*Computers, Materials & Continua*
DOI: 10.32604/cmc.2022.023905
***Article***

Tech Science Press

# Enhanced Robotic Vision System Based on Deep Learning and Image Fusion

**E. A. Alabdulkreem[1], Ahmed Sedik[2], Abeer D. Algarni[3,*], Ghada M. El Banby[4], Fathi E. Abd El-Samie[3,5] and Naglaa F. Soliman[3,6]**

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 84428, Saudi Arabia
[2]Department of the Robotics and Intelligent Machines, Faculty of Artificial Intelligence, KafrelSheikh University, Kafrelsheikh, 33511, Egypt
[3]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 84428, Saudi Arabia
[4]Department of Industrial Electronics and Control Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, 32952, Egypt
[5]Department Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, 32952, Egypt
[6]Department of Electronics and Communications, Faculty of Engineering, Zagazig University, Zagazig, 44519, Egypt
*Corresponding Author: Abeer D. Algarni. Email: adalqarni@pnu.edu.sa
Received: 26 September 2021; Accepted: 30 March 2022

**Abstract:** Image fusion has become one of the interesting fields that attract researchers to integrate information from different image sources. It is involved in several applications. One of the recent applications is the robotic vision. This application necessitates image enhancement of both infrared (IR) and visible images. This paper presents a Robot Human Interaction System (RHIS) based on image fusion and deep learning. The basic objective of this system is to fuse visual and IR images for efficient feature extraction from the captured images. Then, an enhancement model is carried out on the fused image to increase its quality. Several image enhancement models such as fuzzy logic, Convolutional Neural Network (CNN) and residual network (ResNet) pre-trained model are utilized on the fusion results and they are compared with each other and with the state-of-the-art works. Simulation results prove that the fuzzy logic enhancement gives the best results from the image quality perspective. Hence, the proposed system can be considered as an efficient solution for the robotic vision problem with multi-modality images.

**Keywords:** Deep learning; fuzzy logic; image fusion; IR images

## 1 Introduction

It is known that image fusion is concerned with the merging process of multi-source images to attain an image with salient features that comprises complementary features from both source images. Image fusion is one of the recent research fields, which is involved in several applications. For example, medical image fusion [1,2] is carried out on images with different imaging modalities to get the fusion result with features of both images. Moreover, image fusion technology has been extensively implemented in applications such as object recognition [3], face recognition [4,5], remote sensing, and Internet-of-Things (IoT) [6]. The main issue in IoT is the utilization of a diversity of sensors in order to get a diversity of images for the same scene. This issue leads to the consumption of a large transmission bandwidth and storage space, and therefore image fusion is required to address the aforementioned problem.

Another implementation of image fusion is for merging IR and visible images that is considered as a multi-source sensor information fusion process. This fusion process is important for military surveillance and robotic vision [7–9]. It is possible to discriminate targets in IR images from the radiation difference between the target and its background. This difference is not affected by the environmental conditions like light, sand, and smoke. Generally, IR images have some drawbacks, such as unremarkable constituent information, low contrast, and poor visibility characteristics. Unlike IR images, visible images reveal the targets with good appearance, which allows interpretation with the human visual system [7,10]. To benefit from the nature of each of the IR and visible images, efficient image fusion algorithms are required to attain salient image features from both of them [11].

Different types of fusion including pixel-level, feature-level and decision-level exist [6]. The simplest is the pixel-level fusion, which depends on some sort of weighted average in the spatial domain. This type is simple and easy to implement, but the artifacts in both images still exist in the fusion result. On the other hand, feature-level fusion depends on feature extraction from both images, and then merging the extracted features to be used in subsequent classification tasks. The feature extraction may be more robust to the degradation effects in the images. The decision-level fusion is implemented by decision rules, and it achieves a high level of fusion. Firstly, the source images are processed separately to extract features of each of them, and then the most important information is selected based on a specified rule.

A general classification of image fusion algorithms includes spatial-domain algorithms and transform-domain algorithms [12,13]. The spatial-domain fusion begins by block-by-block segmentation of the two images to be merged [4,12,14]. The corresponding blocks in both images are combined together to get new formed blocks that keep the most salient information in both images. This type of fusion algorithms is appropriate, when the source images have the same modality, and it most likely gives artifacts in positions of block or region edges. In contrary, transform-domain image fusion has another concept of operation that depends on firstly transforming the images into an appropriate domain [15]. Multi-scale transformations are good candidates for this task. This type of fusion is appropriate for images of different modalities, and it gives a good performance if the weight coefficients are optimized and selected carefully in the fusion algorithm.

Recently, a new trend in image fusion based on machine learning has emerged [16,17]. Moreover, CNNs have been utilized for the image fusion tasks [12,18]. The convolutional layers used in CNNs have a significant role in the computer vision field to extract comprehensive and valuable features. They perform weighted averaging to generate the feature maps. Therefore, the characteristics of fused images obtained with CNNs are similar to those obtained with transform-domain fusion algorithms. Moreover, CNNs solve the fusion optimization problem in order to maximize the quality metric values of the fusion results [12].

Almost in the existing image fusion algorithms, an input image should have a high quality represented in good contrast and sharp edge details. However, for IR and visible image fusion, the edge details are sometimes weak due to the image acquisition conditions and the variability of environments.

Recently, deep learning has become an effective tool in several research fields of image processing [16], and hence it is recommended to achieve high-quality image fusion.

In this paper, an elegant framework is presented with the objective of integrating the image fusion based on deep learning with image enhancement in order to attain fused images that have high contrast and sharp edge details. The image enhancement is considered as a significant tool in computer vision to improve the quality, especially for acquired images that suffer from degradation problems such as the low contrast [19–22]. Moreover, the edge details are represented with the significant sharpness of edges, especially object edges. It leads to an enhancement of object appearance in images.

The key contributions of this proposed work are threefold. Firstly, operation is mainly concerned with color images that are represented in YCbCr format. The luminance obtained from this representation is treated with a deep learning network to get multi-layer features. Meanwhile, for the chrominance channels, a weighted fusion process is applied. The fused image is restructured by merging the modified luminance and two chrominance components. After that, the fused image is enhanced using a fuzzy logic algorithm, and two image enhancement benchmark models based on deep learning: CNN and ResNet pre-trained model. Finally, comparisons are presented to spot the proposed algorithm efficiency based on different types of evaluation metrics.

The contributions of this work can be listed in the following points:

1. Building a robotic vision system based on deep fusion and fuzzy-logic enhancement models.
2. Building benchmark enhancement models including deep learning and traditional models.
3. Assessment of both proposed and benchmark models.
4. Introduction of a comparison study to highlight the efficiency of the proposed framework.

