



OPEN ACCESS

EDITED BY
Zhiqin Zhu,
Chongqing University of Posts and
Telecommunications, China

REVIEWED BY
Xiaosong Li,
Foshan University, China
Zhiliang Wu,
Zhejiang University, China

*CORRESPONDENCE
Tiande Ma,
✉ 20201303225@stu.xju.edu.cn

RECEIVED 12 February 2025
ACCEPTED 05 March 2025
PUBLISHED 02 April 2025

CITATION
Jia Y and Ma T (2025) Multi-focus image
fusion based on pulse coupled neural
network and WSEML in DTCWT domain.
Front. Phys. 13:1575606.
doi: 10.3389/fphy.2025.1575606

COPYRIGHT
© 2025 Jia and Ma. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Multi-focus image fusion based on pulse coupled neural network and WSEML in DTCWT domain

Yuan Jia¹ and Tiande Ma^{2*}

¹School of Statistics, Renmin University of China, Beijing, China, ²School of Computer Science and Technology, Xinjiang University, Urumqi, China

The goal of multi-focus image fusion is to merge near-focus and far-focus images of the same scene to obtain an all-focus image that accurately and comprehensively represents the focus information of the entire scene. The current multi-focus fusion algorithms lead to issues such as the loss of details and edges, as well as local blurring in the resulting images. To solve these problems, a novel multi-focus image fusion method based on pulse coupled neural network (PCNN) and weighted sum of eight-neighborhood-based modified Laplacian (WSEML) in dual-tree complex wavelet transform (DTCWT) domain is proposed in this paper. The source images are decomposed by DTCWT into low- and high-frequency components, respectively; then the average gradient (AG) motivate PCNN-based fusion rule is used to process the low-frequency components, and the WSEML-based fusion rule is used to process the high-frequency components; we conducted simulation experiments on the public Lytro dataset, demonstrating the superiority of the algorithm we proposed.

KEYWORDS

multi-focus image, image fusion, DTCWT, PCNN, WSEML

1 Introduction

Multi-focus image fusion is a technique in the field of image processing that combines multiple images, each focused on different objects or regions, into a single image that captures the sharp details from all focal points [1]. This approach is particularly useful in applications where the depth of field is limited, such as in macro photography, surveillance, medical imaging, and robotics [2, 3].

In typical photography, a single image can only present objects within a certain range of focus clearly, leaving objects closer or farther away blurry [4, 5]. However, by capturing several images with different focus points and then combining them through image fusion techniques, it is possible to create a final image that maintains sharpness across a wider range of depths [6–8].

The process of multi-focus image fusion generally involves several key steps: image alignment, where all the images are aligned spatially; focus measurement, where the sharpness of various regions in each image is assessed; and fusion, where the sharpest information from each image is retained [9–11]. Advanced fusion algorithms, including pixel-level, transform-domain, and machine learning-based methods, can be employed to optimize the fusion quality and preserve important features from all focused regions. This technology has a broad range of applications. In medical imaging, it helps to create clearer, more detailed visualizations of organs or tissues. In surveillance, it enhances the clarity of

objects at varying distances. In robotics, it contributes to improved perception by enabling robots to focus on multiple objects simultaneously [12, 13]. As computational power and algorithms continue to advance, multi-focus image fusion is expected to play an increasingly significant role in a variety of fields requiring high-quality visual information [14–17].

Currently, image fusion can be categorized into two types: traditional algorithms and deep learning algorithms [18–20]. Traditional algorithms typically rely on handcrafted features and conventional image processing techniques, such as Laplacian pyramid [21], wavelet transform [22], dual-tree complex wavelet transform (DTCWT) [23], contourlet [24–26], shearlet [27, 28] and gradient-based methods [29], to combine focused regions from multiple images. Mohan et al. [30] introduced the multi-focus image fusion method based on quarter shift dual-tree complex wavelet transform (qshiftN DTCWT) and modified principal component analysis (MPCA) in the Laplacian pyramid (LP) domain, and this method outperforms many state-of-the-art techniques in terms of visual and quantitative evaluations. Mohan et al. [31] introduced the image fusion method based on DTCWT combined with stationary wavelet transform (SWT). Lu et al. [32] introduced the multi-focus image fusion using residual removal and fractional order differentiation focus measure, and this algorithm simultaneously employs nonsubsampling shearlet transform and the sum of Gaussian-based fractional-order differentiation. These methods are generally effective in simpler scenarios, but they may struggle with more complex images, especially when dealing with varying levels of focus and noise. Pulse coupled neural network (PCNN) also has extensive applications in the field of image fusion, Xie et al. [33] proposed the multi-focus image fusion method based on sum-modified Laplacian and PCNN in nonsampled contourlet transform domain, and this method excellently improves the focus clarity.

On the other hand, deep learning has extensive applications in image fusion [34–37], image segmentation [38, 39], and video restoration [40–44], and image super-resolution [45, 46]. Deep learning algorithms leverage convolutional neural networks (CNNs), Transformer, Generative adversarial network (GAN), Mamba and other advanced models to automatically learn features and perform fusion in an end-to-end manner [47–49]. These methods can adapt to a wide range of image complexities, providing more accurate and visually appealing fused images, especially in challenging conditions like low light or high noise environments [50, 51]. Deep learning approaches have shown superior performance in recent years, particularly with the availability of large datasets and powerful computational resources [52, 53].

Inspired by the ideas from the algorithm in Reference [33], in this paper, a novel multi-focus image fusion method based on PCNN and weighted sum of eight-neighborhood-based modified Laplacian (WSEML) in DTCWT domain is proposed. The motivation behind this approach is to achieve a more robust and effective fusion method that can handle complex images with varying focus levels and noise, while also being computationally efficient. The source images are decomposed by DTCWT into low- and high-frequency components, respectively; then the average gradient (AG) motivate PCNN fusion rule is used to process the low-frequency components, and the WSEML-based fusion rule is used to process the high-frequency components. The algorithm's superiority is validated through comparative experiments on public Lytro dataset.

2 DTCWT

The dual-tree complex wavelet transform (DTCWT) is an advanced signal processing technique designed to overcome some of the limitations of the traditional discrete wavelet transform (DWT) [54]. It was introduced to provide better performance in tasks such as image denoising, compression, and feature extraction. The DTCWT is particularly useful for applications where directional sensitivity and shift invariance are important.

The DTCWT provides improved directional information compared to the traditional wavelet transforms. It uses two parallel trees of wavelet filters (hence “dual-tree”), one for the real part and one for the imaginary part. This structure allows for better representation of image features, especially edges and textures, in multiple orientations. Unlike the traditional DWT, which suffers from shift variance (i.e., small translations in the signal can cause large changes in the wavelet coefficients), the DTCWT provides a level of shift invariance [55, 56]. This makes it more robust to small shifts or distortions in the input signal, which is critical for many image and signal processing tasks. The transform uses complex-valued coefficients rather than real-valued coefficients. This allows for better capture of phase information in addition to amplitude, providing more detailed and richer representations of the signal or image. The DTCWT significantly reduces the aliasing effect, a common issue in wavelet transforms when high-frequency components mix with low-frequency ones. The dual-tree structure and the use of complex filters help mitigate this problem [57].

