Review

A Survey of Multi-Focus Image Fusion Methods

Youyong Zhou [1], Lingjie Yu [1], Chao Zhi [1], Chuwen Huang [1], Shuai Wang [1], Mengqiu Zhu [1], Zhenxia Ke [1], Zhongyuan Gao [1], Yuming Zhang [2,*] and Sida Fu [3,*]

1   School of Textile Science and Engineering, Xi'an Polytechnic University, Xi'an 710048, China; zhouyouyong@stu.xpu.edu.cn (Y.Z.); lingjie.yu@xpu.edu.cn (L.Y.); zhichao@xpu.edu.cn (C.Z.); cptbtptpr7bo@163.com (C.H.); wangshuai_xpu@163.com (S.W.); zhumengqiu@stu.xpu.edu.cn (M.Z.); kezhenxia@stu.xpu.edu.cn (Z.K.); gaozhongyuan@stu.xpu.edu.cn (Z.G.)
2   School of Textile, Apparel & Art Design, Shaoxing University Yuanpei College, Shaoxing 312000, China
3   China-Australia Institute for Advanced Materials and Manufacturing, Jiaxing University, Jiaxing 314001, China
*   Correspondence: lunwenzhuanyou@126.com (Y.Z.); sidafu@zjxu.edu.cn (S.F.)

Abstract: As an important branch in the field of image fusion, the multi-focus image fusion technique can effectively solve the problem of optical lens depth of field, making two or more partially focused images fuse into a fully focused image. In this paper, the methods based on boundary segmentation was put forward as a group of image fusion method. Thus, a novel classification method of image fusion algorithms is proposed: transform domain methods, boundary segmentation methods, deep learning methods, and combination fusion methods. In addition, the subjective and objective evaluation standards are listed, and eight common objective evaluation indicators are described in detail. On the basis of lots of literature, this paper compares and summarizes various representative methods. At the end of this paper, some main limitations in current research are discussed, and the future development of multi-focus image fusion is prospected.

Keywords: image fusion; multi-focus image; fusion method; evaluation indicators

1. Introduction

Image fusion is a process of generating an image superior to the original image and using a special application based on the study of multiple image features in the same scene by using redundant and complementary information among image data [1]. Using specific algorithms to extract useful feature information from two or more images, image fusion technique can generate a new image with more comprehensive and accurate details. According to the types of input source images, the image fusion can be divided into remote sensing image fusion, medical image fusion, multi-focus image fusion, multi-exposure image fusion, infrared and visible image fusion, etc. [2]. Image fusion technology has been developed for more than 40 years, with more and more research methods and applications rising. Among them, multi-focus image fusion technology has a very broad application prospect in digital photography, computer vision, target tracking, and monitoring, microscopic imaging, and other fields [3,4].

The so-called multi-focus problem can be explained as follows: due to the limited focus range of visible-light imaging systems, it is difficult to clearly capture all objects in the same scene [3,5]. As shown in Figure 1, a point in the scene is projected onto a single point in the focal plane to form a focused image. However, if the sensor plane does not coincide with the focal plane, the image formed on the sensor plane will be a fuzzy disk with a diameter of 2R, which could be called a defocused image.
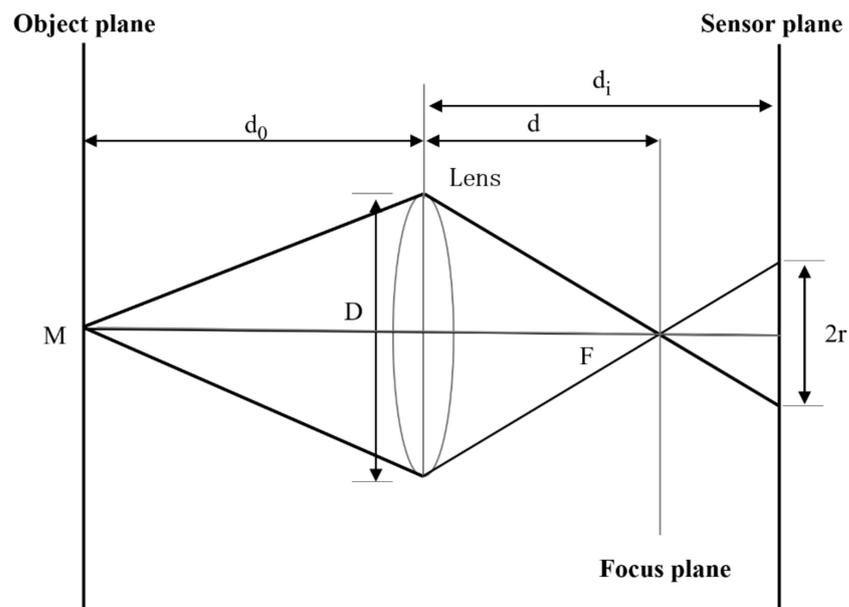
Figure 1. The basic principle of optical imaging characteristics.

According to the above principles, multi-focus images can be divided into two parts, namely the focusing region and the defocusing region. Objects are clearly sharp in the focusing area, while relatively blurred in the defocusing region. Figure 2 shows 14 groups of images, including both gray and color. Each group of images were captured for the same scene but at a different focusing position, forming a multi-focus result, as exhibited in Figure 2.

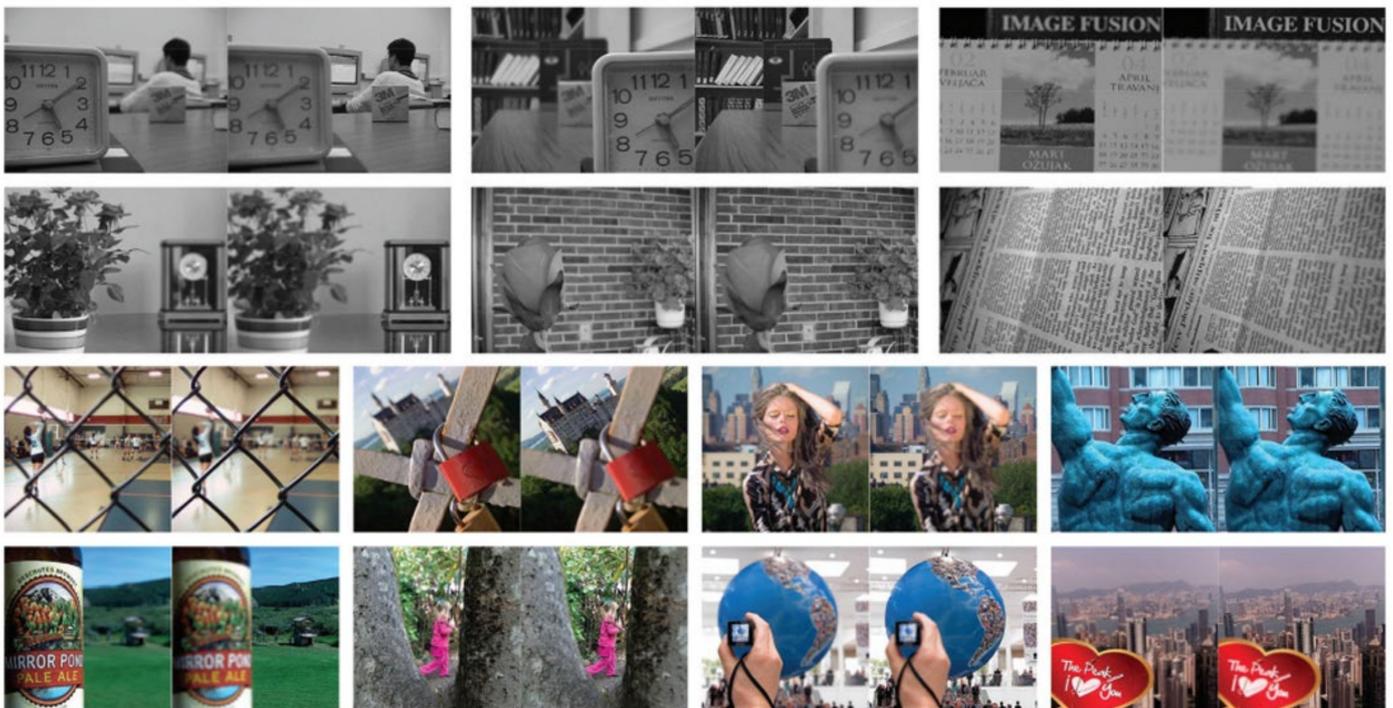**Figure 2.** Multi-focus images in the same scene.