To summarize the paper content, Section 2 presents several works related to image fusion and image enhancement methods. The proposed framework is given and explained in detail in Section 3. The proposed models including the proposed image fusion and image enhancement models are presented in Section 4. In Section 5, the image fusion assessment metrics are given. Simulation results on public datasets and discussions are introduced in Sections 6 and 7, respectively. Finally, the relevant conclusions are introduced in Section 8.

## 2 Related Work

With the variety of image processing tools today, image fusion has gained an essential rule in obtaining optimal image quality with as most useful features as possible. To handle this issue, the researchers have worked on constructing different sorts of algorithms for image fusion that depend on multi-scale representations, adaptive techniques, fuzzy logic techniques, neural and deep neural networks [10]. The fusion based on multi-scale representations is a significant type of fusion. Different types of transforms have been considered for this task including ridgelet, curvelet, Radon, wavelet and contourlet transforms [23–26]. In [23], the authors worked on the fusion of Computed Tomography (CT) and Magnetic Resonance (MR) images using the wavelet transform and the contourlet transform. This algorithm depends on the decomposition of source images with some sort of dual-tree complex wavelet transform. Hence, an energy fusion rule is adopted on the obtained coefficients with the help of the contourlet transform. Chen et al. [6] presented an algorithm for the fusion of visible and IR images with the objective of injecting some of the details in the visual images into the IR images that represent thermal distributions of objects. This algorithm depends on image sub-band decomposition using Laplacian pyramids. The maximum fusion rule has been adopted in this algorithm. The rationale behind the utilization of this rule is to eliminate any effect of blurring in the visual image within the fusion process.

Kanmani et al. [7] introduced a framework for IR and visible image fusion for the target of face recognition. The optimization techniques were used to enhance the face recognition process in order to achieve the highest recognition rates. Three different optimization-based methods have been introduced and compared in this work. Two of them depend on the dual-tree complex wavelet transform and the third one depends on the curvelet transform. The common thread between all three methods is that they all begin with the decomposition process, and then an optimization process is carried out on the obtained coefficients from both images in order to maximize the subsequent recognition rate obtained on the fusion results. Both swarm optimization and brain storm optimization have been investigated and compared in this work.

The utilization of sparse representation techniques has spread in new trends of image processing. Sparse representation allows representing images in the form of blocks based on certain transformation matrices that are composed majorly of few elements and a large number of zeros. These representations can be used in applications such as image super resolution and image fusion [10,15,24]. Sparse representation of multi-modality images such as visible and IR images can be used for the objective of fusion. Zhou et al. [24] presented a method that combines sparse representation of images with dictionary learning in order to fuse both IR and visible images. This method succeeded in obtaining fusion results with as much details as possible. The concept of image super resolution can be used to obtain images with much details based on some sort of dictionary techniques and single image super resolution algorithms. Liu et al. [15] tried to benefit from the super resolution through the multi-modality fusion process. They suggested working on the sub-bands of the decomposed source images with certain super resolution algorithms prior to the fusion process. This strategy succeeded in the fusion of visible and IR images. Still, there is a need to investigate the order in which both image decomposition and super resolution are implemented.

Recently, deep learning based on CNNs has found an outstanding role in image fusion, and it has been adopted in several research works [9,12,16]. Zhang et al. [12] introduced a two-layer CNN for image fusion. The CNN allows informative features of the source images, and it can be easily optimized on the training set. Piao et al. [9] studied IR and visible image fusion using Siamese CNN to extract the features in a weight map representation. Afterwards, the image fusion is performed using wavelet transform decomposition through weighted averaging. Another solution was presented in [16]. In this solution, source images are decomposed into approximation and details. Some components of the images are fused through weighted averaging and the others are fused with the VGG-19 network. The integration of image enhancement and image fusion can enhance the quality of the fusion results. Zhao et al. in [11] introduced a framework for this integration based on spectral total variation method. Moreover, in [26], a fusion method was introduced for images with degraded illumination. Firstly, the illumination is extracted from the source images and enhanced. Then, the fusion process is performed to get high-quality fusion results.

## 3 Proposed Framework

This paper presents a framework for robotic vision that consists of two main phases. The first phase is the fusion of IR and visible images. The proposed scenario includes two sensors to capture the images. The IR images are collected by Raytheon Palm-IR-Pro sensor, while the visible images are collected by Panasonic WV-CP234 sensor. These sensors are assumed to be implemented on the robot machine. The second phase is the enhancement of the fusion results. Both image fusion and enhancement are carried out on the captured images by a central server, which could be connected to one or more robot machines. The objective is to reduce the computational cost through centralized processing. Another reason is to save the energy of robots. Moreover, a central control unit makes it easy to troubleshoot errors that may occur. A disadvantage of the proposed scenario is that a high-speed connection is required to connect the robot machines to the central server without a considerable delay. Fig. 1 shows the proposed framework.