3 The proposed method

The multi-focus image fusion algorithm we proposed can be mainly divided into four steps: image decomposition, low-frequency fusion, high-frequency fusion, and image reconstruction. The structure of the proposed method is shown in Figure 1, and the specific process is as follows.

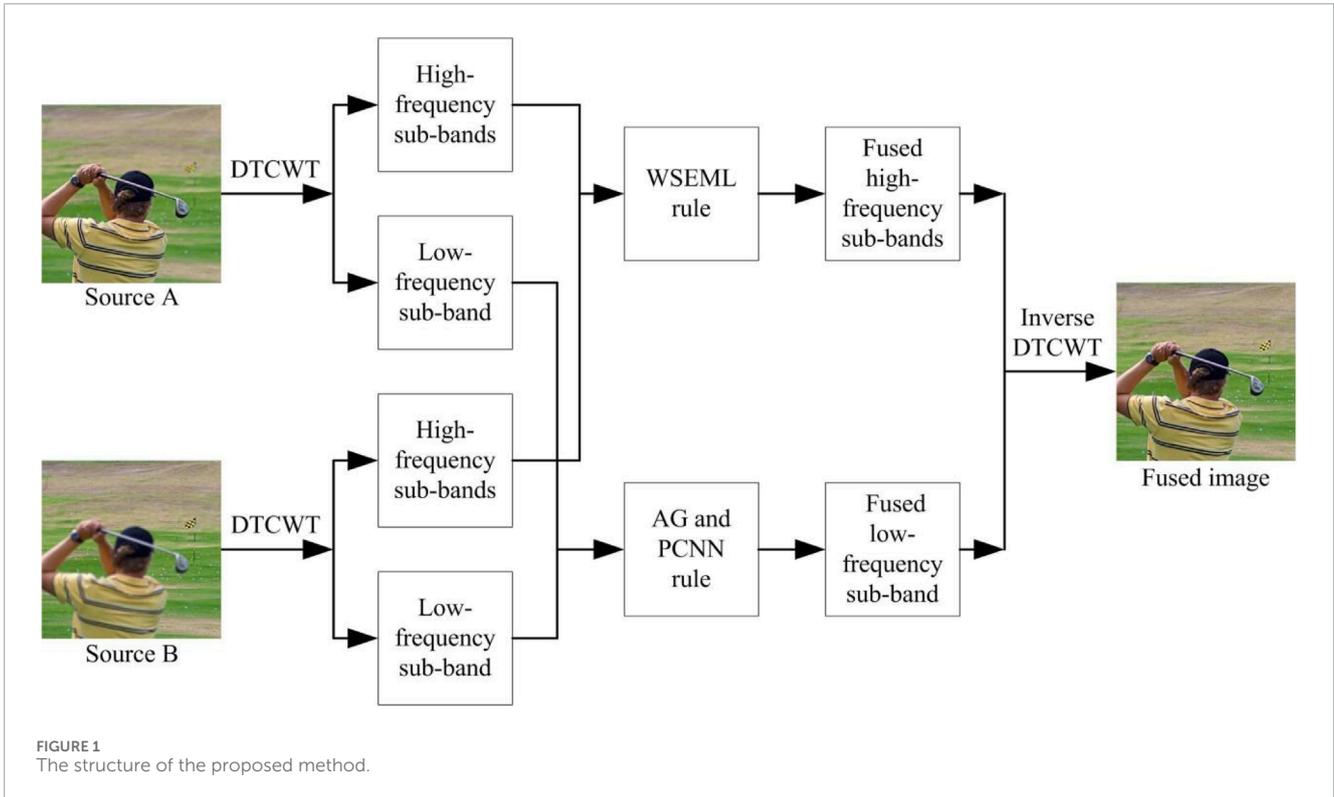
3.1 Image decomposition

The source images A and B are decomposed into low-frequency components $\{L^A, L^B\}$ and high-frequency components $\{H_{l,d}^A, H_{l,d}^B\}$ using DTCWT. The $L^X|X \in (A, B)$ shows the low-frequency, and $H_{l,d}^X|X \in (A, B)$ shows the high-frequency sub-bands l level in the d orientation.

3.2 Low-frequency fusion

The low-frequency component of the image contains the main background information of the image. The average gradient-based (AG) motivate PCNN fusion rule is used to process the low-frequency sub-bands, and the corresponding equations are defined as follows [58, 59]:

$$AG_{ij} = \frac{\sum_i \sum_j ((f(i,j) - f(i+1,j))^2 + (f(i,j) - f(i,j+1))^2)^{\frac{1}{2}}}{mn} \quad (1)$$



$$F_{ij}(n) = AG_{ij} \tag{2}$$

$$L_{ij}(n) = e^{-\alpha_L} L_{ij}(n-1) + V_L \sum_{pq} W_{ij,pq} Y_{ij,pq}(n-1) \tag{3}$$

$$U_{ij}(n) = F_{ij}(n) * (1 + \beta L_{ij}(n)) \tag{4}$$

$$\theta_{ij}(n) = e^{-\alpha_\theta} \theta_{ij}(n-1) + V_\theta Y_{ij}(n-1) \tag{5}$$

$$Y_{ij}(n) = \begin{cases} 1, & \text{if } U_{ij}(n) > \theta_{ij}(n) \\ 0 & \text{else} \end{cases} \tag{6}$$

$$T_{i,j} = T_{i,j}(n-1) + Y_{i,j}(n) \tag{7}$$

In Equation 1, the $f(i,j)$ is pixel intensity at (i,j) and $m \times n$ is the size of the image. In the mathematical model of PCNN in Equations 2–6, the feeding input F_{ij} is equal to the normalized AG_{ij} . The linking input L_{ij} is equal to the sum of neurons firing times in linking range. $W_{ij,pq}$ is the synaptic gain strength and subscripts p and q are the size of linking range in PCNN. α_L is the decay constants. V_L and V_θ are the amplitude gain. β is the linking strength. U_{ij} is total internal activity. θ_{ij} is the threshold. n denotes the iteration times. If U_{ij} is larger than θ_{ij} , then, the neuron will generate a pulse $Y_{ij} = 1$, also called one firing time. In fact, the sum of Y_{ij} in n iteration is often defined as Equation 7, called firing times, to represent image information. Rather than $Y_{ij}(n)$, one often analyzes $T_{ij}(n)$, because neighboring coefficients with similar features representing similar firing times in a given iteration times. AG is input to PCNN to motivate the neurons and generate pulse of

neurons with Equations 2–6. Then, firing times $T_{ij}(n)$ is calculates as Equation 7.