Traditional classification methods include the spatial domain method and the transform domain method [6,7]. With the soaring new multi-focus image fusion methods, it is difficult for the existing classification methods to accurately position all image fusion algorithms. Therefore, the existing multi-focus fusion methods cannot reasonably clas-

sify and summarize the pixel level fusion methods. For instance, the pixel level image fusion method can be simultaneously divided into spatial domain and transform domain according to the choice of the domain [8]. Therefore, this paper innovatively proposes the fusion method type based on boundary segmentation, and classifies the pixel level fusion methods into the fusion method of boundary segmentation. In this paper, the existing multi-focus image fusion methods are reviewed, sorted, and classified, and eight commonly used objective evaluation methods are summarized. Based on a large number of literature, the typical algorithm, fusion process, and key technologies of multi-focus image fusion are discussed, and the fusion results and fusion efficiency are compared and summarized. The image fusion can be divided into four categories: transform domain method, boundary segmentation method, deep learning method, and combination fusion method. The applicability of various methods is summarized. Finally, we analyzed and discussed the challenges faced by this field and proposed the solutions, and the future development of multi-focus image fusion technology has prospected.

The first part of this paper is the introduction, which introduces the concepts of multi-focus image fusion, and summarizes the content of this paper; The second part is the fusion method and analysis, which analyzes and classifies a variety of multi-focus fusion methods; The third part is the evaluation indicators, which introduces the commonly used subjective evaluation and objective evaluation; The fourth part is the limitations, and gives the corresponding solutions according to the common fusion problems; The fifth part is the conclusion, which analyzes the application and development of multi-focus fusion.

2. Fusion Methods and Analysis

As shown in Figure 3, image fusion can be divided into pixel-level image fusion, feature-level image fusion, and decision-level image fusion according to the information representation layer [6]. For the image source classification, image fusion can be divided into remote sensing image fusion, medical image fusion, multi-focus image fusion, multi-exposure image fusion and infrared, and visible image fusion [7]. Referring to the fusion method, image fusion can be divided into spatial domain image fusion and transform domain image fusion [8].
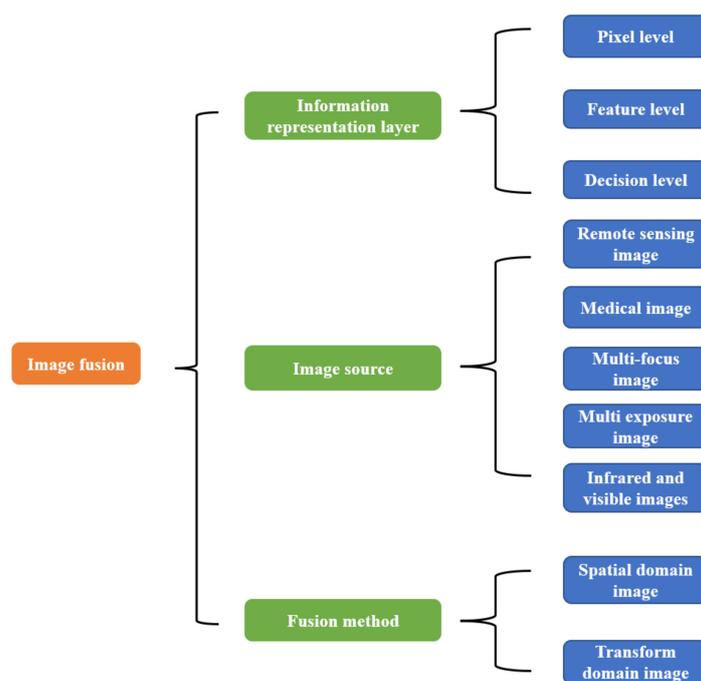

Figure 3. Image fusion classification.

This paper proposes a new classification method, that is boundary segmentation method, dividing the current mainstream multi-focus fusion methods into four categories: transform domain method, boundary segmentation method, deep learning method, and combinatorial fusion method. Table 1 lists the classification method and the current mainstream algorithms.

Table 1. Classification and mainstream algorithms of multi-focus image fusion.

Method	Mainstream Algorithm
Transformation domain	Laplacian pyramid (LP) [9]; gradient pyramid (GD) [10]; contrast pyramid (CP) [11]; region mosaicking on Laplacian pyramids (RMLP) [12]; fast discrete curvelet transform (FDCT) [5]; dual-tree complex wavelet transform (DT-CWT) [13]; nonsubsampled contourlet transform (NSCT) [14]; nonsubsampled shearlet transform (NSST) [15]; cross sparse representation (CSP) [16]; independent component analysis (ICA) [17]; discrete cosine transform (DCT) [18]
Boundary segmentation	adaptive region-segmentation (ARS) [19]; morphology-based focus measure (MBFM) [20]; content adaptive blurring (CAB) [21]; robust principal component analysis (RPCA) [22]; Markov random field (MRF) [23]; optimal defocus estimation (ODE) [24]
Deep learning	convolutional neural networks (CNN) [25]; CNN based [26]; PCNN-Pulse Coupled Neural Network (PCNN) [27]; PCNN based [28]; Generative Adversarial Networks (GAN) [29]
Combination	CTD+SR [30]; NSCT+SR [31]; SF-PAPCNN+NSST+ ISML [32]; CNN+SR [33]