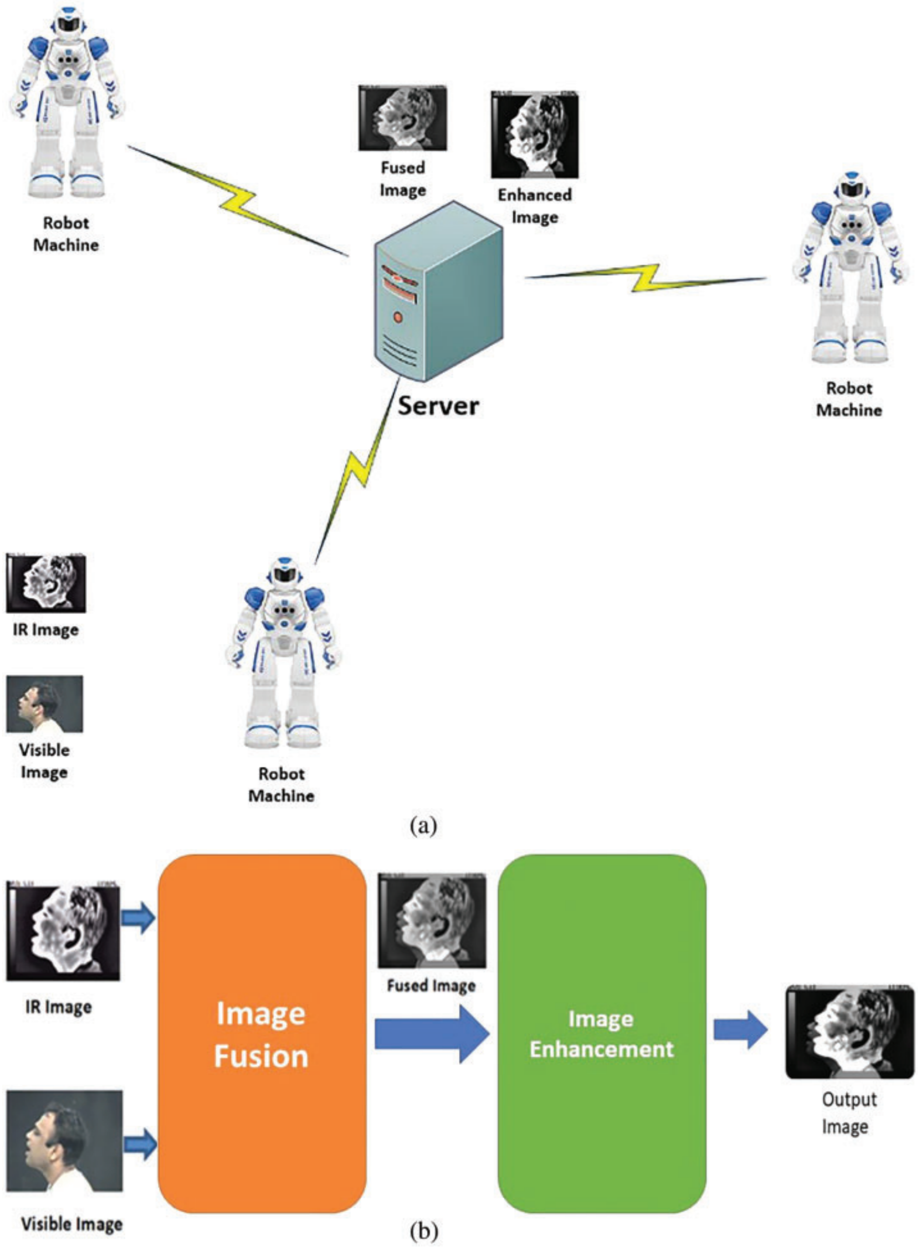


**Figure 1:** (a) General hierarchy of the proposed framework, (b) Block diagram of the proposed framework

### *3.1 Image Fusion Based on Deep Learning*

A CNN is considered as an efficient candidate for image fusion [27]. The main idea to design such an efficient CNN model is to train the network to predict an output close to the real-state target. This closeness can be guaranteed based on loss minimization. For an input $x$ to be mapped to a desired output $y$ through a function $f$, a selected loss function need to be minimized through a feed-forward operation with some sort of error back-propagation. In most cases, the mean square error between the real output and the desired output represents the cost function to be minimized. This paper is based on a Multi-Exposure Fusion Structural Similarity Index Metric (MEF SSIM) as a loss function [27]. A loss related to structural integrity and luminance consistency on multiple scales is evaluated and injected into the optimization process of the CNN.

Fig. 2 illustrates the proposed framework for IR and visible image fusion. Firstly, the $YC_bC_r$ color image transformation is applied on both images. The CNN-based image fusion is applied on the luminance components of both images due to the presence of much details and variations in these components. On the other hand, the chrominance components of both images ($C_b$ and $C_r$) are fused through weighted averaging as they are poor in details. Finally, the fusion result is transformed back to the RGB color coordinate system.

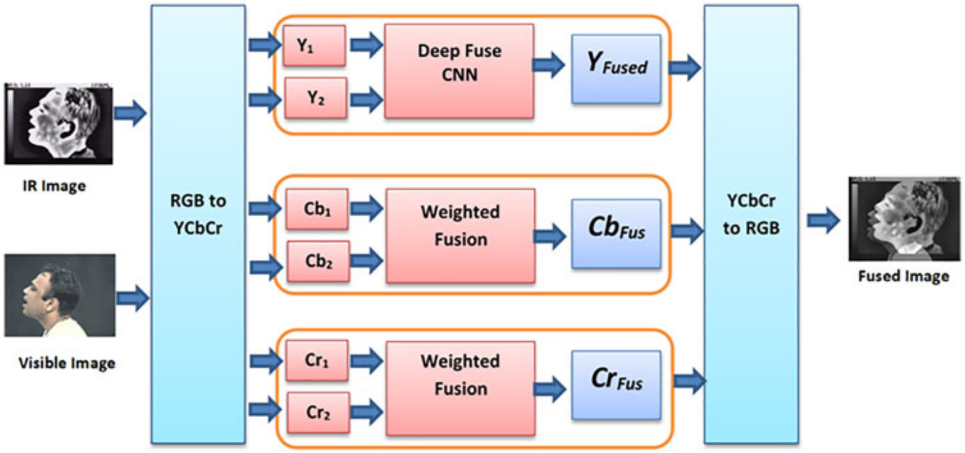


**Figure 2:** The proposed image fusion model based on deep learning

Fig. 3 shows the proposed CNN fusion model. Assume the IR image as $Y_1$ and the visible image as $Y_2$. Both inputs are enrolled into a pair of convolutional layers ($C_{11}$, $C_{21}$) and ($C_{12}$, $C_{22}$) in order to extract the features from them. Both $C_{11}$ and $C_{12}$ consist of 16 filters with a size of $5 \times 5$. In addition, $C_{21}$ and $C_{22}$ consist of 32 filters with a size of $7 \times 7$. The fusion of feature maps is performed with an addition layer. The obtained feature map is finally reconstructed using three convolutional layers ($C_3$, $C_4$, $C_5$). $C_3$ consists of 32 filters with a size of $7 \times 7$. In addition, $C_4$ consists of 16 filters with a size of $5 \times 5$. Furthermore, $C_5$ consists of a single filter with a size of $5 \times 5$. The proposed deep fusion model is trained on 4000 image pairs from the IRIS thermal visible face dataset (TRIS-TVFD) with 100 epochs, a batch size of 60 and a learning rate of $10^{-4}$.

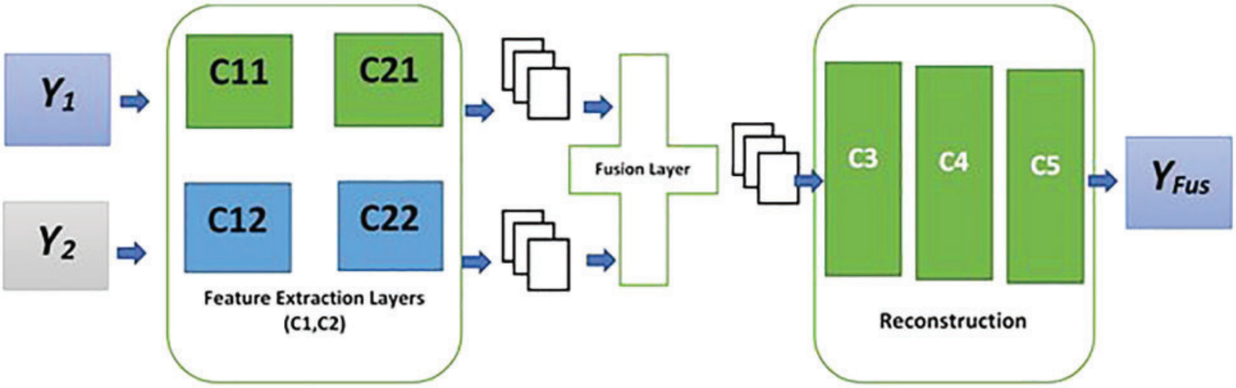


**Figure 3:** Stages of image fusion based on CNN (Feature extraction, Fusion, and Reconstruction)

### *3.1.1 MEF SSIM Loss Function*

The proposed loss function is the MEF SSIM [27,28]. Assume that $\{y_k\} = \{y_k | k = 1, 2\}$ denotes the set of image patches extracted at a location $p$ of a certain pixel from the input image pairs. In addition, assume that $y_f$ denotes the patch extracted from the fused image at the location $p$. A fusion score is obtained based on the input patches $y_k$ and the fused patch $y_f$. The SSIM indicates the degree of similarity between the input patches $y_k$ and the obtained fused image patch $y_f$. There are three aspects of similarity: contrast ($c$), luminance ($l$), and structure ($s$), and their product is used to calculate the overall index.

