A dual-weighted polarization image fusion method based on quality assessment and attention mechanisms

Jin Duan*, Hao Zhang, Ju Liu, Meiling Gao, Cai Cheng and Guangqiu Chen

College of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, China

This paper proposes a dual-weighted polarization image fusion method based on quality assessment and attention mechanisms to fuse the intensity image (S0) and the degree of linear polarization (DoLP). S0 has high contrast and clear details, and DoLP has an outstanding ability to characterize polarization properties, so the fusion can achieve an effective complementation of superior information. We decompose S0 and DoLP into base layers and detail layers. In the base layers, we build a quality assessment unit combining information entropy, no-reference image quality assessment, and local energy to ensure the fused image has high contrast and clear and natural visual perception; in the detail layer, we first extract depth features using the pre-trained VGG19, then construct an attention enhancement unit combining space and channels, and finally effectively improve the preservation of detail information and edge contours in the fused image. The proposed method is able to perceive and retain polarization image features sufficiently to obtain desirable fusion results. Comparing nine typical fusion methods on two publicly available and own polarization datasets, experimental results show that the proposed method outperforms other comparative algorithms in both qualitative comparison and quantitative analysis.

KEYWORDS

image fusion, polarization image, double weighting, quality assessment, attention mechanisms

1 Introduction

Image fusion techniques aim to synthesize images by fusing complementary information from multiple source images captured by different sensors [1]. In recent years, many fusion methods have been proposed. According to [2], the classical fusion methods mainly include multi-scale transform-based methods [3, 4], sparse representation-based methods [5, 6], and neural network-based methods [7, 8]. Most of these methods mainly involve three key operations, such as image transformation, activity level measurement, and fusion strategy design. However, since these operations are mainly designed in a manual way, they may not be suitable for different situations, thus limiting the accuracy of the fusion results.

Fusion methods for deep learning [9, 10] have been widely studied and applied, with better fusion effects than traditional methods by virtue of their powerful feature extraction capabilities. Some scholars have shown that the combination of CNN with various traditional fusion methods not only has outstanding effects but can also effectively

reduce the workload and save computational resources by taking advantage of the migratory nature of CNN-encoded information. For example, Li et al. [11] proposed a fusion algorithm using ResNet50 and ZCA with a weighted average strategy to reconstruct the fused images, which significantly improved the fusion effect of the images. However, since they designed a simple fusion strategy, it may lead to the problem of insufficient information combination during the fusion process. Li et al. [12] used a densely connected network combined with an attention mechanism at the same time to enable the network to better capture the structural information of the source image. However, this method does not take into account the information differences at different scales, so the fused image may suffer from the loss of detailed information. Meanwhile, both methods are for the fusion task of infrared images, and the fusion effect is not satisfactory when applied directly to polarized images. At present, there are relatively few studies on polarization image fusion. Wang et al. [13] combined NSCT and CNN to propose a polarization image fusion network. Although it has some enhancement effect, conventional strategies such as weighted average and local energy are still used in the fusion process, and there is no analysis of polarization images or a more reasonable strategy design.

To solve the problems of the above methods, we propose a dual-weighted polarization image fusion method based on quality assessment and attention mechanisms. The S0 and DoLP images are decomposed into base layers and detail layers, and different strategies are constructed for fusion, respectively. Among them, S0 can reflect the spectral information of the object and is mainly used to describe the reflectance and transmittance; DoLP can reflect the difference in polarization characteristics between different material substances and provide information such as surface shape, shadow, and roughness. The fusion of S0 and DoLP can make up for the disadvantage that S0 cannot provide sufficient information in certain scenes and thus improve the target recognition ability in complex backgrounds. The implementation process of our method mainly includes: in the base layer, a quality assessment unit is constructed to achieve a balanced and reasonable fusion effect. Through comprehensive assessment and calculation of image quality, the best fusion relationship between the base layers can be obtained, and then a clear and natural fusion base layer can be obtained; in the detail layers, the depth features are first extracted using the pre-trained VGG19, and then an attention enhancement unit is constructed to enhance the polarization image detail layers from different dimensions, which can effectively combine global contextual information and improve the structural features of the fused detail layer. Using the fused base layer and the detail layer for reconstruction, the obtained fused images have rich texture details with high enough contrast and natural visual perception. The main contributions to this paper are as follows.