Get the decision map D_{ij} based on Equation 8 and select the coefficients with Equation 9, which means that coefficients with large firing times are selected as coefficients of the fused. The fusion rule is designed as follows:

$$D_{E,ij} = \begin{cases} 1 & \text{If } T_{A,ij}(n) \geq T_{B,ij}(n) \\ 0 & \text{else} \end{cases} \tag{8}$$

$$L^F(i,j) = \begin{cases} L_A(i,j) & \text{If } D_{ij}(n) = 1 \\ L_B(i,j) & \text{If } D_{ij}(n) = 0 \end{cases} \tag{9}$$

where L^F shows the fused low-frequency sub-band.

3.3 High-frequency fusion

The high-frequency component of the image contains the detailed information of the image. The weighted sum of eightneighborhood-based modified Laplacian (WSEML) is used to process the high-frequency sub-bands with Equations 10–12 [60]:

$$WSEML_X(i,j) = \sum_{m=-r}^r \sum_{n=-r}^r \Phi(m+r+1, n+r+1) \times EML_X(i+m, j+n) \tag{10}$$

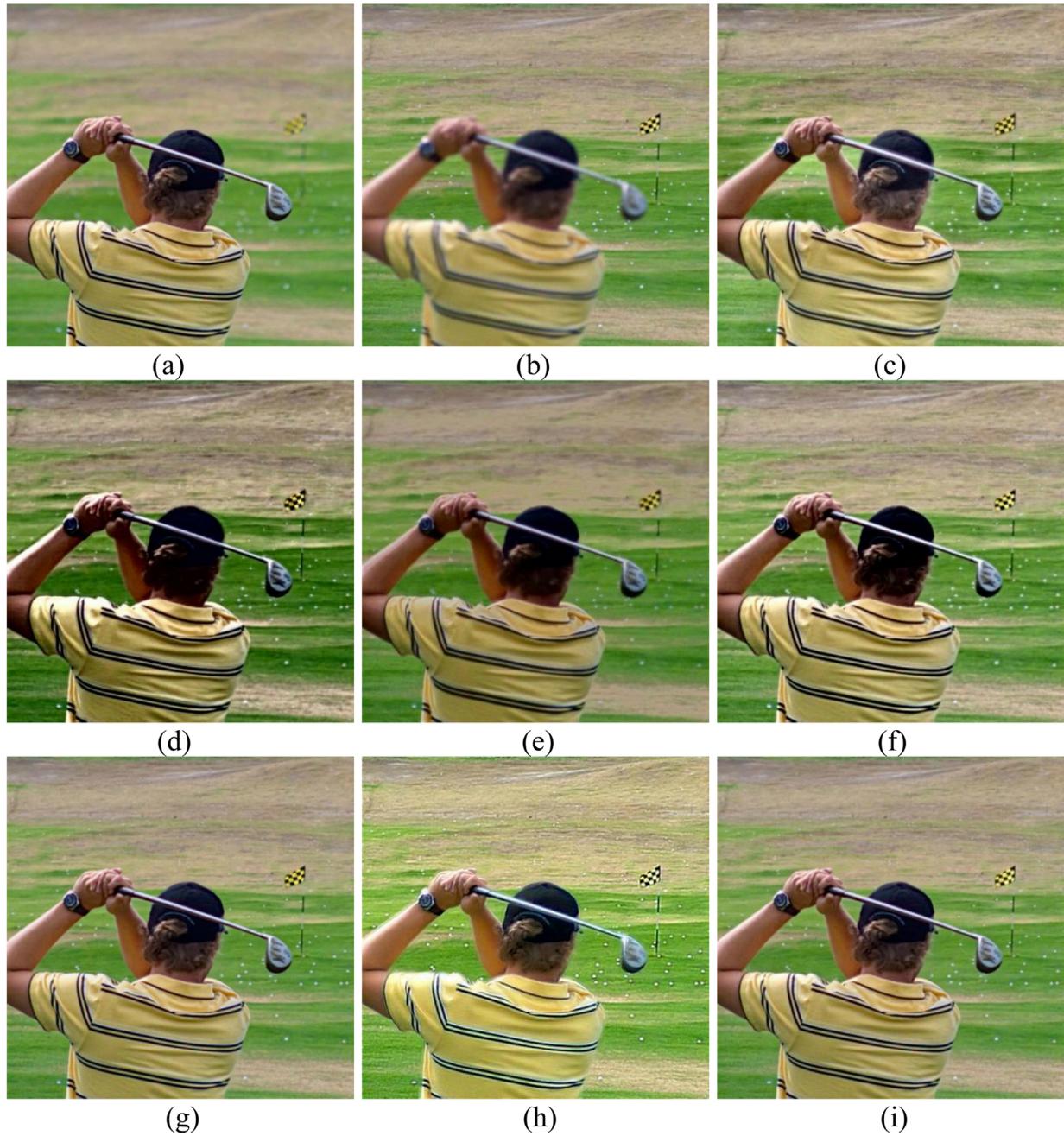


FIGURE 2 Fusion results on Lytro-01. (a) Source A; (b) Source B; (c) GD; (d) FusionDN; (e) PMGI; (f) U2Fusion; (g) ZMFF; (h) UUDFusion; (i) Proposed.

$$\begin{aligned}
 EML_X(i,j) = & |2X(i,j) - X(i-1,j) - X(i+1,j)| \\
 & + |2X(i,j) - X(i,j-1) - X(i,j+1)| \\
 & + \frac{1}{\sqrt{2}} |2X(i,j) - X(i-1,j-1) - X(i+1,j+1)| \\
 & + \frac{1}{\sqrt{2}} |2X(i,j) - X(i-1,j+1) - X(i+1,j-1)|
 \end{aligned} \quad (11)$$

where $X \in \{A, B\}$, and Φ is a $(2r + 1) \times (2r + 1)$ weighting matrix with radius r . For each element in Φ , its value is set to 2^{2r-d} , where d is its four-neighborhood distance to the center. As an example, the 3×3

normalized version of Φ is

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

The fused high-frequency sub-bands are defined as follows:

$$H_{l,d}^F(i,j) = \begin{cases} H_{l,d}^A(i,j) & \text{if } WSEML_{H_{l,d}^A}(i,j) \geq WSEML_{H_{l,d}^B}(i,j) \\ H_{l,d}^B(i,j) & \text{else} \end{cases} \quad (12)$$

where $H_{l,d}^F(i,j)$ shows the fused high-frequency sub-bands.