$$l\left(y_k,\ y_f\right) = \frac{2\mu_{y_k}\mu_{y_f} + C_1}{\mu_{y_k}^2 + \mu_{y_f}^2 + C_1} \tag{1}$$

$$c\left(y_k,\ y_f\right) = \frac{2\sigma_{y_k}\sigma_{y_f} + C_2}{\sigma_{y_k}^2 + \sigma_{y_f}^2 + C_2} \tag{2}$$

$$s\left(y_k,\ y_f\right) = \frac{\sigma_{y_k y_f} + C_3}{\sigma_{y_k}\sigma_{y_f} + C_3} \tag{3}$$

$$SSIM\left(y_k, y_f\right) = \left[l\left(y_k, y_f\right)\right] \cdot \alpha \cdot \left[c\left(y_k, y_f\right)\right] \cdot \beta \cdot \left[s\left(y_k, y_f\right)\right]\gamma \tag{4}$$

where $\mu_{y_k}$, $\mu_{y_f}$, $\sigma_{y_k}$, $\sigma_{y_f}$, and $\sigma_{y_k y_f}$ represent local means, standard deviations, and cross-covariance for input image patch $y_k$ and output image patch $y_f$. $C_1$, $C_2$, and $C_3$ are stabilization constants. With $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, the SSIM is given as:

$$SSIM\left(y_k, y_f\right) = \frac{\left(2\mu_{y_k}\ \mu_{y_f} + C_1\right)\left(2\sigma_{y_k y_f} + C_2\right)}{\left(\mu_{y_k}^2 + \mu_{y_f}^2 + C_1\right)\left(\sigma_{y_k}^2 + \sigma_{y_f}^2 + C_2\right)} \tag{5}$$

The obtained score at $p$ is given as:

$$Score\left(p\right) = SSIM\left(y_k, y_f\right) \tag{6}$$

Hence, the total loss is calculated as:

$$Loss = \frac{1}{N} \sum_{p \in P} score\,(p) \tag{7}$$

where $N$ represents the image size in pixels and $P$ is the set of all pixels in the input image.

The inherent operation to estimate the MEF SSIM is based on a gradient descent optimizer, and this leads to some sort of similarity between fusion results and the original source images.

### *3.2 Image Enhancement Based on Fuzzy Logic*

In IR images, due to low contrast, fine details and several structures may not be visible, and consequently several areas or edges are unclear and fuzzy in nature. Hence, IR image enhancement is an essential demand. Fundamentally, IR image enhancement includes contrast enhancement and rim enhancement to enhance the dynamic range of IR images. This process allows to discriminate objects or facial expressions from the IR images. There are several crisp approaches for image enhancement. One of the most common approaches is histogram equalization. However, due to the interference between pixel values of objects and the background, crisp enhancement techniques have limited performance. For this reason, fuzzy set methods have been presented to overcome the vagueness in pixels. There are different publications on image enhancement based on fuzzy theory [29]. Enhancement methods based on fuzzy logic provide high-quality images in a short time. Fuzzy image processing is a transformation approach applied on the input image in gray-scale domain to get a transformed output image in the fuzzy domain, which is processed, modified and defuzzified to provide an enhanced output image in the gray-scale domain. There are three main stages for any fuzzy image processing algorithm: image fuzzification, membership value modification and defuzzification. The merit of fuzzy image enhancement lies in the middle step of the membership value modification. The block diagram describing the main stages of fuzzy image processing is displayed in Fig. 4 below.

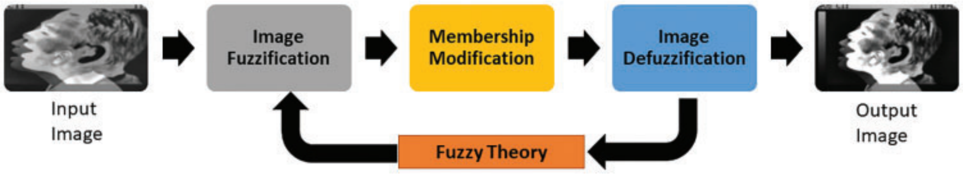


**Figure 4:** The proposed fuzzy image enhancement model

Membership plane is considered as the main part of fuzzy image processing, where the membership degree refers to the degree of belongingness of any pixel to an image. Fuzzification process is the transformation of the intensity crisp value of a pixel to a membership degree using a specific function. Appropriate fuzzy methods are selected to perform modification of membership values based on user requirements. In this paper, an intensification operator is used for IR image enhancement. Hence, the membership values are treated with an intensifier [30]. Firstly, the membership degree of each pixel is calculated using a selected membership function. This is followed by the modification process, which transforms the degree of membership values above 0.5 to much higher degrees and the degree of membership values less than 0.5 to much lower degrees to achieve a better contrast in the image. The membership function value is defined as $\mu(i,j)$ [31–34].

### 3.3 *Image Enhancement Benchmark Models*

Image enhancement is implemented by different models. The first model is based on a CNN. The second model is based on a pre-trained ResNet. The third model is based on the interpolation function in OpenCV Python library.

#### *3.3.1 Deep Learning Benchmark Models*

Deep learning is involved in several image processing applications. This paper covers the CNN-based image fusion. In addition, it presents two enhancement models based on deep learning. The first one is based on ResNet. It consists of two pairs of convolutional (Conv.) and Batch Normalization (BN) layers. In addition, a Rectified Linear Unit (ReLU) activation function is implemented after the first pair, while an addition function is performed after the second pair. An addition is carried out between both the original image and the feature map generated from the sequence. Fig. 5 shows the enhancement model based on ResNet.