2.1. Multi-Focus Fusion Methods Based on Transform Domain

Most traditional fusion methods in multi-focus image fusion are based on the transform domain [34]. As shown in Figure 4, the transformation domain method mainly operates the decomposition coefficient after image transformation, which mainly includes three fusion stages: image transformation, coefficient decomposition, and inverse transformation reconstruction. Firstly, the source image is transformed into the transform domain by an image decomposition algorithm; then various fusion strategies are used to fuse different coefficients; finally, the corresponding inverse transformation of the fusion coefficient is processed to obtain the final fusion image. The more layers of decomposition are used, the more detailed the information will be however, the efficiency will decrease. Therefore, the fusion effect will be greatly improved by properly handling the relationship between decomposition layers and execution efficiency. In terms of the different transformations, the transform domain method is further divided into the method based on multi-scale decomposition (MSD), the method based on sparse representation (SR), and the method based on gradient domain (GD) in this paper.
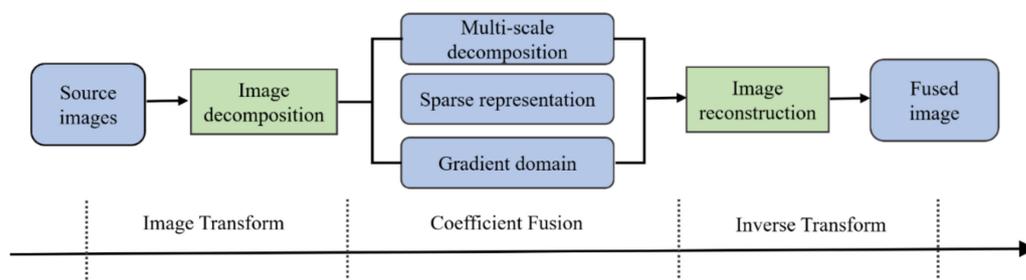

Figure 4. The general schematic diagram of transform domain methods.

2.1.1. Multi-Scale Decomposition (MSD)-Based Methods

Multi-scale decomposition (MSD)-based methods have always been the mainstream of image fusion. Pyramid transform and wavelet transform are the first-used MSD methods in image fusion. In addition, the multi-scale geometric analysis method (the improved wavelet transform method) also achieves an excellent fusion effect in multi-focus image fusion, superior to pyramid transform and wavelet transform in feature representation.

(1) Pyramid transformation

The Laplacian pyramid is the earliest multiscale decomposition method [9]. In this method, the absolute value of the decomposition coefficient is measured by its activity level, and the fusion coefficient is obtained by the choosing-Max rule. The greater the absolute value of the coefficient is, the more information it contains. In 2018, Sun et al. [12] proposed a Region Mosaic (RMLP) method based on The Laplace pyramid to fuse microscopically captured multi-focus images. The method firstly used The Laplacian operator to measure the focus level of multi-focus images. Then, a density-based region growth algorithm was used to segment the focused region mask of each image. Finally, the mask was decomposed into a mask pyramid to supervise the regional stitching of the Laplacian pyramid.

Due to the different forms of the tower structure, pyramid transformation can be divided into gradient pyramid [10], contrast pyramid [11], morphological pyramid [35], etc. The fusion method based on pyramid transformation has the advantages of high fusion efficiency while retaining sufficient original information. However, the decomposition method and the decomposition layer number have a great influence on the final result. The greater number of decomposition layers, the more blurred the fusion image boundary would be.

(2) The wavelet transforms

The wavelet transform can decompose the original image into high frequency coefficient and low frequency coefficient. The high frequency coefficient includes vertical, horizontal, and diagonal information. The fusion effect of wavelet transform is better than that of pyramid transform. However, the wavelet transform is not displacement invariant for the feature representation; thus, the fusion effect is not satisfactory for the image with poor registration. To solve this problem, many improved wavelet transform methods are proposed. Yang et al. [5] introduced a multi-focus image fusion method based on fast discrete curve-wave transform (FDCT), which solved the problem of block effect in texture selection and spatial fusion. Yu et al. [13] extracted the six-dimensional feature vectors of the source image using the dual-tree complex wavelet transform (DT-CWT) coefficient sub-bands and then projected them onto the class tags by training a two-class (focused and unfocused) support vector machine (SVM).

The common methods based on wavelet transform include contour wave [36], shear wave [37], non-subsampled contour wave transform (NSCT) [14], and non-subsampled shear wave transform (NSST) [15], etc. NSCT method and NSST method are two widely used methods at present. Figure 5 shows the image decomposition process of NSCT.
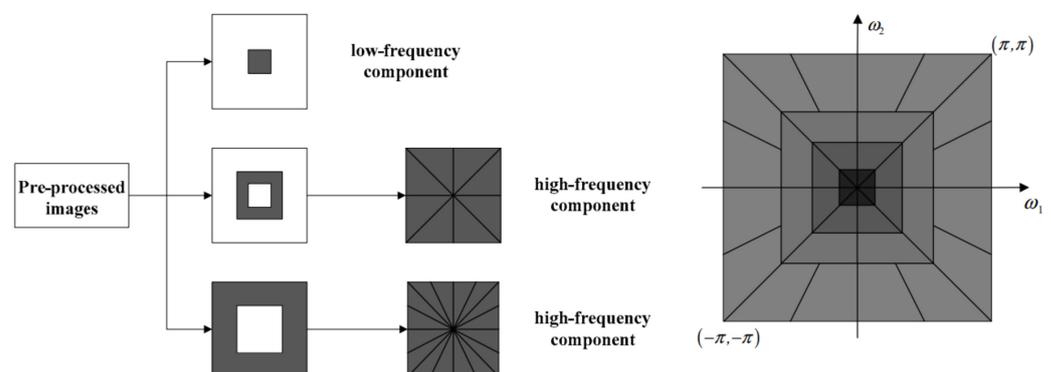

Figure 5. Image decomposition process of NSCT [38]. Reprinted/adapted with permission from Ref. [38]. 2019, MDPI.

In general, the advantage of the MSD method lies in extracting more accurate feature information and having a better fusion effect. However, the decomposition information is too much, leading to a large amount of calculation.

2.1.2. Sparse Representation (SR)-Based Methods

Sparse representation is a new image fusion method. By processing the natural sparsity of image signals, a signal is approximately represented as a linear combination of several atoms in a redundant dictionary. Atomic libraries of sparse representations are provided by over-complete dictionaries. By selecting some atoms in the over-complete dictionary and using a linear combination to reconstruct the image, the dependence between data dimensions and feature vectors can be reduced. Figure 6 shows a multi-focus image fusion framework based on sparse representation.
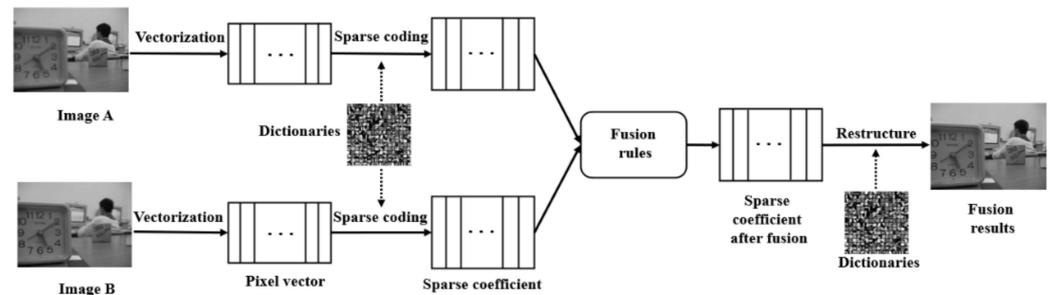

Figure 6. A multi-focus image fusion framework based on sparse representation. Image A and Image B are tow images captured of the same scene; The right area of Image A is focused while the left area of Image B is focused.