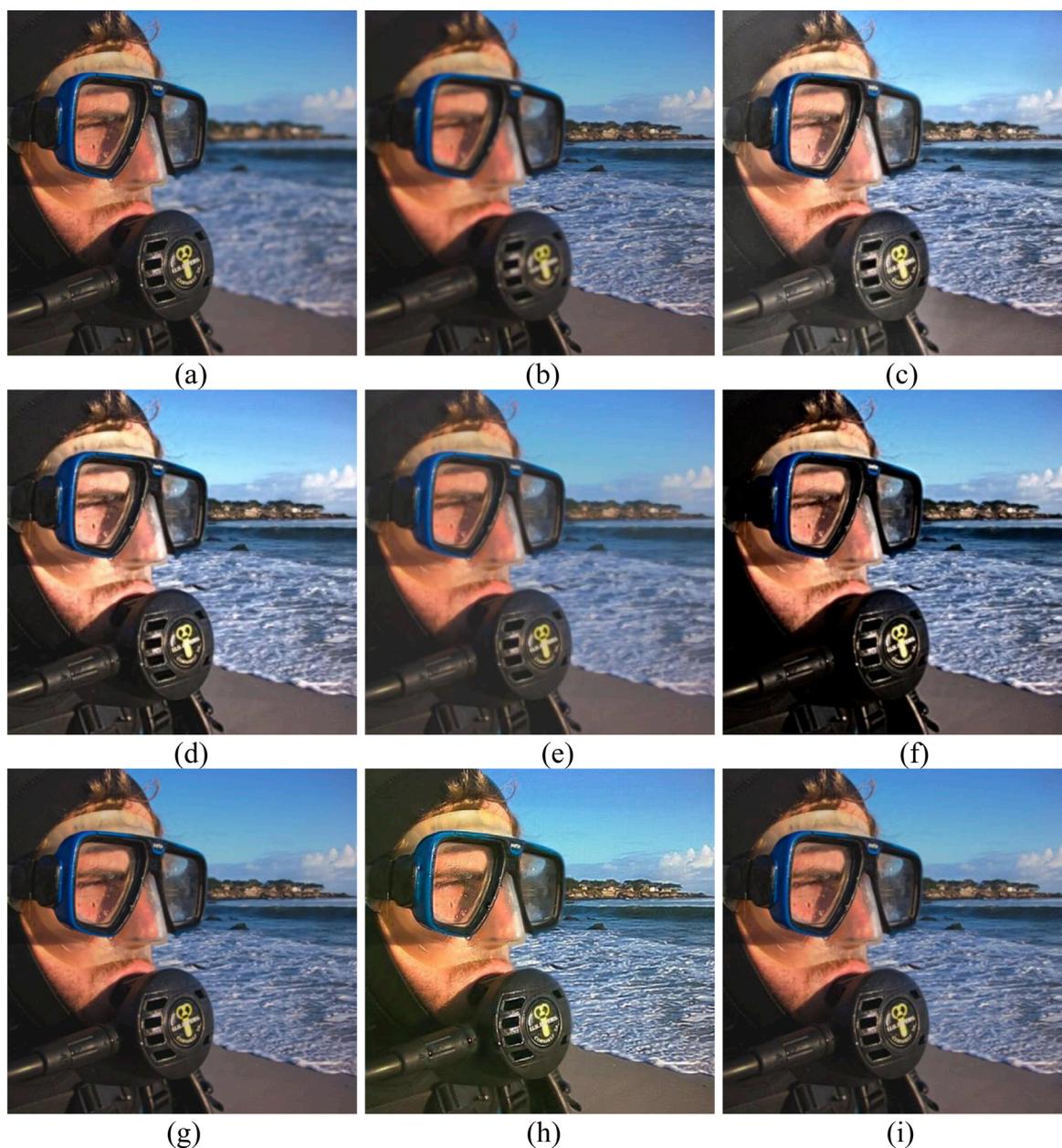


FIGURE 3 Fusion results on Lytro-02. (a) Source A; (b) Source B; (c) GD; (d) FusionDN; (e) PMGI; (f) U2Fusion; (g) ZMFF; (h) UUDFusion; (i) Proposed.

3.4 Image reconstruction

The fused image F is obtained by the inverse DTCWT on $L^F(i, j)$ and $H_{1,d}^F(i, j)$.

4 Experimental results and analysis

To demonstrate the effectiveness of our algorithm, we conducted simulation experiments on the commonly used public Lytro dataset [61] and compared it with six classic image fusion algorithms, namely, GD [29], FusionDN [62], PMGI [63], U2Fusion [64], ZMFF

[65], and UUDFusion [66]. Additionally, we employed six objective evaluation metrics to qualitatively assess the experimental results, namely, edge-based similarity measurement $Q_{AB/F}$ [59], mutual information metric Q_{MI} [59], nonlinear correlation information entropy Q_{NCIE} [67], Chen-Blum metric Q_{CB} [67], image fusion metric-based on phase congruency Q_p [67] and gradient-based fusion performance Q_G [67]. The higher these metric values, the better the fusion effect. We adopt a combined subjective and objective evaluation approach to measure the effectiveness of the algorithms. The parameters of the comparison algorithms were set according to the original papers, while in our algorithm, the decomposition level of DTCWT was set to 4 layers; parameters of



FIGURE 4 Fusion results on Lytro-03. (a) Source A; (b) Source B; (c) GD; (d) FusionDN; (e) PMGI; (f) U2Fusion; (g) ZMFF; (h) UUDFusion; (i) Proposed.

PCNN is set as $p \times q$, $\alpha_L = 0.06931$, $\alpha_\theta = 0.2$, $\beta = 0.2$, $V_L = 1.0$, $V_\theta = 20$, $\Phi = \begin{bmatrix} 0.707 & 1 & 0.707 \\ 1 & 0 & 1 \\ 0.707 & 1 & 0.707 \end{bmatrix}$, and the maximal iterative number is $n = 200$.

Figure 2 shows the fused results with different methods on Lytro-01. The GD method retains significant focus information from both the foreground and background. However, some blending artifacts are visible, and the focus transitions may not be smooth. The FusionDN algorithm preserves structural details well but exhibits some loss of sharpness in the golfer and background.