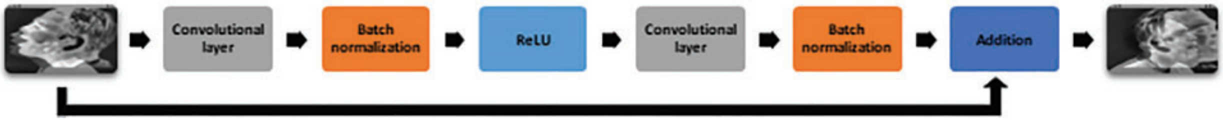


**Figure 5:** Image enhancement benchmark model based on ResNet

The second benchmark model is based on a CNN. It consists of three convolutional layers to enlarge the input image. The model consists of two main stages. The first stage is performed to prepare the input images in order to be suitable for the nature of the deep learning model. In this stage, the RGB color image representation is adopted. In addition, the RGB images are transformed into tensors to be enrolled into the deep learning model. Fig. 6 shows the sequence of layers performed in this model.

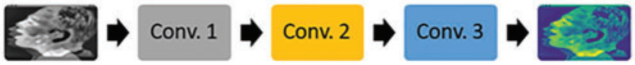


**Figure 6:** Image enhancement benchmark model based on a CNN

#### *3.3.2 Benchmark Model Based on Interpolation*

Another benchmark model is based on interpolation. The interpolation process is performed to upscale the input image. In this case, the input image is represented as the fused image. In order to understand the necessity of the interpolation process, the relationship between the fused image and the required high-resolution image can be represented as:

$$\mathbf{g} = \mathbf{Df} + \mathbf{v} \tag{8}$$

where **g, f, D**, and **v** represent the fused image, the high-resolution image, the decimation matrix, and the noise, respectively.

The matrix **D** is defined as:

$$\mathbf{D} = \mathbf{D}_1 \otimes \mathbf{D}_1 \tag{9}$$

where ⊗ refers to a Kronecker product operation [35] with:

$$\mathbf{D}_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \tag{10}$$

Fig. 7 shows how a low-resolution image is related to the high-resolution imaged through the explained decimation model that needs to be inverted in order to acquire the required high-resolution image.

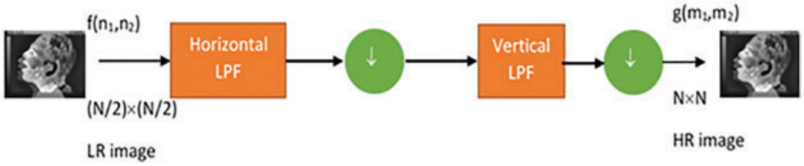


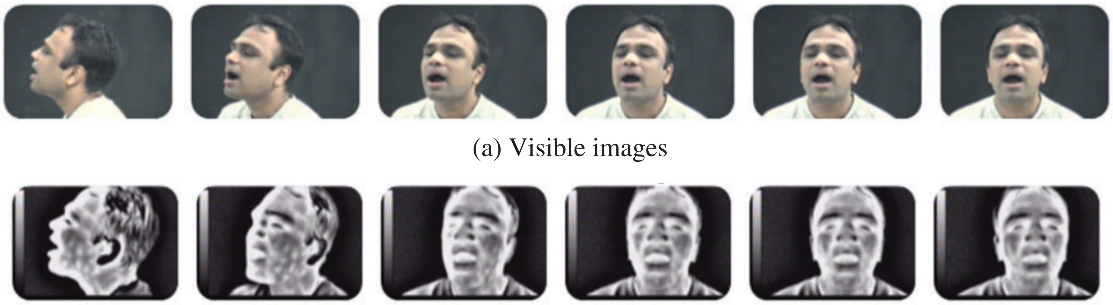
**Figure 7:** Proposed image enhancement model based on interpolation

## 4 Simulation Experiments

Several experiments have been conducted to assess the suggested image enhancement and fusion framework. Different evaluation metrics are adopted in the assessment process [36,37]. These metrics include entropy to represent the amount of information in the fusion results, in addition to average gradient, contrast and edge intensity to reflect the edge details in images.

### *4.1 Dataset Description*

The proposed models have been carried out on a part of the IRIS thermal visible face dataset (TRIS-TVFD) [38]. This dataset includes poses of expressions. The selected images belong to the first expression of a single person. The visible images are collected by Panasonic WV-CP234 sensor, while the IR images are collected by Raytheon Palm-IR-Pro sensor. Fig. 8 shows the selected visible and IR images.



(a) Visible images

(b) IR images

**Figure 8:** Samples of visible and IR image datasets

### 4.2 Simulation Results

For simulation experiments, a local machine with Intel core i7 8th edition CPU, 16 GB DDR5 RAM and 4GB DDR5 GPU with CUDA has been used. This paper introduces two main contributions in image fusion and image enhancement. The first contribution is a deep learning model for image fusion. The second contribution is selecting an efficient technique for image enhancement. Tabs. 1 and 2 show the quality metrics of visible and IR images. Although the visible images have high values of entropy, the IR images have high values for other quality metrics such as contrast, edge intensity and average gradient. So, the fused images are expected to have the advantages of both types of images.

**Table 1:** Quality metrics of visible images

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.4993 | 0.5455 | 0.6058 | 0.6256 | 0.6369 | 0.6437 |
| Edge intensity | 30.6819 | 35.9218 | 38.5243 | 38.7900 | 38.8297 | 38.8387 |
| Entropy | 7 | 7 | 7 | 7 | 7 | 7 |
| Average gradient | 3.1882 | 3.6489 | 3.8624 | 3.8716 | 3.8732 | 3.8736 |

**Table 2:** Quality metrics of IR images

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.9776 | 0.9682 | 0.9803 | 0.9751 | 0.9684 | 0.9695 |
| Edge intensity | 89.7657 | 83.0753 | 79.3130 | 78.5465 | 78.6558 | 78.8585 |
| Entropy | 4 | 4 | 4 | 4 | 4 | 4 |
| Average gradient | 8.5763 | 7.9739 | 7.6336 | 7.5678 | 7.5759 | 7.5882 |

#### 4.2.1 Results of Image Fusion

Fig. 9 shows the obtained images from the image fusion process. Tab. 3 shows the evaluation metrics of the fused images. It can be observed that the fused images combine high entropy from visible images and high contrast, edge intensity and average gradient from IR images. In addition, the fused images score a measurement of enhancement (EME) in the range of 15 to 18. So, the fused images reveal a considerable enhancement for both visible and IR images.