1. Dual-weighted fusion method. We propose a dual-weighted polarization image fusion method with strong perception and retention of features of polarization images, which is more suitable for the fusion task of polarization images.
2. Quality assessment-based fusion strategy. We construct a quality assessment unit consisting of information entropy, no-reference image quality assessment, and local energy. The

optimal fusion weight is obtained by assessing the information quality of the base layers, which is used to ensure the high contrast and natural visual perception of the fused image.
3. Attention enhancement-based fusion strategy. We construct an attention enhancement unit consisting of space and channels to enhance the global features of the detail layers in two dimensions, which can effectively improve the detail information and texture contours to obtain a fused image that fully combines intensity information and polarization characteristics.

The rest of this paper is organized as follows. Section 2 briefly reviews the research related to image fusion methods based on multiscale transforms and attention mechanisms. In Section 3, the details of our proposed method are described. Section 4 conducts experiments on the public dataset and our polarization dataset, and the experimental results are analyzed. The paper is summarized in Section 5.

2 Related work

2.1 Multiscale transform for image fusion

The fusion method based on multiscale transformation has the advantages of simplicity, efficiency, and outstanding effect compared with other methods, so it is most widely studied and applied. The main implementation process is to first decompose the source image into several different scales, then fuse the images of different scales according to specific fusion rules, then perform the corresponding multi-scale inverse transform, and finally obtain the fused image. Usually, many methods divide the image into high-frequency and low-frequency parts, or basis and detail parts. Among them, the low-frequency part and the basis part represent the energy distribution of the image, and the commonly used fusion rules include average value, local energy maximum, etc.; the high-frequency part and the detail part represent the edge and detailing features of the image, and the fusion rules include absolute value maximum, adaptive weighting, etc. Since the fusion strategy has a great influence on the fusion effect, it is important and one of the most challenging studies to design a more reasonable strategy to improve the fusion effect.

Fusion methods based on multiscale transformations have been widely studied in recent years. Wang et al. [13] proposed a fusion algorithm for polarized images. Noise removal and pre-fusion processing are performed first, and then the polarization and intensity images obtained by pre-fusion are decomposed by NSCT, and then the fusion strategies for high and low frequencies are developed separately, and finally the target fusion image is obtained by inverse transformation. Zhu et al. [14] proposed a fusion method based on image cartoon-texture decomposition and sparse representation. After decomposing the source image into cartoon and texture parts, the proposed spatial morphological structure preservation method and the sparse representation-based method are used for fusion, respectively. Zhu et al. [15] proposed a multimodal medical image fusion method. The high-pass and low-pass subbands were fused using phase congruency-based and local Laplace energy-based rules,

respectively, and the effectiveness of the proposed method was verified experimentally. Li et al. [16] proposed a fusion framework that decomposes the source image into a base part and a detail part. Among them, the base part uses a weighted average fusion strategy, and the detail part uses a maximum selection strategy to fuse the extracted multilayer depth features. Finally, the fused base and detail parts are combined to obtain a clear and natural fused image; Liu et al. [17] proposed an infrared polarization image fusion method. A multi-decomposition latent low-rank representation is used to decompose the source image into low-rank and significant parts, and different strategies are used to process the weight map, and finally the fused image is reconstructed. Hu et al. [18] proposed an improved hybrid multiscale fusion algorithm. The image is first decomposed into low-frequency and high-frequency parts using the support value transform, and then the prominent edges are further extracted from these support value images using the shearlet transform of NSST. Zou et al. [19] proposed a visible and near-infrared image fusion method based on a multiscale gradient guided edge-smoothing model and local gradient weighting, which has obvious advantages in maintaining edge details and color naturalness.

It can be found that the fusion rule of existing methods rarely analyzes the images, and most of them still use manually designed rules that do not consider the differences between images. To solve these problems, we assess the image quality and combine the attention mechanism to design two novel fusion strategies and apply them to different layers of polarization images. Among them, the quality assessment unit is applied to the base layers, which can obtain the best fusion weight based on the image quality, and the attention enhancement unit is applied to the detail layers, which can enhance the detail features by combining global contextual information.

2.2 Attention mechanism for image fusion

The attention mechanism is consistent with the human visual system and has been widely studied because it can better perceive and extract image features [20, 21]. The purpose of fusion is to combine the superior information from different images, and more weight needs to be given to salient parts during the fusion process, such as features like detailed textures and edge contours. The ability to maintain the integrity of the salient target regions using attention mechanisms can effectively improve the quality of fused images.

