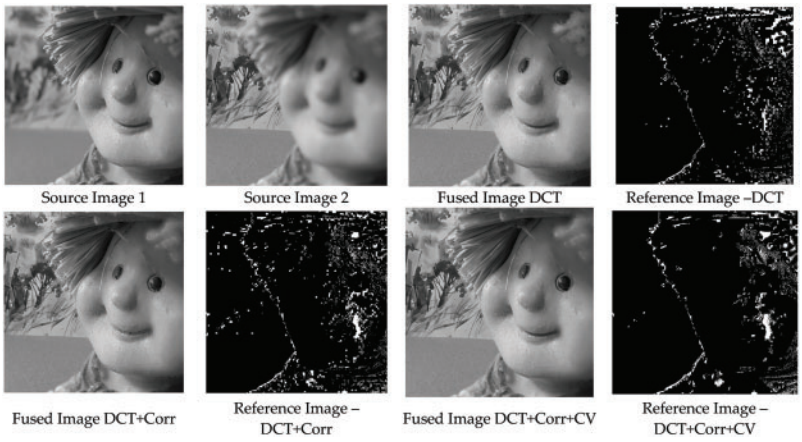

Figure 3: (Continued)

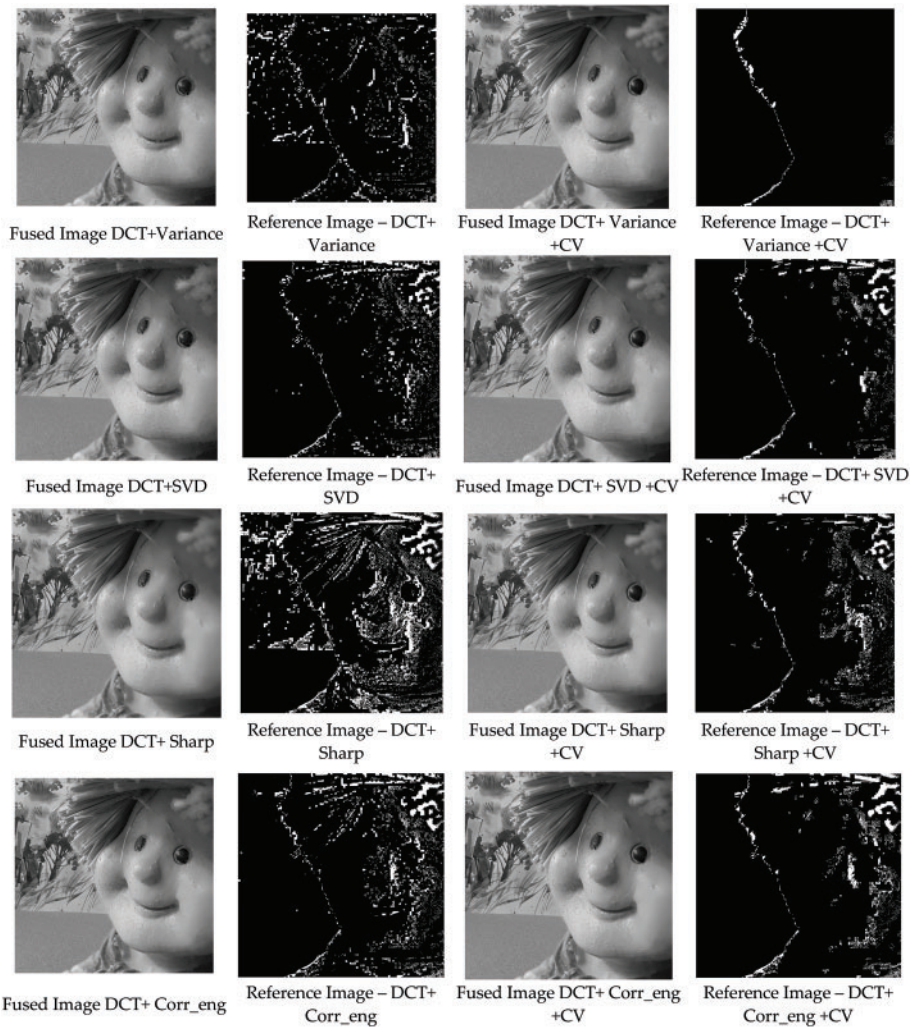

Figure 3: The fused images and corresponding error images produced by different DCT-based methods for test Image 2

The quantitative analysis is conducted using six different evaluation metrics, revealing some intriguing insights. The statistical values are notably close to one another, indicating that all DCT-based methods demonstrate strong performance.