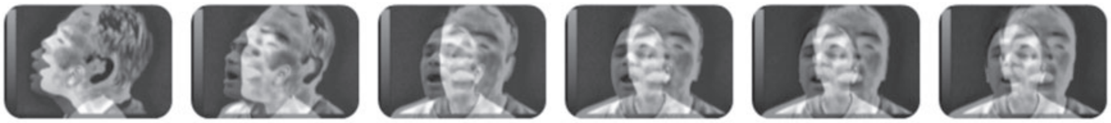


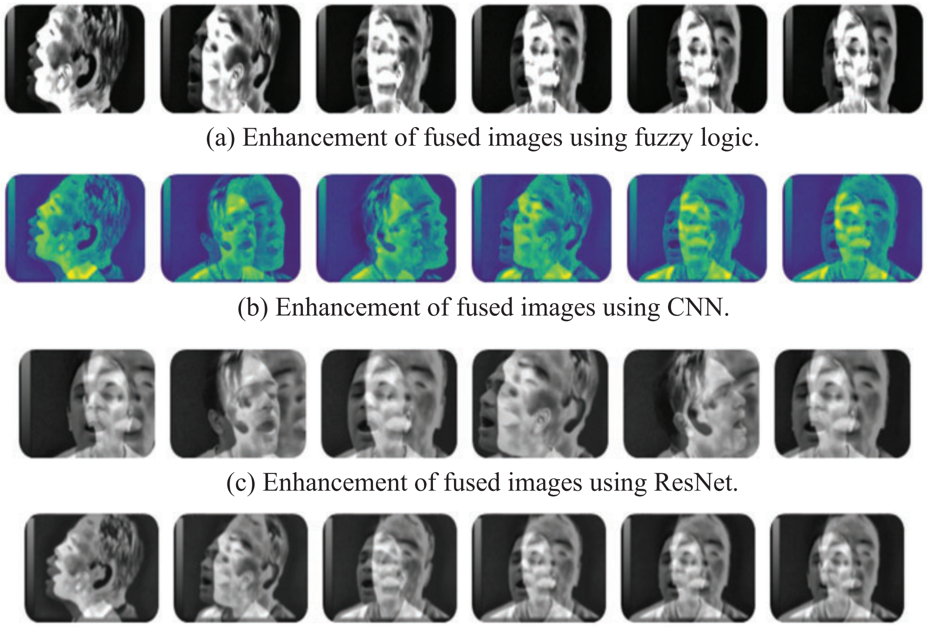
**Figure 9:** Fused images using deep learning

**Table 3:** Quality metrics of fused images

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.7656 | 0.7224 | 0.7470 | 0.7567 | 0.8064 | 0.8052 |
| Edge intensity | 57.5001 | 56.206 | 55.8283 | 55.6551 | 56.6229 | 56.8730 |
| Entropy | 7 | 7 | 7 | 7 | 7 | 7 |
| Average gradient | 5.54603 | 5.419 | 5.3777 | 5.3623 | 5.4588 | 5.4793 |
| EME | 16.64539 | 15.54766 | 15.36652 | 16.09842 | 18.7631 | 18.44208 |

### *4.2.2 Results of Image Enhancement*

This section presents the results of different image enhancement techniques including fuzzy logic, CNN, ResNet, and interpolation models. The proposed image enhancement models are carried out on the fused images obtained with the proposed image fusion technique. The proposed models are evaluated in order to achieve an optimal performance based on the evaluation metrics. The aim is to obtain a high performance in terms of entropy, contrast, edge intensity, average gradient and EME. Fig. 10 shows the images, which result from each enhancement model. In addition, Tabs. 4–7 show the simulation results of each enhancement model. The simulation results reveal that the proposed fuzzy logic enhancement model has a superior performance. The proposed fuzzy logic image enhancement model achieves an average EME value of 30, which is a high value compared with the other EME values of CNN, ResNet and interpolation models.



(a) Enhancement of fused images using fuzzy logic.

(b) Enhancement of fused images using CNN.

(c) Enhancement of fused images using ResNet.

(d) Enhancement of fused images using interpolation.

**Figure 10:** Images resulting from each enhancement model

**Table 4:** Quality metrics of enhanced images based on fuzzy logic

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.9458 | 0.9357 | 0.9561 | 0.9548 | 0.9691 | 0.9696 |
| Edge intensity | 90.8577 | 93.5528 | 96.9116 | 95.4729 | 93.0612 | 93.8710 |
| Entropy | 7 | 7 | 8 | 7 | 7 | 7 |
| Average gradient | 8.67740 | 8.9251 | 9.2037 | 9.0689 | 8.8395 | 8.9142 |
| EME | 28.95886 | 28.99249 | 30.68904 | 30.45498 | 33.68213 | 33.37140 |

**Table 5:** Quality metrics of enhanced images based on CNN

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.6826 | 0.7130 | 0.7092 | 0.6392 | 0.6625 | 0.6748 |
| Edge intensity | 56.9894 | 56.5766 | 57.1255 | 55.0496 | 54.1555 | 54.2197 |
| Entropy | 7 | 7 | 7 | 7 | 7 | 7 |
| Average gradient | 5.5206 | 5.426 | 5.5611 | 5.3328 | 5.2457 | 5.2544 |
| EME | 14.19053 | 14.65736 | 14.67437 | 13.24740 | 12.8365 | 13.71884 |

**Table 6:** Quality metrics of enhanced images based on ResNet

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.8071 | 0.7827 | 0.7610 | 0.7210 | 0.7792 | 0.8081 |
| Edge intensity | 55.4584 | 54.1746 | 53.9137 | 55.6464 | 53.0078 | 54.9567 |
| Entropy | 2.5 | 2.6 | 2.6 | 2.57 | 2.57 | 2.57 |
| Average gradient | 5.2741 | 5.1848 | 5.1070 | 5.2734 | 5.1039 | 5.2092 |
| EME | 16.3224 | 14.60077 | 13.55203 | 14.11655 | 14.2093 | 15.43813 |

**Table 7:** Quality metrics of enhanced images based on interpolation

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.7351 | 0.6853 | 0.7102 | 0.7179 | 0.7660 | 0.7634 |
| Edge intensity | 46.2368 | 45.5861 | 45.2091 | 45.1612 | 45.9436 | 46.0907 |
| Entropy | 2.57 | 2.57 | 2.57 | 2.57 | 2.57 | 2.57 |
| Average gradient | 4.2323 | 4.1660 | 4.1284 | 4.1234 | 4.1952 | 4.2098 |
| EME | 13.18903 | 12.29305 | 12.10548 | 12.46011 | 14.31898 | 14.11047 |

#### *4.2.3 Results of Simple Enhancement Models*

To highlight the performance of the proposed robotic vision system, we deployed simple enhancement models including median filter and histogram equalization. This deployment is carried out to clarify the importance of using fuzzy logic enhancement rather than the existing simple models. Tabs. 8

and 9 illustrate the quality assessment metrics of median filter and histogram equalization, respectively. We can observe that the fuzzy logic image enhancement model achieves a high performance compared to those of the existing models including median filter and histogram equalization in terms of quality assessment metrics. So, it can be considered as an efficient enhancement model for IR and visible robotic vision.