The fusion quality is moderate, with slight blurring at focus boundaries. The PMGI method achieves reasonable fusion but struggles with preserving contrast and sharpness, especially in the golfer’s details. The background appears slightly oversmoothed. The ZMFF method performs well in maintaining the focus of both the foreground (golfer) and background. The details are well-preserved, but minor artifacts can be noticed in the focus transition areas. The UUDFusion method produces an average fusion result, with noticeable blurring in both the foreground and background. The image lacks the clarity and sharpness needed for an effective all-focus image. The proposed method delivers the best results. Both the golfer (foreground) and the background are sharply focused, with smooth

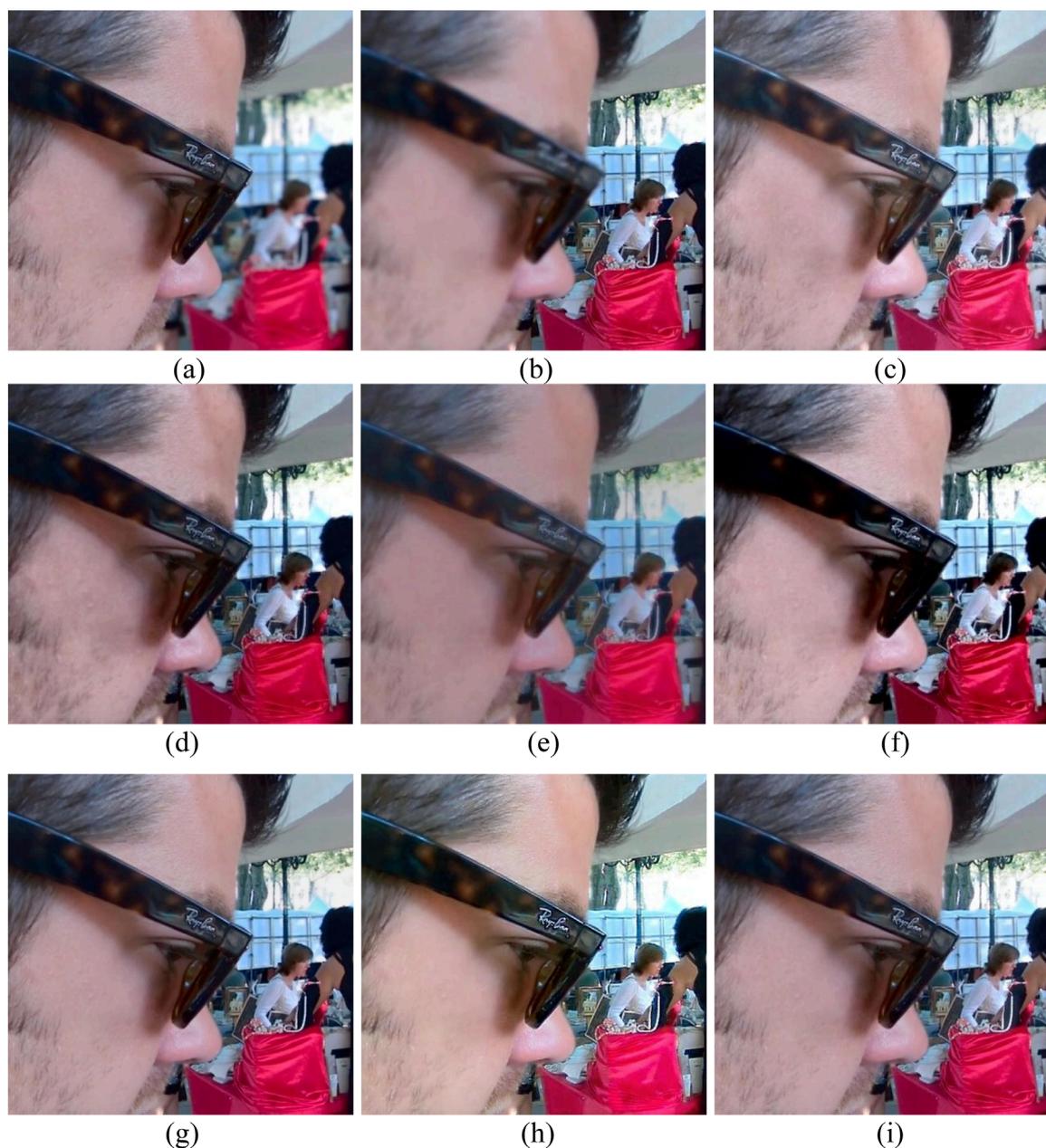


FIGURE 5
Fusion results on Lytro-04. (a) Source A; (b) Source B; (c) GD; (d) FusionDN; (e) PMGI; (f) U2Fusion; (g) ZMFF; (h) UUDFusion; (i) Proposed.

transitions between the focus regions. The image appears natural and well-balanced, with no noticeable artifacts.

Figure 3 presents fusion results for various algorithms applied to the Lytro-02 dataset, aiming to create an all-focus image by combining the near-focus (foreground) and far-focus (background) regions. The proposed method clearly outperforms all other methods, producing a sharp and balanced image where both the diver's face and the background are well-preserved. The transitions between focus regions are smooth and free of noticeable artifacts, resulting in a natural-looking image. ZMFF demonstrates competitive performance, preserving sharpness in both the diver's face and the background. However, slight artifacts and less refined transitions between focus

regions make it less effective than the proposed method. Similarly, FusionDN and U2Fusion provide moderate results, balancing focus between the foreground and background but lacking the sharpness and clarity of the best-performing algorithms. PMGI maintains good detail in the background but struggles with sharpness in the foreground, leading to an imbalanced fusion result. GD performs adequately, but the diver's face appears softened, and overall sharpness is inconsistent. Finally, UUDFusion produces the weakest fusion result, with significant blurring in both focus areas, making it unsuitable for generating high-quality all-focus images. In summary, the proposed method achieves the most visually appealing and technically superior fusion result, while ZMFF serves as a strong

TABLE 1 The average metric values of different methods on Lytro dataset.

	Year	$Q_{AB/F}$	Q_{MI}	Q_{NCIE}	Q_{CB}	Q_P	Q_G
GD	2016	0.7034	3.8521	0.8139	0.6115	0.7466	0.6987
FusionDN	2020	0.6018	5.7908	0.8221	0.6008	0.6221	0.5952
PMGI	2020	0.3901	5.8641	0.8225	0.5656	0.4620	0.3857
U2Fusion	2022	0.6143	5.7765	0.8221	0.5682	0.6657	0.6093
ZMFF	2023	0.7087	6.6271	0.8271	0.7412	0.7853	0.7030
UUDFusion	2024	0.5107	4.8412	0.8178	0.5989	0.5630	0.5055
Proposed		0.7409	7.1960	0.8313	0.7504	0.8137	0.7385

Notes: Bold font indicates the optimal values.

alternative with slight limitations. Other algorithms exhibit varying levels of performance but fall short of achieving the balance and detail provided by the proposed method.