In 2010, Yang et al. [39] introduced SR into multi-focus image fusion for the first time. In this method, sliding window technology was used to segment each source image into multiple overlapping small blocks, and orthogonal matching pursuit (OMP) algorithm was used to perform sparse decomposition for each small block. The sparse coefficient vectors after fusion were obtained by using the maximum selection fusion rule. Subsequently, a variety of improved algorithms based on sparse representation follow. Ma et al. [40] obtained an adaptive dictionary based on rough k-means singular value decomposition. Then the fixed dictionary was combined with the adaptive dictionary to obtain the joint dictionary. The final joint dictionary was used to sparsely encode the source image to separate the complementary and redundant components. In addition, there are also cross sparse representation [16], group sparse representation [41], K-SVD [42], and other methods.

Sparse representation can better solve the problem of fused image noise. However, the processing effect of image details (edge, texture, etc.) is not ideal and easy to blur. In addition, the method has high complexity, low computational efficiency, and poor real-time performance.

2.1.3. Gradient Domain (GD)-Based Methods

The gradient domain (GD)-based method is to fuse the gradient representation of the source image, limiting the gradient of the fused image within a particular threshold. Therefore, it is crucial to obtain gradient information of the image for this method. Paul et al. [43] input the gradient of the image component at each image pixel. They solved the Poisson equation at each resolution to achieve boundary continuity in the gradient domain. Wang et al. [44] proposed a gradient domain image fusion method based on the structure tensor, in which source images were stacked into a multi-valued image, and the structure tensor of each source image was calculated according to its gradient graph.

The method of gradient domain image fusion can improve the image's visual effect, retaining the details and structural information of the source image. This method can be applied not only to multi-focus image fusion but also to multi-exposure image fusion.

Despite the above fusion methods, many other multi-focus image fusion methods based on transform domain, such as independent component analysis (ICA) [17], high order singular value decomposition (HOSVD) [45], discrete cosine transform (DCT) [18], compressed sensing (CS) [46], cartoon-texture decomposition (CTD) [47] and other methods have been successfully applied to multi-focus image fusion.

2.2. Multi-Focus Fusion Methods Based on Boundary Segmentation

This paper proposes a new classification method named the boundary segmentation method. According to the spatial characteristics of the source image, this method mainly generates a weight map for each source image by processing the region of pixels. It then calculates the fused image by the weighted average method or maximum method. This method has high operation efficiency, and the fused image can retain the image information of each local section. However, due to the improper boundary segmentation, many algorithms based on boundary segmentation often lost the edge, contour, and other image details. Therefore, strengthening the extraction of image boundary can effectively improve the quality of fusion. In this paper, the boundary segmentation method is further divided into block-based fusion method, region-based fusion method, and pixel-based fusion method.

2.2.1. Block-Based Methods

The earliest block-based segmentation scheme is to divide the source image into several fixed-size blocks, obtain the fused block by using the threshold based adaptive fusion rule, and finally use the consistency test method to achieve the fused image. Due to the fixed block size, the boundary of the multi-focus fusion image is prone to the fuzzy phenomenon.

Zhang et al. [19] proposed a multi-focus image fusion method based on adaptive region segmentation, which decomposed pre-registered source images into approximate coefficients and detail coefficients using the Laplace pyramid transform. In order to avoid the defect of fixed block size, an adaptive differential evolution algorithm is designed to calculate the optimal block size. Figure 7 shows a fusion framework for this approach. This method can effectively reduce the noise with high computational efficiency. De et al. [20] adopted the quadtree structure realized adaptive segmentation to fuse multi-focus images. The varied block sizes were determined by each specific content, which effectively solved the problem of the block effect.
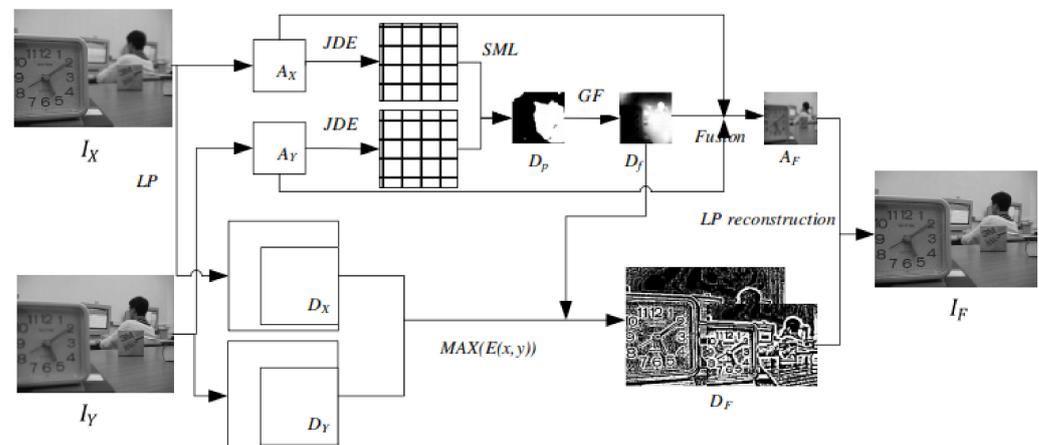

Figure 7. Multi-focus image fusion framework based on adaptive region segmentation [19]. Reprinted/adapted with permission from Ref. [19]. 2019, WSPC.

It can be seen from the above that the size of the block has a crucial influence on the final fusion effect, and it is prone to the phenomenon of fuzzy boundary. In addition, the compatibility between adjacent blocks needs to be considered.

2.2.2. Region-Based Methods

In order to improve the flexibility of source image segmentation, the region-based image fusion method came into being. The zone-based approach is similar to the block-based approach. The main difference is that the activity level is measured in each irregularly sized segmented area rather than a block. Li et al. [48] initially proposed a region-based

multi-focus image fusion method. Farid et al. [21] proposed a multi-focus image fusion method based on content adaptive fuzzy (CAB), in which the absolute difference between the original image and the CAB blurred image was used to generate the initial segmentation graph, and morphological operators and graph cutting techniques were used to improve the segmentation accuracy. Xiao et al. [49] proposed an adaptive initialization method for image depth estimation. The image depth was approximated by the iterative solution of the partial differential equation. The target image was adaptively divided into three regions: clear region, fuzzy region, and transition region. Finally, the multi-focus image fusion was realized by extracting the pixels of the clear region and fusing the pixels of the transition region.