**Table 8:** Quality metrics of enhanced images based on median filter

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.9 | 0.8229 | 0.8469 | 0.849 | 0.8683 | 0.8550 |
| Edge intensity | 88.54 | 89.19 | 89.17 | 89.4153 | 89.068 | 88.147 |
| Entropy | 5.94 | 5.87 | 5.8 | 5.867 | 5.8732 | 5.8752 |
| Average gradient | 8.6 | 9.02 | 9.06 | 9.0974 | 9.0788 | 8.9725 |
| EME | 21.29 | 23.26 | 22.993 | 23.1927 | 23.596 | 23.46162 |

**Table 9:** Quality metrics of enhanced images based on histogram equalization

| Quality metric | Case_1 | Case_2 | Case_3 | Case_4 | Case_5 | Case_6 |
|---|---|---|---|---|---|---|
| Contrast | 0.7508 | 0.6966 | 0.724 | 0.7334 | 0.7817 | 0.779 |
| Edge intensity | 49.539 | 48.69 | 48.28 | 48.267 | 49.08 | 49.33 |
| Entropy | 7.2492 | 7.206 | 7.158 | 7.13 | 7.155 | 7.15 |
| Average gradient | 4.5815 | 4.496 | 4.45 | 4.4551 | 4.53 | 4.55 |
| EME | 13.978 | 12.94 | 12.82 | 13.308 | 15.31 | 15.04 |

## 5 Result Discussion

This paper presented a computer vision system for efficient robotic vision. The proposed framework comprises image fusion and image enhancement. It begins with image fusion in the first stage. Moreover, the image enhancement stage can be implemented with different enhancement models. This stage could be built with fuzzy logic, CNN, ResNet or interpolation. In order to evaluate these models, we used various quality evaluation metrics. The main evaluation metric used is the EME. This quality evaluation metric indicates the amount of enhancement in the image. Fig. 11 shows a visual comparison between the proposed fuzzy logic model, CNN, ResNet and interpolation models. Furthermore, it includes a comparison with and without enhancement. The comparison reveals that the proposed framework based on deep learning image fusion with fuzzy logic enhancement achieves EME values of 28, 28. 30, 30, 33, 33 for Case_1, Case_2, Case_3, Case_4 and Case_5, respectively. To provide more clarification on the performance of the proposed framework, it has been compared with the works in the literature. Tab. 10 shows this comparison between the proposed framework and the previous ones for efficient image fusion.

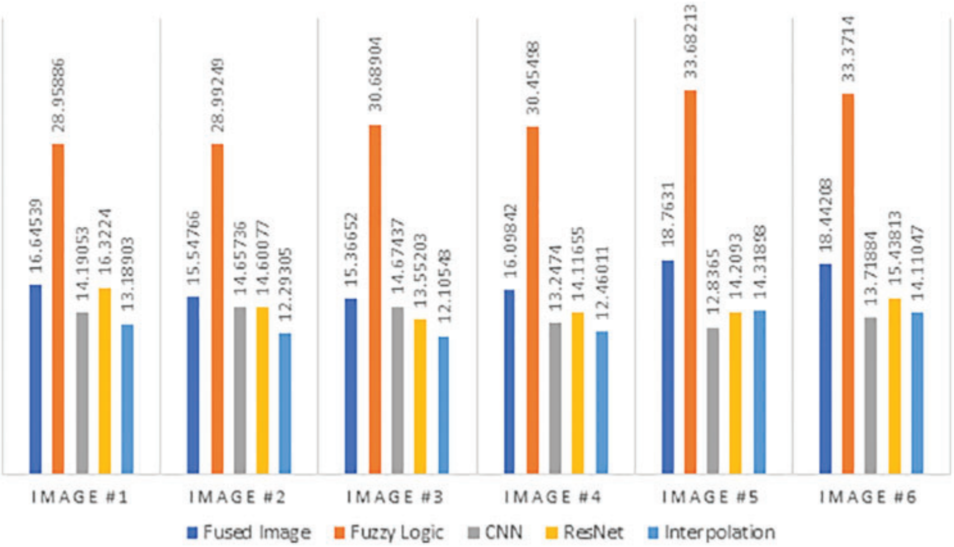


**Figure 11:** EME comparison between the proposed models of enhancement

**Table 10:** Comparison between the proposed framework and the traditional works in terms of quality metrics

| Quality metric | SWT [39] | NSCT [40] | PCA [41] | DWT [42] | Curvelet [43] | Median filter | Histogram equalization | Deep fusion + Fuzzy logic (proposed) |
|---|---|---|---|---|---|---|---|---|
| Average gradient | 0.0683 | 9.8019 | 0.0382 | 0.0639 | 0.0902 | 8.79145 | 4.5104 | 8.9381 |
| Contrast | 0.7474 | 0.6711 | 0.6650 | 0.7443 | 1.1792 | 0.85701 | 0.74425 | 0.95518 |
| Entropy | 7.7436 | 7.5815 | 7.5646 | 7.7377 | 7.6022 | 5.8709 | 7.1477 | 7.16667 |

## 6 Conclusions

This paper discussed the problem of computer vision for robot devices that work on IR and visible images. The proposed framework consists of two stages. The first stage is image fusion based on deep learning, while the second stage is image enhancement. Different image enhancement models have been investigated in this paper including fuzzy logic, CNN, ResNet, and image interpolation. In addition, the proposed framework has been carried out on both IR and visible images in order to obtain high-quality fusion results. The simulation results reveal that the proposed framework based on image fusion and fuzzy logic image enhancement achieves an optimal image quality for robotic vision of IR scenes, when merged with visible scenes. Furthermore, in the future work, we will investigate image fusion with motion artifacts.

**Acknowledgement:** The authors would like to thank the support of the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University.

**Funding Statement:** This research was funded by the Deanship of Scientific Research at Princess Nourah Bint Abdulrahman University through the Research Funding Program (Grant No# FRP-1440-23).

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

## References

[1] A. Dogra, B. Goyal and S. Agrawal, “From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications,” *IEEE Access*, vol. 5, pp. 16040–16067, 2017.

[2] J. Du, W. Li, K. Lu and B. Xiao, “An overview of multi-modal medical image fusion,” *Neurocomputing*, vol. 215, pp. 3–20, 2016.

[3] Y. Liu, Y. Li, X. Ma and R. Song, “Facial expression recognition with fusion features extracted from salient facial areas,” *Sensors*, vol. 17, no. 4, pp. 712, 2017.

[4] H. Tang, B. Xiao, W. Li and G. Wang, “Pixel convolutional neural network for multi-focus image fusion,” *Information Sciences*, vol. 433, pp. 125–141, 2018.

[5] Y. Zhang and Q. Ji, “Active and dynamic information fusion for facial expression understanding from image sequences,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699–714, 2005.

[6] J. Chen, X. Li, L. Luo, X. Mei and J. Ma, “Infrared and visible image fusion based on target-enhanced multiscale transform decomposition,” *Information Sciences*, vol. 508, pp. 64–78, 2020.

[7] M. Kanmani and V. Narasimhan, “Optimal fusion aided face recognition from visible and thermal face images,” *Multimedia Tools and Applications*, vol. 79, pp. 17859–17883, 2020.