Figure 4 compares the fusion results of multiple algorithms on the Lytro-03 dataset. Each algorithm demonstrates varying capabilities in handling multi-focus image fusion, balancing sharpness, color fidelity, and detail preservation. These are the two input images with distinct focal regions. Source A focuses on the foreground, while Source B highlights the background. The goal of fusion algorithms is to combine these focal regions into a single, sharp image. The GD method struggles with detail preservation and produces a fused image that appears slightly blurred, especially around the edges of the child's face. The colors also seem less vibrant, which detracts from the overall quality. As a deep learning-based approach, FusionDN performs well in preserving details and maintaining sharpness. The child's face and the Cartoon portrait are both clear, with vivid colors. However, minor edge artifacts are noticeable, which slightly impacts the naturalness of the result. The PMGI approach achieves a good balance between sharpness and detail integration. However, it slightly lacks precision in integrating the finest details. The U2Fusion provides decent sharpness and color fidelity but occasionally fails to balance focus across regions. For example, the child's face is slightly less sharp compared to the background, resulting in a less seamless fusion. Some areas also become very dark, resulting in severe information loss. This ZMFF method exhibits noticeable limitations. The fused image lacks sharpness, and the details in both the foreground and background are not well-preserved. The colors are also muted, leading to an overall decrease in visual quality. The image produced by UUDFusion exhibits severe distortion and artifacts, with significant color information loss and poor fusion performance. The proposed method outperforms all others in this comparison. It successfully combines the sharpness and details of both the child's face and the gingerbread figure. The colors are vibrant and natural, with no visible artifacts or blurriness. The transitions between the foreground and background are smooth, creating a visually seamless result.

Figure 5 compares the fusion results of various algorithms on the Lytro-04 dataset, focusing on how well the algorithms preserve details, manage focus regions, and maintain color fidelity. Figure 5a focuses on the foreground, specifically the man's face and sunglasses, while the background is blurred. Figure 5b focuses on the background (the person and chair) but blurs the foreground. Figures 5c-i represent

the fusion results of different algorithms. The GD exhibits moderate sharpness in both the foreground and background. However, some details in the man's sunglasses and the background elements appear slightly smoothed, reducing overall clarity. The color representation is acceptable but lacks vibrancy compared to other methods. As a deep learning-based method, FusionDN achieves good sharpness and color fidelity. The man's face and sunglasses are well-preserved, and the background details are clear. However, subtle edge artifacts are noticeable around the foreground and background transitions, slightly affecting the fusion quality. The PMGI fails to preserve sufficient details in both the foreground and background. The man's sunglasses appear blurred, and the background lacks clarity. The overall image looks less vibrant and exhibits significant information loss, making it one of the weaker methods in this comparison. The overall quality of the fused image is subpar. The U2Fusion method achieves decent fusion but struggles with focus balance. The foreground (sunglasses and face) is slightly less sharp, while the background elements are relatively clear. The ZMFF method produces relatively good fusion results, but the brightness and sharpness of the image still need improvement. The UUDFusion generates noticeable artifacts and distortions, particularly in the background. The details in the foreground (the man's face and sunglasses) are not clear, with significant color distortion, resulting in poor fusion performance. The proposed method demonstrates the best performance among the algorithms. Both the foreground (man's face and sunglasses) and the background (chair and person) are sharp, with vibrant and natural colors. The transitions between the focused regions are smooth, and there are no visible artifacts or distortions. It successfully preserves all critical details, making it the most effective fusion approach in this comparison.

Table 1 shows the average metric values of different algorithms in the simulation experiments on 20 data sets from the Lytro dataset. Table 1 compares the performance of various algorithms on the Lytro dataset across six evaluation metrics: $Q_{AB/F}$, Q_{MI} , Q_{NCIE} , Q_{CB} , Q_P and Q_G . Each metric highlights different aspects of image fusion quality. Among the listed methods, the proposed method demonstrates the best overall performance. It achieves the highest scores in all metrics, such as $Q_{AB/F} = 0.7409$, $Q_{MI} = 7.1960$, $Q_{NCIE} = 0.8313$, $Q_{CB} = 0.7504$, $Q_P = 0.8137$ and $Q_G = 0.7385$. These results suggest that the proposed method is highly robust and effective, delivering superior results across multiple dimensions of evaluation. ZMFF also shows competitive performance. The FusionDN and

U2Fusion maintain balanced performance but fail to excel in any particular metric. UUDFusion performs consistently lower across all metrics, indicating limited effectiveness compared to other algorithms. In summary, the proposed method clearly outperforms all other algorithms, providing the best fusion performance. The ZMFF and GD are strong competitors in specific metrics, but their inconsistencies in other areas limit their overall efficacy. This comparison highlights the superiority of the proposed method for image fusion tasks on the Lytro dataset. These results are consistent with the objective evaluation shown in Figures 2–5.

5 Conclusion

In this paper, a novel multi-focus image fusion method based on pulse coupled neural network and WSEML in DTCWT domain is proposed. The source images are decomposed by DTCWT into low- and high-frequency components, respectively; then the AG and pulse coupled neural network-based fusion rule is used to process the low-frequency components, and the WSEML-based fusion rule is used to process the high-frequency components. The experimental results show that our method achieves better performance in terms of both visual quality and objective evaluation metrics compared to several state-of-the-art image fusion algorithms. The proposed approach effectively preserves important details and edges while reducing artifacts and noise, leading to more accurate and reliable fused images. Future work will focus on further exploring its potential in other image processing tasks.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

References

- Bai X, Zhang Y, Zhou F, Xue B. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Inf Fusion* (2015) 22:105–18. doi:10.1016/j.inffus.2014.05.003
- Li H, Shen T, Zhang Z, Zhu X, Song X. EDMF: a new benchmark for multi-focus images with the challenge of exposure difference. *Sensors* (2024) 24:7287. doi:10.3390/s24227287
- Zhang Y, Bai X, Wang T. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure. *Inf Fusion* (2017) 35:81–101. doi:10.1016/j.inffus.2016.09.006
- Li X, Zhou F, Tan H, Chen Y, Zuo W. Multi-focus image fusion based on nonsubsampling contourlet transform and residual removal. *Signal Processing* (2021) 184:108062. doi:10.1016/j.sigpro.2021.108062
- Liu Y, Qi Z, Cheng J, Chen X. Rethinking the effectiveness of objective evaluation metrics in multi-focus image fusion: a statistic-based approach. *IEEE Trans Pattern Anal Machine Intelligence* (2024) 46:5806–19. doi:10.1109/tpami.2024.3367905
- Zheng K, Cheng J, Liu Y. Unfolding coupled convolutional sparse representation for multi-focus image fusion. *Inf Fusion* (2025) 118:102974. doi:10.1016/j.inffus.2025.102974
- Li X, Li X, Ye T, Cheng X (2024). Bridging the gap between multi-focus and multi-modal: a focused integration framework for multi-modal image fusion. In *Proceedings of the 2024 IEEE winter conference on applications of computer vision (WACV 2024), waikoloa, HI, United states*, January 4–January 8, 2024, 4–8.
- Li X, Li X, Cheng X, Wang M, Tan H. MCDFFD: multifocus image fusion based on multiscale cross-difference and focus detection. *IEEE Sensors J* (2023) 23:30913–26. doi:10.1109/jsen.2023.3330871
- Zhou Z, Li S, Wang B. Multi-scale weighted gradient-based fusion for multi-focus images. *Inf Fusion* (2014) 20:60–72. doi:10.1016/j.inffus.2013.11.005
- Liu Y, Liu S, Wang Z. Multi-focus image fusion with dense SIFT. *Inf Fusion* (2015) 23:139–55. doi:10.1016/j.inffus.2014.05.004