From Table 3, the DCT + Variance method appears to yield the best results for the Disk image set. In Table 4, the PEF metric performs well with DCT + SVD, while entropy values are higher for DCT + Sharpen + CV. Meanwhile, the Lab image set results indicate that most metrics favor DCT + Variance.

For the Doll image set, DCT + Variance achieves the best scores for RMSE and PEF, while the remaining metrics show higher values for DCT + Variance + CV. Based on both visual and statistical comparisons, our analysis suggests that DCT + Variance and DCT + Variance + CV stand out as the most effective methods among all DCT-based approaches.

Table 3: The quantitative results of DCT-based approaches using Lab image sets

Matrices	DCT + Corr	DCT + Corr + CV	DCT + Variance	DCT + Variance + CV	DCT + SDV	DCT + SDV + CV	DCT + Sharp	DCT + Sharp + CV	DCT + Corr_eng	DCT + Corr_eng + CV
RMSE	12.1542	12.2989	**11.9258**	12.5200	12.0189	12.4451	12.0213	12.3603	12.0557	12.2923
PFE	3.5469	3.6464	3.6432	3.7566	**3.2436**	3.6778	3.5299	3.7156	3.5387	3.6574
Entropy	0.1348	0.1196	0.1403	0.1064	0.1246	0.1103	0.1608	**0.1935**	0.1128	0.0818
SNR	20.1086	20.0058	**20.2734**	19.8511	20.2059	19.9032	20.2041	19.9626	20.1793	20.0105
PSNR	37.3175	37.2661	**37.3999**	37.1888	37.3662	37.2148	37.3653	37.2445	37.3529	37.2685
Correlation	0.9957	0.9956	**0.9959**	0.9954	0.9958	0.9955	0.9958	0.9955	0.9958	0.9956

Note: The bold values show the best results among all.

Table 4: The quantitative results of DCT-based approaches using Lab image sets

Matrices	DCT + Corr	DCT + Corr + CV	DCT + Variance	DCT + Variance + CV	DCT + SDV	DCT + SDV + CV	DCT + Sharp	DCT + Sharp + CV	DCT + Corr_eng	DCT + Corr_eng + CV
RMSE	8.0140	8.1007	**8.0013**	8.8278	8.4002	8.5047	8.0017	8.5352	8.3854	8.5264
PFE	1.3696	1.4370	**1.3251**	1.7146	1.4882	1.6153	1.3688	1.6131	1.6260	1.6531
Entropy	0.0417	0.0522	0.0528	**0.0686**	0.0618	0.0634	0.0476	0.0618	0.0680	0.0663
SNR	23.6820	23.5885	22.8977	**23.8419**	23.2732	23.1658	23.6954	23.1347	23.2885	23.1437
PSNR	39.1263	39.0796	38.7342	**39.7063**	38.9219	38.8682	39.1330	38.8526	38.9296	38.8571
Correlation	0.9981	0.9980	0.9977	**0.9986**	0.9979	0.9978	0.9981	0.9978	0.9979	0.9978

Note: The bold values show the best results among all.

4.3 Simulation Conditions

The simulations for the DCT-based methods were conducted using MATLAB R2020b on a system equipped with an Intel Core i7-6700K processor (4.00 GHz) and 8 GB of RAM. The simulation codes for all the methods were obtained from an online database, originally provided by Haghighat et al. [15] and further contributed by Naji et al. [16,18,20,21].

5 Conclusion

In this article, a comparative analysis is conducted to evaluate DCT-based methods. The DCT-based methods of multi-focused image fusion are proven to be more suitable and time-saving in real-time systems for still images or videos. This comparative analysis will broaden the knowledge about a bunch of techniques and explain the technical modifications in these methods for emerging researchers. So, all DCT-based methods are technically (mathematically) elaborated, and the results are compared using six commonly used metrics on two different image sets. According to experiments, we are assuming that based on visual and statistical results, the DCT + Variance and DCT + Variance + CV methods performed well and achieved the most successful results for all of the image sets. The investigation provided a framework for further study in multi-focus image fusion on DCT-based methods. The improvements to the above disadvantages are left as future work.