2.2.3. Pixel-Based Methods

Beyond 2012, the pixel-based fusion method has become a popular direction of multi-focus image fusion. The main reason is that this method can obtain accurate pixel-weighted images. Most pixel-based approaches' core problem is obtaining a weight map for each source image. In these methods, the activity level measurement is first adopted, and then the focus values obtained from different source images are compared to generate pixel-level weight maps. Weight graphs are also known as decision graphs because multi-focus image fusion can be viewed as a classification problem in which each pixel's focus attributes (focusing and defocusing) are determined. In some methods, the source image is also divided into regions with different points (such as focus/defocus/border, texture/smooth). Different fusion rules are applied according to their characteristics.

Du et al. [50] proposed a more focused image fusion algorithm based on image segmentation; the task of decision graph detection was regarded as the image segmentation between the focus region and the defocus region, and the feature images on the boundaries of the focus region and the defocus region were obtained by the multi-scale convolutional neural network. Then the initial segmentation, morphological operation, and watershed processing were performed on the fused image to get the segmentation graph and decision graph. This method proved that the decision graph obtained by a multi-scale convolutional neural network is reliable and can produce high-quality fusion images. Ma et al. [51] proposed a dual-scale multi-focus image fusion algorithm based on an enhanced random walk. Using the complementary characteristics of dual-scale measurement can better align boundaries and solve noise problems, thus achieving a more robust fusion.

Other pixel-based fusion methods involve robust principal component analysis (RPCA) [22], random field [23], morphological filtering [24], etc. The pixel-level fusion method has fast fusion speed and real-time solid performance. However, operating pixels are susceptible to noise, which will reduce the signal-to-noise ratio and contrast of the image.

2.3. Multi-Focus Fusion Method Based on Deep Learning

Beyond 2014, deep learning methods have developed rapidly with special effects and excellent applications. In general, deep learning models mainly use the learnability of the network to extract features from multi-focus images and separate focused and defocused regions to generate full-focus fusion images. At present, convolutional neural networks (CNNs) are one of the popular models in this field. In addition, pulse-coupled neural networks (PCNN) and generative adversarial networks (GAN) also have many applications.

2.3.1. Convolutional Neural Network Model

The convolutional neural network is one of the most popular deep learning models. It can realize parallel computing and has high speed and high efficiency characteristics [52,53]. CNN is widely used in medical image analysis, remote sensing image analysis, noise signal analysis, and other fields. Javed Awan et al. [54] used a customized 14-layer convolutional neural network resnet-14 architecture to automatically detect and evaluate ACL injuries of athletes. Zhang et al. [55] proposed a new method more suitable for farmland vacancy segmentation, using the improved RESNET network as the backbone of signal transmission. Lopac et al. [56] proposed a method for the classification of noisy non-stationary time-series

signals based on Cohen's class of their time-frequency representations (TFRs) and deep learning algorithms. The proposed approach combining deep CNN architectures with Cohen's class TFRs yields high values of performance metrics and significantly improves the classification performance compared to the base model.

In the field of image fusion, CNN can learn the feature representation mechanism of different abstraction levels for source images, and it is trainable. CNN extracts the features of input images by learning filters to obtain different feature maps of each level. Each unit or coefficient in the feature maps is called a neuron. Generally, three calculation methods, filtering convolution activation function and pooling, are used to connect feature maps between adjacent levels [57]. The typical structure diagram of CNN is shown in Figure 8.
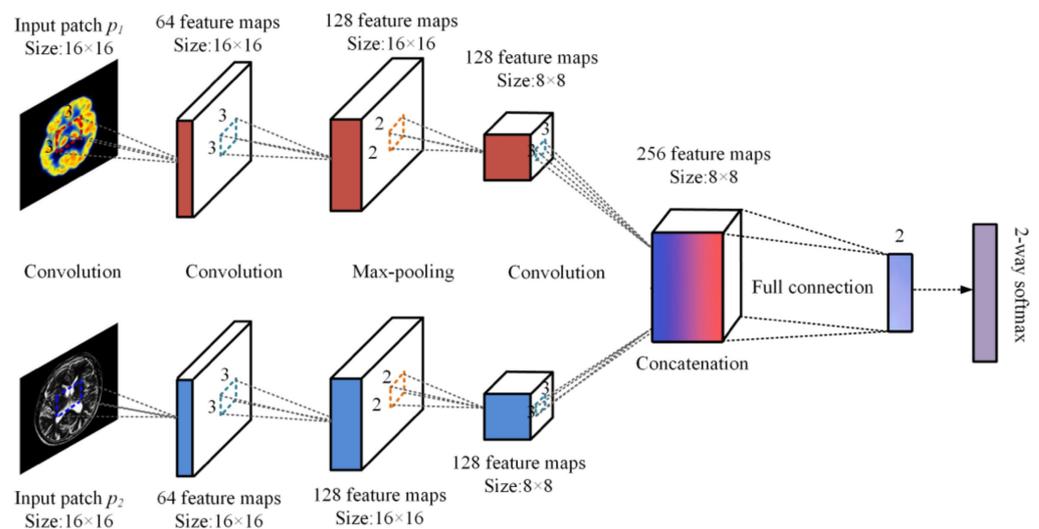

Figure 8. The typical structure diagram of CNN [58]. Reprinted/adapted with permission from Ref. [58]. 2020, MDPI.

In 2017, Liu et al. [25] successfully applied CNN to the field of image fusion for the first time. By extending the classification idea of artificial neural networks, a convolutional neural network is used instead of an artificial neural network to classify the pixels of source images. According to the mapping value, the score graph of the source image can be obtained. Then the decision graph can be built by the consistency verification of the score graph. Amin-Naji et al. [26] proposed a new CNN-based integrated learning approach to pursue data diversity to reduce the over-fitting problem. This fusion method based on CNN integration is better than the single CNN fusion method. Zhang et al. [59] proposed a full end-to-end convolution layer network model. This model chose the feature level fusion, canceled all pooling layers, adopt the strategy of full connection layer directly, gave up on the mapping of the source image pixel resolution, and judged the fusion result loss, thus achieving the purpose of the end-to-end output of fusion image. This method is more concise and effective and avoids the complicated follow-up processing problems of the CNN model.

The advantages of the multi-focus image fusion method based on CNN lie in the ability of layered learning features, more diversity of feature expression, strong discrimination ability, and better generalization performance. The disadvantage is that the training takes a long time, and there is no special training set, which usually requires particular training and image preprocessing.