Author contributions

YJ: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Supervision, Writing–original draft, Writing–review and editing. TM: Data curation, Formal Analysis, Funding acquisition, Methodology, Software, Supervision, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Tianshan Talent Training Project–Xinjiang Science and Technology Innovation Team Program (2023TSYCTD).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

11. Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process* (2013) 22:2864–75. doi:10.1109/TIP.2013.2244222
12. Wang W, Deng L, Vivone G. A general image fusion framework using multi-task semi-supervised learning. *Inf Fusion* (2024) 108:102414. doi:10.1016/j.inffus.2024.102414
13. Wang W, Deng L, Ran R, Vivone G. A general paradigm with detail-preserving conditional invertible network for image fusion. *Int J Comput Vis* (2024) 132:1029–54. doi:10.1007/s11263-023-01924-5
14. Wu X, Cao Z, Huang T, Deng L, Chanussot J, Vivone G. Fully-connected transformer for multi-source image fusion. *IEEE Trans Pattern Anal Machine Intelligence* (2025) 47:2071–88. doi:10.1109/tpami.2024.3523364
15. Li J, Li X, Li X, Han D, Tan H, Hou Z, et al. Multi-focus image fusion based on multiscale fuzzy quality assessment. *Digital Signal Process.* (2024) 153:104592. doi:10.1016/j.dsp.2024.104592
16. Wan H, Tang X, Zhu Z, Xiao B, Li W. Multi-focus color image fusion based on quaternion multi-scale singular value decomposition. *Front Neurobot* (2021) 15:695960. doi:10.3389/fnbot.2021.695960
17. Li X, Li X, Tan H, Li J (2024). SAMF: small-area-aware multi-focus image fusion for object detection. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP), Seoul, Korea*, pp. 3845–9.
18. Basu S, Singhal S, Singh D. Multi-focus image fusion: a systematic literature review. *SN COMPUT SCI* (2025) 6:150. doi:10.1007/s42979-025-03678-y
19. Li J, Chen L, An D, Feng D, Song Y. A novel method for CSAR multi-focus image fusion. *Remote Sens.* (2024) 16:2797. doi:10.3390/rs16152797
20. Zhang X. Deep learning-based multi-focus image fusion: a survey and a comparative study. *IEEE Trans Pattern Anal Mach Intell* (2022) 44:4819–38. doi:10.1109/tpami.2021.3078906
21. Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf Fusion* (2015) 24:147–64. doi:10.1016/j.inffus.2014.09.004
22. Giri A, Sagan V, Alifu H, Maiwulanjiang A, Sarkar S, Roy B, et al. A wavelet decomposition method for estimating soybean seed composition with hyperspectral data. *Remote Sens* (2024) 16:4594. doi:10.3390/rs16234594
23. Wang F, Chen T. A dual-tree-complex wavelet transform-based infrared and visible image fusion technique and its application in tunnel crack detection. *Appl Sci* (2024) 14:114. doi:10.3390/app14010114
24. Vivone G, Deng L, Deng S, Hong D, Jiang M, Li C, et al. Deep learning in remote sensing image fusion methods, protocols, data, and future perspectives. *IEEE Geosci Remote Sensing Mag* (2024) 2–43. doi:10.1109/mgrs.2024.3495516
25. Wang G, Li J, Tan H, Li X. Fusion of full-field optical angiography images via gradient feature detection. *Front Phys* (2024) 12:1397732. doi:10.3389/fphy.2024.1397732
26. Zhu Z, Zheng M, Qi G, Wang D, Xiang Y. A phase congruency and local laplacian energy based multi-modality medical image fusion method in NSCT domain. *IEEE Access* (2019) 7:20811–24. doi:10.1109/access.2019.2898111
27. Chen H, Wu Z, Sun Z, Yang N, Menhas M, Ahmad B. CsdIFusion: an infrared and visible image fusion method based on LatLRR-NSST and compensated saliency detection. *J Indian Soc Remote Sens* (2025) 53:117–34. doi:10.1007/s12524-024-024-01987-y
28. Ramakrishna Y, Agrawal R. Pan-sharpening through weighted total generalized variation driven spatial prior and shearlet transform regularization. *J Indian Soc Remote Sens* (2024) 53:681–91. doi:10.1007/s12524-024-02006-w
29. Paul S, Sevcenco I, Agathoklis P. Multi-exposure and multi-focus image fusion in gradient domain. *J Circuits Syst Comput* (2016) 25:1650123. doi:10.1142/s0218126616501231
30. Mohan CR, Chouhan K, Rout RK, Sahoo KS, Jhanjhi NZ, Ibrahim AO, et al. Improved procedure for multi-focus images using image fusion with qshiftN DTCWT and MPCA in Laplacian pyramid domain. *Appl Sci* (2022) 12:9495. doi:10.3390/app12199495
31. Mohan CR, Kiran S, Vasudeva. Improved procedure for multi-focus image quality enhancement using image fusion with rules of texture energy measures in the hybrid wavelet domain. *Appl Sci* (2023) 13:2138. doi:10.3390/app13042138
32. Lu J, Tan K, Li Z, Chen J, Ran Q, Wang H. Multi-focus image fusion using residual removal and fractional order differentiation focus measure. *Signal Image Video Process.* (2024) 18:3395–410. doi:10.1007/s11760-024-03002-w
33. Xie Q, Yi B. Multi-focus image fusion based on SML and PCNN in NSCT domain. *Computer Sci* (2017) 44:266–9.
34. Wu Z, Zhang K, Xuan H, Yuan X, Zhao C. Divide-and-conquer model based on wavelet domain for multi-focus image fusion. *Signal Processing: Image Commun* (2023) 116:116982. doi:10.1016/j.image.2023.116982
35. Liu Y, Shi Y, Mu F, Cheng J, Chen X. Glioma segmentation-oriented multi-modal MR image fusion with adversarial learning. *IEEE/CAA J Automatica Sinica* (2022) 9:1528–31. doi:10.1109/jas.2022.105770
36. Zhang K, Wu Z, Yuan X, Zhao C. CFNet: context fusion network for multi-focus images. *The Institution of Engineering and Technology.* 16 (2022) 499–508.
37. Shi Y, Liu Y, Cheng J, Wang Z, Chen X. VDMUFusion: a versatile diffusion model-based unsupervised framework for image fusion. *IEEE Trans Image Process* (2025) 34:441–54. doi:10.1109/tip.2024.3512365
38. Zhu Z, He X, Qi G, Li Y, Cong B, Liu Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf Fusion* (2023) 91:376–87. doi:10.1016/j.inffus.2022.10.022
39. Zhu Z, Wang Z, Qi G, Mazur N, Yang P, Liu Y. Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction. *Pattern Recognition* (2024) 153:110553. doi:10.1016/j.patcog.2024.110553
40. Wu Z, Sun C, Xuan H, Zhang K, Yan Y. Divide-and-conquer completion network for video inpainting. *IEEE Trans Circuits Syst Video Technology* (2023) 33:2753–66. doi:10.1109/tcsvt.2022.3225911
41. Wu Z, Sun C, Xuan H, Liu G, Yan Y. WaveFormer: wavelet transformer for noise-robust video inpainting. *AAAI Conf Artif Intelligence* (2024) 38:6180–8. doi:10.1609/aaai.v38i6.28435
42. Wu Z, Chen K, Li K, Fan H, Yang Y. BVINet: unlocking blind video inpainting with zero annotations. *arXiv* (2025). doi:10.48550/arXiv.2502.01181
43. Chen K, Wu Z, Hou W, Li K, Fan H, Yang Y. Prompt-aware controllable shadow removal. *arXiv* (2025). doi:10.48550/arXiv.2501.15043
44. Wang F, Guo D, Li K, Zhong Z, Wang M (2024). Frequency decoupling for motion magnification via multi-level isomorphic architecture. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), Seattle, WA*, 18984–94.
45. Li J, Zheng K, Gao L, Han Z, Li Z, Chanussot J. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Trans Geosci Remote Sensing* (2025) 63:1–18. doi:10.1109/tgrs.2025.3531646
46. Li J, Zheng K, Gao L, Ni L, Huang M, Chanussot J. Model-informed multistage unsupervised network for hyperspectral_enspensp_super-resolution. *IEEE Trans Geosci Remote Sensing* (2024) 62:5516117. doi:10.1109/TGRS.2024.3391014
47. Li S, Huang S. AFA-Mamba: adaptive feature alignment with global-local mamba for hyperspectral and LiDAR data classification. *Remote Sensing* (2024) 16:4050. doi:10.3390/rs16214050
48. Ouyang Y, Zhai H, Hu H, Li X, Zeng Z. FusionGCN: multi-focus image fusion using superpixel features generation GCN and pixel-level feature reconstruction CNN. *Expert Syst Appl* (2025) 262:125665. doi:10.1016/j.eswa.2024.125665
49. Liu Y, Chen X, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. *Inf Fusion* (2017) 36:191–207. doi:10.1016/j.inffus.2016.12.001
50. Feng S, Wu C, Lin C, Huang M. RADFNet: an infrared and visible image fusion framework based on distributed network. *Front Plant Sci* (2023) 13:1056711. doi:10.3389/fpls.2022.1056711
51. Li H, Ma H, Cheng C, Shen Z, Song X, Wu X. Conti-Fuse: a novel continuous decomposition-based fusion framework for infrared and visible images. *Inf Fusion* (2025) 117:102839. doi:10.1016/j.inffus.2024.102839
52. Zhang Y, Liu Y, Sun P, Yan H, Zhao X, Zhang L. IFCNN: a general image fusion framework based on convolutional neural network. *Inf Fusion* (2020) 54:99–118. doi:10.1016/j.inffus.2019.07.011
53. Gao X, Liu S. BCMFIFuse: a bilateral cross-modal feature interaction-based network for infrared and visible image fusion. *Remote Sens* (2024) 16:3136. doi:10.3390/rs16173136
54. Selesnick I, Baraniuk R, Kingsbury N. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag* (2005) 22:123–51. doi:10.1109/msp.2005.1550194
55. Jiang J, Zhai H, Yang Y, Xiao X, Wang X. Multi-focus image fusion method based on adaptive weighting and interactive information modulation. *Multimedia Syst* (2024) 30:290. doi:10.1007/s00530-024-01506-6
56. Vishwanatha JS, Srinivasa Pai P, D'Mello G, Sampath Kumar L, Bairy R, Nagaral M, et al. Image-processing-based model for surface roughness evaluation in titanium based alloys using dual tree complex wavelet transform and radial basis function neural networks. *Sci Rep* (2024) 14:28261. doi:10.1038/s41598-024-75194-7
57. Ghosh T, Jayanthi N. Multimodal fusion of different medical image modalities using optimised hybrid network. *Int J Ad Hoc Ubiquitous Comput* (2025) 48:19–33. doi:10.1504/ijahuc.2025.143546
58. Shreyamsha Kumar BK. Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video Process.* (2015) 9:1193–204. doi:10.1007/s11760-013-0556-9
59. Qu X, Yan J, Xiao H, Zhu ZQ. Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain. *Acta Autom* (2008) 34:1508–14. doi:10.1016/s1874-1029(08)60174-3
60. Yin M, Liu X, Liu Y, Chen X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Trans Instrumentation Meas* (2019) 68:49–64. doi:10.1109/tim.2018.2838778