While the comparative results provide valuable insight into the strengths of DCT-based methods, the study acknowledges that these methods are primarily traditional and may have limitations in handling highly dynamic or complex real-world scenarios. The absence of deep learning-based comparisons, adaptive thresholding techniques, and fusion performance under noisy conditions remains a gap that should be addressed. In future work, we plan to extend this study by:

- Integrating deep learning approaches (e.g., CNNs and Transformers) with DCT-based methods to enhance fusion quality.
- Introducing optimization frameworks for dynamic parameter tuning in DCT variants.
- Conducting large-scale evaluations on diverse and real-time datasets, including noisy, medical, and aerial images.
- Incorporating statistical significance testing (e.g., p-values) to validate differences in performance.
- Exploring hybrid frameworks that combine spatial and frequency domain fusion techniques.

These directions will not only overcome the existing limitations but also offer a more robust, generalizable, and modern approach to multi-focus image fusion.

Acknowledgement: We would like to acknowledge the support of our team members and the Kaggle repository for providing the data. Our sincere gratitude goes team leader for his invaluable guidance and feedback throughout this research.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Muhammad Osama and Sarwar Shah Khan; data collection: Sajid Khan and Reshma Khan; analysis and interpretation of results: Muhammad Osama, Sarwar Shah Khan, Sajid Khan, Muzammil Khan and Mian Muhammad Danyal; draft manuscript preparation: Muhammad Osama, Sarwar Shah Khan, Reshma Khan, Muzammil Khan and Mian Muhammad Danyal. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in the Kaggle repository at: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset (accessed on 17 July 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Mohamed MA, El-Den BM. Implementation of image fusion techniques for multi-focus images using FPGA. In: Proceedings of the 2011 28th National Radio Science Conference (NRSC); 2011 Apr 26–28; Cairo, Egypt. p. 1–11. doi:10.1109/NRSC.2011.5873618.
2. Siddiqui AB, Jaffar MA, Hussain A, Mirza AM. Block-based feature-level multi-focus image fusion, Busan, Republic of Korea. In: Proceedings of the 2010 5th International Conference on Future Information Technology; 2010 May 21–23; Busan, Republic of Korea. p. 1–7. doi:10.1109/FUTURETECH.2010.5482718.
3. Masood S, Sharif M, Yasmin M, Shahid MA, Rehman A. Image fusion methods: a survey. J Eng Sci Technol Rev. 2017;10(6):186–95. doi:10.25103/jestr.106.24.
4. Jiang S, Yu S. Refined Multi-Focus image fusion on using multi-scale neural network with SpSwin Autoencoder-based matting. Expert Syst Appl. 2025;276:126980. doi:10.1016/j.eswa.2025.126980.
5. Kumari AV. A review of image enhancement techniques. Int J Trend Res Dev. 2017;2(7):232–35.
6. Javed U, Riaz MM, Ghafoor A, Ali SS, Cheema TA. MRI and PET image fusion using fuzzy logic and image local features. Sci World J. 2014;2014:708075. doi:10.1155/2014/708075.
7. Bhosekar S, Singh P, Garg D. A recent advancements in multi-modal medical image fusion techniques using SWT, NSST, NSCT, and CNN. In: Proceedings of the 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI); 2025 Jan 16–18; Greater Noida, India. Piscataway, NJ, USA: IEEE. p. 928–32.
8. Wang M, Shang X. A fast image fusion with discrete cosine transform. IEEE Signal Process Lett. 2020;27:990–4. doi:10.1109/LSP.2020.2999788.