2.3.2. Pulse Coupled Neural Network

The pulse coupled neural network model [60] is proposed based on the analysis of synchronous pulse oscillations of visual cortex neurons in cats and has been widely used in image fusion. Each neuron in the PCNN model corresponds to a pixel, whose definition is determined by the firing times of the neuron. The more the firing times are, the clearer the

corresponding pixel points would be. As shown in Figure 9, PCNN consists of three parts: feeding input field, modulation field, and firing subsystem. The stimulus is received by the feeding input field and fed back to the firing subsystem through modulation field.
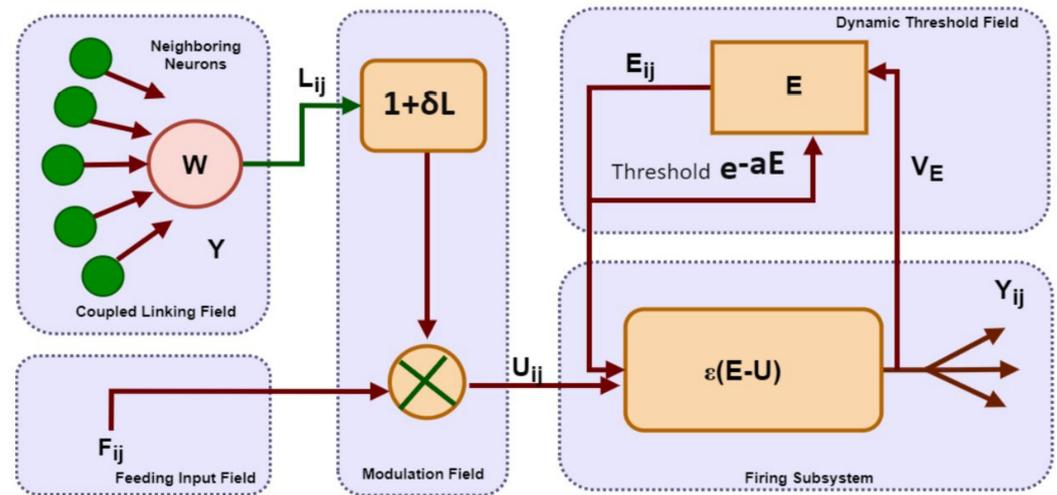

Figure 9. Typical structure diagram of PCNN [61]. Reprinted/adapted with permission from Ref. [61]. 2022, MDPI.

To make the fused image clearer, Wang et al. [27] proposed a multi-focus image fusion method based on PCNN and random walk. The technique used PCNN to measure the sharpness of the source image and constructed an initial fusion image. The random walk method was then used to improve the accuracy of fusion region detection, and the final fusion image was generated according to the probability calculated by a random walk. An improved PCNN method [28] was proposed to fuse the source image with the guide filter. The improved PCNN was excited by the intermediate fusion image to generate the fusion image. This method created fusion images several times and fused the fusion images with PCNN to make the fusion results more accurate.

The PCNN model can extract local details effectively and recognize image content well. However, the massive iterative calculation and configuration parameters make this method high coupling and time-consuming.

In addition to the above two models, a generative adversarial network (GAN) has also been used in multi-focus image fusion. Guo et al. [29] proposed a multi-focus image fusion method based on least square GAN. In the fusion process, the final fusion decision graph was obtained by binary segmentation and the refinement of the focus graph.

2.4. Combinatorial Fusion Method

As seen from the above, different fusion methods have different fusion characteristics. The combination fusion method combines two or more methods to take their strengths. For example, the transform domain method can be combined with the region segmentation method, which can extract more details and enhance the fusion efficiency.

Zhu et al. [30] proposed an image fusion scheme based on image cartoon texture decomposition and sparse representation. Aiming at the proposed sparse representation-based fusion method, they trained a dictionary with a strong representation ability to fuse the animation and texture components. Yang et al. [31] proposed a multi-focus image fusion framework based on non-subsampled contourlet transform form (NSCT) and sparse representation (SR). Li et al. [32] proposed a multi-focus image fusion algorithm based on spatial frequency-driven parameter adaptive pulse-coupled neural network (SF-PAPCNN) and an improved non-subsampled Shear-wave transform (NSST) domain summing modified Laplace Transform (ISML).

3. Evaluation Indicators

3.1. Subjective Evaluation

The subjective evaluation depends on the observer to evaluate the quality of the image, including the edge, whether the content is clear and whether it contains noise, etc. However, subjective evaluation is not applicable for the following reasons: observers need to possess relevant professional knowledge, and it is difficult to observe the details of the image with naked eyes. Equipment environment, such as lighting, display brightness, etc., affects the observer's judgment; In order to evaluate the accuracy, it usually needs to organize evaluation meetings, and is time-consuming and labor-consuming.

3.2. Objective Evaluation

Objective evaluation is to calculate the image quality through some algorithm, and the calculation results are used as evaluation criteria. Liu et al. [62] divided 12 popular image fusion evaluation indexes into four categories: evaluation indexes based on information theory, evaluation indexes based on image features, evaluation indexes based on image structure similarity, and evaluation indexes based on human perception. Many researchers welcome this classification method.

The commonly used objective evaluation indexes are shown in Table 2. When evaluating the quality of fused images, multiple evaluation indexes are often needed to be calculated. The better the evaluation effect, the better the corresponding fusion effect is.

Table 2. Common evaluation indicators.

Evaluation Indicators	Evaluation of the Effect
Information entropy (EN)	Reflect the amount of information the image carries; the larger the value, the richer the amount of information, the better the quality.
Mean gradient (AG)	Measure the clarity of the image, the greater the value, the higher the clarity, the better the quality.
MEAN	Measure the average brightness of the image, the average is moderate, the better the quality.
Standard deviation (STD)	Reflect the richness of image information, the larger the value, the more scattered the gray level distribution, the better the quality.
Mean square error (RMSE)	Reflect the spatial details of the image, the smaller the value, the smaller the difference, the better the quality.
Signal to noise (SNR)	To measure the proximity between fusion image and ideal image, the greater the value, the higher the similarity.
Normalized mutual information (QMI)	Reflects the amount of information retained by the source image in the fusion image, the larger the value, the better the quality.
Structural similarity (SSIM)	By comparing the structural similarity between source image and fusion image, the closer the value is to 1, the better the quality is.

4. Limitations

In the last ten years, multi-focus image fusion technology has been developed. However, there are still some urgent problems that need to be addressed.

(1)   Image registration

Most of the current fusion methods focus on feature extraction of source images, paying little attention to the image scene consistency, content deformation, and other registration problems. The actual source images are not as accurate as the experimental samples. Thus, the fusion effect would be greatly affected.

In our view, the multi-view registration method can be studied to address the above problem. To be specific, capturing images of similar objects or scene from multiple perspectives can obtain a better representation of the scanned object. The multi-view registration

can realized by various algorithms such as image mosaic, 3D model reconstruction from 2D image, etc.

(2) Fusion efficiency

Many scholars pursue the applicability and quality of fusion methods, but ignore the efficiency of fusion. However, we believe the fusion efficiency is of great value in practical application. The difficulty may be alleviated by immerging several fusion stages into a one-stop rapid stage, thus simplifying the sophisticated fusion process.

(3) Application scenarios

Although there are many multi-focus image fusion methods, most of them are studied and tested in public image libraries. We think it is helpful to collect and build image libraries in many specific industrial fields. Based on the specific image library, with the help of state-of-the-art mathematical theory or models, researchers can develop multi-focus image fusion methods suitable for actual application.

5. Conclusions

This paper describes four kinds of multi-focus image fusion methods: transform domain method, boundary segmentation method, deep learning method, and combinatorial fusion method. Each method is deeply classified and the advantages and disadvantages of each method are compared. For different scenarios, it is necessary to choose the appropriate method. In addition, the commonly used evaluation indicators are listed, and the objective evaluation is more accurate than the subjective evaluation, which takes less energy and time. Finally, the solution is discussed based on the analysis of the shortcomings of current applications and methods.