[8] J. Ma, Y. Ma and C. Li, “Infrared and visible image fusion methods and applications: A survey,” *Information Fusion*, vol. 45, pp. 153–178, 2019.

[9] J. Piao, Y. Chen and H. Shin, “A new deep learning based multi-spectral image fusion method,” *Entropy*, vol. 21, no. 6, pp. 570, 2019.

[10] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo *et al.,* “Infrared and visible image fusion via detail preserving adversarial learning,” *Information Fusion*, vol. 54, pp. 85–98, 2020.

[11] W. Zhao, H. Lu and D. Wang, “Multisensor image fusion and enhancement in spectral total variation domain,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 866–879, 2017.

[12] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao *et al.,* “IFCNN: A general image fusion framework based on convolutional neural network,” *Information Fusion*, vol. 54, pp. 99–118, 2020.

[13] Y. Zhang, X. Bai and T. Wang, “Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure,” *Information Fusion*, vol. 35, pp. 81–101, 2017.

[14] Z. Zhou, S. Li and B. Wang, “Multi-scale weighted gradient-based fusion for multi-focus images,” *Information Fusion*, vol. 20, pp. 60–72, 2014.

[15] Y. Liu, S. Liu and Z. Wang, “A general framework for image fusion based on multi-scale transform and sparse representation,” *Information Fusion*, vol. 24, pp. 147–164, 2015.

[16] H. Li, X. -J. Wu and J. Kittler, “Infrared and visible image fusion using a deep learning framework,” in *2018 24th Int. Conf. on Pattern Recognition (ICPR)*, Milan, Italy, pp. 2705–2710, 2018.

[17] Y. Liu, X. Chen, H. Peng and Z. Wang, “Multi-focus image fusion with a deep convolutional neural network,” *Information Fusion*, vol. 36, pp. 191–207, 2017.

[18] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward *et al.,* “Deep learning for pixel-level image fusion: Recent advances and future prospects,” *Information Fusion*, vol. 42, pp. 158–173, 2018.

[19] Y. S. Moon, B. G. Han, H. S. Yang and H. G. Lee, “Low contrast image enhancement using convolutional neural network with simple reflection model,” *Advances in Science, Technology and Engineering Systems*, vol. 4, no. 1, pp. 159–164, 2019.

[20] Y. Choi and R. Krishnapuram, “Image enhancement based on fuzzy logic,” *Proc., Int. Conf. on Image Processing*, vol. 1, pp. 167–170, 1993.

[21] I. Kaur and N. Neeru, “An improved method for image enhancement of remote sensed images using fusion methods,” *International Journal of Advanced Research in Computer Science*, vol. 8, no. 7, pp. 928–931, 2017.

[22] K. Koteswararao and K. Veera Swamy, "Multimodal medical image fusion using nsct and dwt fusion frame work," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 2278–3075, 2019.

[23] J. Wang, J. Peng, X. Feng, G. Heand, J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Physics & Technology*, vol. 67, pp. 477–489, 2014.

[24] Z. Zhou, M. Dong, X. Xie and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Applied Optics*, vol. 55, no. 23, pp. 6480–6490, 2016.

[25] C. Qi, Q. Li, Y. Liu, J. Ni, R. Ma *et al.,* "Infrared image segmentation based on multi-information fused fuzzy clustering method for electrical equipment," *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, pp. 1729881420909600, 2020.

[26] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding *et al.,* "A Fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[27] F. Qian, J. Guo, T. Sun and T. Wang, "Quantitative assessment of laser-dazzling effects through wavelet-weighted multi-scale SSIM measurements," *Optics & Laser Technology*, vol. 67, pp. 183–191, 2015.

[28] Y. Tang, F. Ren and W. Pedrycz, "Fuzzy C-means clustering through SSIM and patch for image segmentation," *Applied Soft Computing*, vol. 87, pp. 105928, 2020.

[29] T. Chaira, "Medical image processing: Advanced fuzzy set theoretic techniques," Boca Raton, Florida, USA, CRC Press, 2015.

[30] A. K. Gupta, S. S. Chauhan and M. Shrivastava, "Low contrast image enhancement technique by using fuzzy method," *International Journal of Engineering Research and General Science*, vol. 4, no. 2, pp. 518–526, 2016.

[31] T. Chaira and A. K. Ray, "Fuzzy Image PreProcessing," in "*Fuzzy Image Processing and Applications with MATLAB*", William Francis, Richard Taylor, United Kingdom, Taylor & Francis Group, LLC, pp. 45–68, 2009.

[32] T. Chaira, "Image Enhancement," in "*Medical Image Processing-Advanced Fuzzy Set Theoretic Techniques*", William Francis, Richard Taylor, United Kingdom, CRC Press, Taylor & Francis Group, LLC, pp. 83–108, 2015.

[33] W. Zhang, J. Li and Z. Hua, "Near-infrared shadow detection based on HDR image," *Multimed Tools Appl*, 2022. https://doi.org/10.1007/s11042-022-12996-9.

[34] P. Jain and R. Meenu, "Automatic contrast enhancement using fuzzy logic for real time object recognition system," *International Journal of Scientific & Engineering Research*, vol. 8, no. 3, pp. 762–765, 2017.

[35] S. E. El-Khamy, M. M. Hadhoud, M. I. Dessouky, B. M. Salam and F. E. Abd El-Samie, "Efficient implementation of image interpolation as an inverse problem," *Digital Signal Processing*, vol. 15, no. 2, pp. 137–152, 2005.

[36] S. E. El-Khamy, M. M. Hadhoud, M. I. Dessouky, B. M. Salam and F. E. -S. Abd El-Samie, "New techniques to conquer the image resolution enhancement problem," *Progress in Electromagnetics Research B*, vol. 7, pp. 13–51, 2008.

[37] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquat. Procedia*, vol. 4, no. Icwrcoe, pp. 133–142, 2015.

[38] https://www.trb.org/InformationServices/ResourcesfortheTRISDatabases.aspx.

[39] L. Yang, B. L. Guoand, W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.

[40] P. K. Atrey, M. A. Hossain, A. El Saddik and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Systems*, vol. 16, no. 6, pp. 345–379, 2010.

[41] T. A. Tuan, H. V. Long, R. Kumar, I. Priyadarshini and N. T. K. Son, "Performance evaluation of botnet DDoS attack detection using machine learning," *Evolutionary Intelligence*, vol. 13, pp. 1–12, 2019.

[42] A. Wang, H. Sun and Y. Guan, "The application of wavelet transform to multi-modality medical image fusion," in *IEEE Int. Conf. on Networking, Sensing and Control*, Lauderdale, FL, USA, pp. 270–274, 2006.

[43] A. Ardeshir Goshtasby and S. Nikolov, "Guest editorial: Image fusion: Advances in the state of the art," *Information Fusion*, vol. 8, no. 2, pp. 114–118, 2007.