9. Oral M, Turgut SS. A comparative study for image fusion. In: Proceedings of the 2018 Innovations in Intelligent Systems and Applications Conference (ASYU); 2018 Oct 4–6; Adana, Türkiye. p. 1–6. doi:10.1109/asyu.2018.8554000.
10. Zhou Y, Yu L, Zhi C, Huang C, Wang S, Zhu M, et al. A survey of multi-focus image fusion methods. Appl Sci. 2022;12(12):6281. doi:10.3390/app12126281.
11. Zhang H, Xu H, Tian X, Jiang J, Ma J. Image fusion meets deep learning: a survey and perspective. Inf Fusion. 2021;76:323–36. doi:10.1016/j.inffus.2021.06.008.
12. Wang Z, Ziou D, Armenakis C, Li D, Li Q. A comparative analysis of image fusion methods. IEEE Trans Geosci Remote Sens. 2005;43(6):1391–402. doi:10.1109/TGRS.2005.846874.
13. Burger W, Burge MJ. The discrete cosine transform (DCT). In: Digital image processing. Cham, Switzerland: Springer International Publishing; 2022. p. 589–97. doi:10.1007/978-3-031-05744-1_20.
14. Amin-Naji M, Aghagolzadeh A. Multi-focus image fusion using VOL and EOL in DCT domain. arXiv:1710.06511. 2017.
15. Haghighat MBA, Aghagolzadeh A, Seyedarabi H. Real-time fusion of multi-focus images for visual sensor networks. In: Proceedings of the 2010 6th Iranian Conference on Machine Vision and Image Processing; 2010 Oct 27–28; Isfahan, Iran. p. 1–6. doi:10.1109/IranianMVIP.2010.5941140.
16. Naji MA, Aghagolzadeh A. Multi-focus image fusion in DCT domain based on correlation coefficient. In: Proceedings of the 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI); 2015 Nov 5–6; Tehran, Iran. p. 632–9. doi:10.1109/KBEI.2015.7436118.
17. Nie X, Xiao B, Bi X, Li W, Gao X. A focus measure in discrete cosine transform domain for multi-focus image fast fusion. Neurocomputing. 2021;465:93–102. doi:10.1016/j.neucom.2021.08.109.
18. Naji MA, Aghagolzadeh A. Multi-focus image fusion in DCT domain using variance and energy of Laplacian and correlation coefficient for visual sensor networks. J AI Data Min. 2018;6(2):233–50. doi:10.1109/kbei.2015.7436118.
19. Osama M, Khan SS, Khan S. High-quality multi-focus image fusion: a comparative analysis of DCT-based approaches with their variants. ICCK J Image Anal Process. 2025;1(1):27–35.
20. Naji MA, Ranjbar-Noiey P, Aghagolzadeh A. Multi-focus image fusion using singular value decomposition in DCT domain. In: Proceedings of the 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP); 2017 Nov 22–23; Isfahan, Iran. p. 45–51. doi:10.1109/IranianMVIP.2017.8342367.
21. Naji MA, Aghagolzadeh A. A new multi-focus image fusion technique based on variance in DCT domain. In: Proceedings of the 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI); 2015 Nov 5–6; Tehran, Iran. p. 478–84. doi:10.1109/KBEI.2015.7436092.
22. Danyal MM, Khan S, Khan RS, Jan S, Rahman N. Enhancing multi-modality medical imaging: a novel approach with Laplacian filter + discrete Fourier transform pre-processing and stationary wavelet transform fusion. J Intell Med Healthc. 2024;2:35–53. doi:10.32604/jimh.2024.051340.
23. Khan SS, Khan M, Alharbi Y, Haider U, Ullah K, Haider S. Hybrid sharpening transformation approach for multifocus image fusion using medical and nonmedical images. J Healthc Eng. 2021;2021:7000991. doi:10.1155/2021/7000991.
24. Dogra A, Kumar S. Multi-modality medical image fusion based on guided filter and image statistics in multidirectional shearlet transform domain. J Ambient Intell Humaniz Comput. 2023;14(9):12191–205. doi:10.1007/s12652-022-03764-6.
25. Moushmi S, Sowmya V, Soman KP. Empirical wavelet transform for multifocus image fusion. In: Proceedings of the International Conference on Soft Computing Systems; 2015; New Delhi, India: Springer. p. 257–63. doi:10.1007/978-81-322-2671-0_25.
26. Singh S, Singh H, Bueno G, Deniz O, Singh S, Monga H, et al. A review of image fusion: methods, applications and performance metrics. Digit Signal Process. 2023;137:104020. doi:10.1016/j.dsp.2023.104020.
27. Khan SS, Khan M, Ran Q. Multi-focus color image fusion using Laplacian filter and discrete Fourier transformation with qualitative error image metrics. In: Proceedings of the 2nd International Conference on Control and Computer Vision; 2019; Jeju, Republic of Korea: ACM. p. 41–5. doi:10.1145/3341016.3341019.

28. Haghighat MBA, Aghagolzadeh A, Seyedarabi H. Multi-focus image fusion for visual sensor networks in DCT domain. Comput Electr Eng. 2011;37(5):789–97. doi:10.1016/j.compeleceng.2011.04.016.
29. Shah M, Khan M, Khan SS, Ali S. Multi-focus image fusion using unsharp masking with discrete cosine transform. In: Proceedings of 1st International Conference on Computing Technologies, Tools and Applications (ICTAPP-23); 2023 May 9–11; Peshawar, Pakistan.