Multi-focus image fusion can effectively solve the depth of field problem in optical lens areas and has a wide application space in many fields, such as medicine, security, photography, etc. It has been successfully applied in the areas of microscopic imaging [63], image deblurring [64], focusing shape [65], and information forensics [66].

To sum up, multi-focus image fusion needs further development. Solving the problem of image registration will improve the universality of the method and expand the scope of fusion. In the pursuit of fusion quality, it is necessary to pursue time efficiency, taking real-time fusion as the ultimate goal.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. James, A.P.; Dasarathy, B.V. Medical image fusion: A survey of the state of the art. *Inf. Fusion* **2014**, *19*, 4–19. [CrossRef]
2. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [CrossRef]
3. Li, Q.; Yang, X.; Wu, W.; Liu, K.; Jeon, G. Multi-focus image fusion method for vision sensor systems via dictionary learning with guided filter. *Sensors* **2018**, *18*, 2143. [CrossRef] [PubMed]
4. Meher, B.; Agrawal, S.; Panda, R.; Abraham, A. A survey on region based image fusion methods. *Inf. Fusion* **2019**, *48*, 119–132. [CrossRef]
5. Yang, Y.; Tong, S.; Huang, S.; Lin, P.; Fang, Y. A Hybrid method for multi-focus image fusion based on fast discrete curvelet transform. *IEEE Access* **2017**, *5*, 14898–14913. [CrossRef]
6. Xiao, G.; Bavirisetti, D.P.; Liu, G.; Zhang, X. Decision-Level Image Fusion. In *Image Fusion*; Springer: Singapore, 2020; pp. 149–170. ISBN 978-981-15-4866-6.
7. Li, H.; Liu, L.; Huang, W.; Yue, C. An improved fusion algorithm for infrared and visible images based on multi-scale transform. *Infrared Phys. Technol.* **2016**, *74*, 28–37. [CrossRef]
8. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]
9. Burt, P.J.; Adelson, E.H. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **1987**, *31*, 671–679. [CrossRef]
10. Petrović, V.; Xydeas, C. Gradient-based multiresolution image fusion. *IEEE Trans. Image Process.* **2004**, *13*, 228–237. [CrossRef]
11. Toet, A.; Van Ruyven, L.J.; Valeton, J.M. Merging thermal and visual images by a contrast pyramid. *Opt. Eng.* **1989**, *28*, 789–792. [CrossRef]
12. Sun, J.; Han, Q.; Kou, L.; Zhang, L.; Zhang, K.; Jin, Z. Multi-focus image fusion algorithm based on Laplacian pyramids. *J. Opt. Soc. Am. A* **2018**, *35*, 480–490. [CrossRef]
13. Yu, B.; Jia, B.; Ding, L.; Cai, Z.; Wu, Q.; Law, R.; Huang, J.; Song, L.; Fu, S. Hybrid dual-tree complex wavelet transform and support vector machine for digital multi-focus image fusion. *Neurocomputing* **2016**, *182*, 1–9. [CrossRef]
14. Bhatnagar, G.; Wu, Q.J.; Liu, Z. Directive contrast based multimodal medical image fusion in NSCT domain. *IEEE Trans. Multimed.* **2014**, *9*, 1014–1024. [CrossRef]
15. Liu, X.; Mei, W.; Du, H. Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion. *Neurocomputing* **2017**, *235*, 131–139. [CrossRef]
16. Zhang, Y.; Prasad, S. Multisource geospatial data fusion via local joint sparse representation. *IEEE Trans. Geosci. Remote* **2016**, *54*, 3265–3276. [CrossRef]
17. Mitianoudis, N.; Stathaki, T. Pixel-based and region-based image fusion schemes using ICA bases. *Inf. Fusion* **2007**, *8*, 131–142. [CrossRef]
18. Haghighat, M.B.A.; Aghagolzadeh, A.; Seyedarabi, H. Multi-focus image fusion for visual sensor networks in DCT domain. *Comput. Electr. Eng.* **2011**, *37*, 789–797. [CrossRef]
19. Zhang, L.; Zeng, G.; Wei, J. Adaptive region-segmentation multi-focus image fusion based on differential evolution. *Int. J. Pattern Recognit. Artif. Intell.* **2019**, *33*, 1954010. [CrossRef]
20. De, I.; Chanda, B. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Inf. Fusion* **2013**, *14*, 136–146. [CrossRef]
21. Farid, M.S.; Mahmood, A.; Al-Maadeed, S.A. Multi-focus image fusion using content adaptive blurring. *Inf. Fusion* **2018**, *45*, 96–112. [CrossRef]
22. Wan, T.; Zhu, C.; Qin, Z. Multifocus image fusion based on robust principal component analysis. *Pattern Recognit. Lett.* **2013**, *34*, 1001–1008. [CrossRef]
23. Sun, J.; Zhu, H.; Xu, Z.; Han, C. Poisson image fusion based on Markov random field fusion model. *Inf. Fusion* **2013**, *14*, 241–254. [CrossRef]
24. Aslantas, V.; Toprak, A.N. Multi-focus image fusion based on optimal defocus estimation. *Comput. Electr. Eng.* **2017**, *62*, 302–318. [CrossRef]
25. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* **2017**, *36*, 191–207. [CrossRef]
26. Amin-Naji, M.; Aghagolzadeh, A.; Ezoji, M. Ensemble of CNN for multi-focus image fusion. *Inf. Fusion* **2019**, *51*, 201–214. [CrossRef]
27. Wang, Z.; Wang, S.; Guo, L. Novel multi-focus image fusion based on PCNN and random walks. *Neural Comput. Appl.* **2018**, *29*, 1101–1114. [CrossRef]
28. Wang, Z.; Wang, S.; Zhu, Y. Multi-focus Image Fusion Based on the Improved PCNN and Guided Filter. *Neural Process. Lett.* **2017**, *45*, 75–94. [CrossRef]
29. Guo, X.; Nie, R.; Cao, J.; Zhou, D.; Mei, L.; He, K. FuseGAN: Learning to fuse multi-focus image via conditional generativeadversarial network. *IEEE Trans. Multimed.* **2019**, *21*, 1982–1996. [CrossRef]
30. Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf. Sci.* **2018**, *432*, 516–529. [CrossRef]
31. Yang, Y.; Ding, M.; Huang, S.; Que, Y.; Wan, W.; Yang, M.; Sun, J. Multi-focus image fusion via clustering PCA based joint dictionary learning. *IEEE Access* **2017**, *5*, 16985–16997. [CrossRef]

32. Li, L.; Si, Y.; Wang, L.; Jia, Z.; Ma, H. A novel approach for multi-focus image fusion based on SF-PAPCNN and ISML in NSST domain. *Multimed. Tools Appl.* **2020**, *79*, 24303–24328. [CrossRef]

33. Wei, B.; Feng, X.; Wang, K.; Gao, B. The multi-focus-image-fusion method based on convolutional neural network and sparse representation. *Entropy* **2021**, *23*, 827. [CrossRef]