61. Nejati M, Samavi S, Shirani S. Multi-focus image fusion using dictionary-based sparse representation. *Inf Fusion* (2015) 25:72–84. doi:10.1016/j.inffus.2014.10.004
62. Xu H, Ma J, Le Z, Jiang J, Guo X. (2020). FusionDN: a unified densely connected network for image fusion. *Proc Thirty-Fourth AAAI Conf Artif Intelligence (Aaai)*. 34. 12484–91. doi:10.1609/aaai.v34i07.6936
63. Zhang H, Xu H, Xiao Y, Guo X, Ma J. (2020). Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity. *Proc AAAI Conf Artif Intelligence*. 34. 12797–804. doi:10.1609/aaai.v34i07.6975
64. Xu H, Ma J, Jiang J, Guo X, Ling H. U2Fusion: a unified unsupervised image fusion network. *IEEE Trans Pattern Anal Mach Intell* (2022) 44:502–18. doi:10.1109/tpami.2020.3012548
65. Hu X, Jiang J, Liu X, Ma J. ZMFF: zero-shot multi-focus image fusion. *Inf Fusion* (2023) 92:127–38. doi:10.1016/j.inffus.2022.11.014
66. Wang X, Fang L, Zhao J, Pan Z, Li H, Li Y. UUD-Fusion: an unsupervised universal image fusion approach via generative diffusion model. *Computer Vis Image Understanding* (2024) 249:104218. doi:10.1016/j.cviu.2024.104218
67. Liu Z, Blasch E, Xue Z, Zhao J, Laganieri R, Wu W. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. *IEEE Trans Pattern Anal Mach Intell* (2012) 34:94–109. doi:10.1109/tpami.2011.109