34. Liu, Y.; Wang, L.; Cheng, J.; Li, C.; Chen, X. Multi-focus image fusion: A Survey of the state of the art. *Inf. Fusion* **2020**, *64*, 71–91. [CrossRef]

35. Ali, H.; Elmogy, M.; El-Daydamony, E.; Atwan, A. Multi-resolution MRI brain image segmentation based on Morphological pyramid and fuzzy C-mean clustering. *Arab. J. Sci. Eng.* **2015**, *40*, 3173–3185. [CrossRef]

36. Upla, K.P.; Joshi, M.V.; Gajjar, P.P. An edge preserving multiresolution fusion: Use of contourlet transform and MRF prior. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3210–3220. [CrossRef]

37. Easley, G.; Labate, D.; Lim, W.Q. Sparse directional image representations using the discrete shearlet transform. *Appl. Comput. Harmon. Anal.* **2008**, *25*, 25–46. [CrossRef]

38. Du, G.; Dong, M.; Sun, Y.; Li, S.; Mu, X.; Wei, H.; Lei, M.; Liu, B. A new method for detecting architectural distortion in mammograms by NonSubsampled contourlet transform and improved PCNN. *Appl. Sci.* **2019**, *9*, 4916. [CrossRef]

39. Yang, B.; Li, S. Multifocus Image Fusion and Restoration With Sparse Representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892. [CrossRef]

40. Ma, X.; Hu, S.; Liu, S.; Fang, J.; Xu, S. Multi-focus image fusion based on joint sparse representation and optimum theory. *Signal Process. Image Commun.* **2019**, *78*, 125–134. [CrossRef]

41. Li, S.; Yin, H.; Fang, L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [CrossRef]

42. Yin, H.; Li, Y.; Chai, Y.; Liu, Z.; Zhu, Z. A novel sparse-representation-based multi-focus image fusion approach. *Neurocomputing* **2016**, *216*, 216–229. [CrossRef]

43. Paul, S.; Sevcenco, I.S.; Agathoklis, P. Multi-exposure and multi-focus image fusion in gradient domain. *J. Circ. Syst. Comput.* **2016**, *25*, 1650123. [CrossRef]

44. Wang, Y.; Wang, Y. Fusion of 3-D medical image gradient domain based on detail-driven and directional structure tensor. *J. X-ray Sci. Technol.* **2020**, *28*, 1001–1016. [CrossRef] [PubMed]

45. Luo, X.; Zhang, Z.; Zhang, C.; Wu, X. Multi-focus image fusion using HOSVD and edge intensity. *J. Vis. Commun. Image Represent.* **2017**, *45*, 46–61. [CrossRef]

46. Kang, B.; Zhu, W.P.; Yan, J. Fusion framework for multi-focus images based on compressed sensing. *IET Image Process.* **2013**, *7*, 290–299. [CrossRef]

47. Liu, Z.; Chai, Y.; Yin, H.; Zhou, J.; Zhu, Z. A novel multi-focus image fusion approach based on image decomposition. *Inf. Fusion* **2017**, *35*, 102–116. [CrossRef]

48. Li, M.; Cai, W.; Tan, Z. A region-based multi-sensor image fusion scheme using pulse-coupled neural network. *Pattern Recognit. Lett.* **2006**, *27*, 1948–1956. [CrossRef]

49. Xiao, J.; Liu, T.; Zhang, Y.; Zou, B.; Lei, J.; Li, Q. Multi-focus image fusion based on depth extraction with inhomogeneous diffusion equation. *Signal Process.* **2016**, *125*, 171–186. [CrossRef]

50. Du, C.; Gao, S. Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. *IEEE Access* **2017**, *5*, 15750–15761. [CrossRef]

51. Ma, J.; Zhou, Z.; Wang, B.; Miao, L.; Zong, H. Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps. *Neurocomputing* **2019**, *335*, 9–20. [CrossRef]

52. Chen, M.; Yu, L.; Zhi, C.; Sun, R.; Zhu, S.; Gao, Z.; Ke, Z.; Zhu, M.; Zhang, Y. Improved faster R-CNN for fabric defect detection based on Gabor filter with Genetic Algorithm optimization. *Comput. Ind.* **2022**, *134*, 103551. [CrossRef]

53. Kowsher, M.; Alam, M.A.; Uddin, M.J.; Ahmed, F.; Ullah, M.W.; Islam, M.R. Detecting third umpire decisions & automated scoring system of cricket. In Proceedings of the 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 11–12 July 2019; pp. 1–8.

54. Javed Awan, M.; Mohd Rahim, M.S.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. *Diagnostics* **2021**, *11*, 105. [CrossRef]

55. Zhang, X.; Yang, Y.; Li, Z.; Ning, X.; Qin, Y.; Cai, W. An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field. *Entropy* **2021**, *23*, 435. [CrossRef]

56. Lopac, N.; Hržić, F.; Vuksanović, I.P.; Lerga, J. Detection of non-stationary GW signals in high noise from cohen's class of time-frequency representations using deep learning. *IEEE Access* **2021**, *10*, 2408–2428. [CrossRef]

57. Tang, H.; Xiao, B.; Li, W.; Wang, G. Pixel convolutional neural network for multi-focus image fusion. *Inf. Sci.* **2018**, *433*, 125–141. [CrossRef]

58. Wang, K.; Zheng, M.; Wei, H.; Qi, G.; Li, Y. Multi-modality medical image fusion using convolutional neural network and contrast pyramid. *Sensors* **2020**, *20*, 2169. [CrossRef]

59. 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. [CrossRef]

60. Yang, Y.; Yang, M.; Huang, S.; Ding, M.; Sun, J. Robust sparse representation combined with adaptive PCNN for multifocus image fusion. *IEEE Access* **2018**, *6*, 20138–20151. [CrossRef]

61. Basar, S.; Waheed, A.; Ali, M.; Zahid, S.; Zareei, M.; Biswal, R.R. An efficient defocus blur segmentation scheme based on Hybrid LTP and PCNN. *Sensors* **2022**, *22*, 2724. [CrossRef]
62. Liu, Z.; Blasch, E.; Xue, Z.; Zhao, J.; Laganiere, 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.* **2011**, *34*, 94–109. [CrossRef]
63. Tsai, D.C.; Chen, H.H. Reciprocal focus profile. *IEEE Trans. Image Process.* **2012**, *21*, 459–468. [CrossRef] [PubMed]
64. Ferris, M.H.; Blasen, E.; McLaughlin, M. Extension of no-reference deblurring methods through image fusion. In Proceedings of the 2014 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2014), Washington, DC, USA, 14–16 October 2014; pp. 1–6.
65. Hariharan, R.; Rajagopalan, A.N. Shape-from-focus by tensor voting. *IEEE Trans. Image Process.* **2012**, *21*, 3323–3328. [CrossRef]
66. Hu, Y.; Sirlantzis, K.; Howells, G. Signal-level information fusion for less constrained iris recognition using sparse-error low rank matrix factorization. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1549–1564. [CrossRef]