In recent years, many attention-based or saliency-based fusion methods have been proposed. Wang et al. [22] proposed a two-branch network based on an attention mechanism in the fusion block while using an attention model with large perceptual fields in the decoder to effectively improve the quality of fused images. Li et al. [23] proposed a generative adversarial based on multiscale feature migration and a deep attention mechanism for the fusion of infrared and visible images and achieved excellent fusion results. Liu et al. [24] designed a two-stage enhancement framework based on attention mechanisms and a feature-linking model with the advantage of being able to suppress noise effectively. Zhang et al. [25] proposed an iterative visual saliency map to retain more details of

the infrared image and calculate the weight map based on the designed multiscale bootstrap filter and saliency map, which in turn guides the texture fusion. Cao et al. [26] proposed a fusion network based on multi-scale and attention mechanisms, and the advantages of the proposed method were verified by experiments on two datasets. Wang et al. [27] proposed a multimodal image fusion framework that was designed mainly using multiscale gradient residual blocks and a pyramid split attention module.

At present, there is no polarization image fusion method that extracts features using pre-trained CNNs in a multi-scale layer while combining an attention mechanism. Specifically, we combine a deep learning framework with an attention mechanism. A pre-trained VGG19 is used to extract the depth features of the detail layers, while an attention enhancement unit consisting of space and channels is constructed to enhance the global features of the detail layers in two dimensions, which is used to improve the detail information and texture contours of the fused images.

3 Proposed method

Our method consists of three parts, and the framework is shown in Figure 1.

(1) Decomposition: the method of literature [16] is used to decompose S0 and DoLP into base layers and detail layers.
(2) Fusion: for the base layers, the fused base layer is obtained by weighting calculation using the quality assessment unit; for the detail layers, the fused detail layer is obtained by using the attention enhancement unit.
(3) Reconstruction: the fusion polarization images are reconstructed using the fused base layers and the detail layers, which can be formulated as in Eq. 1.
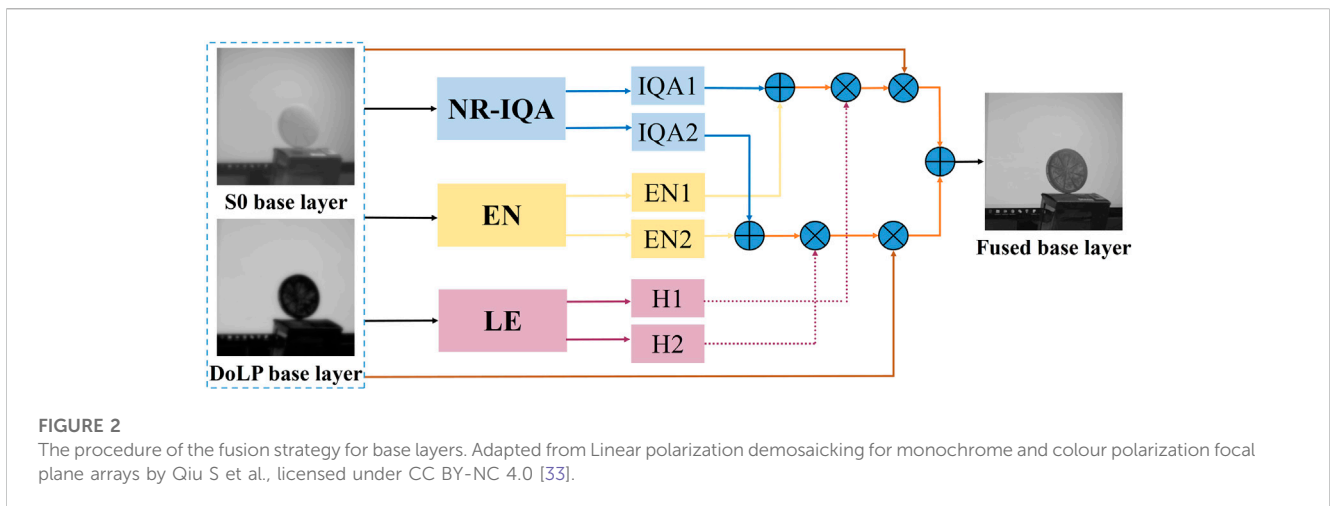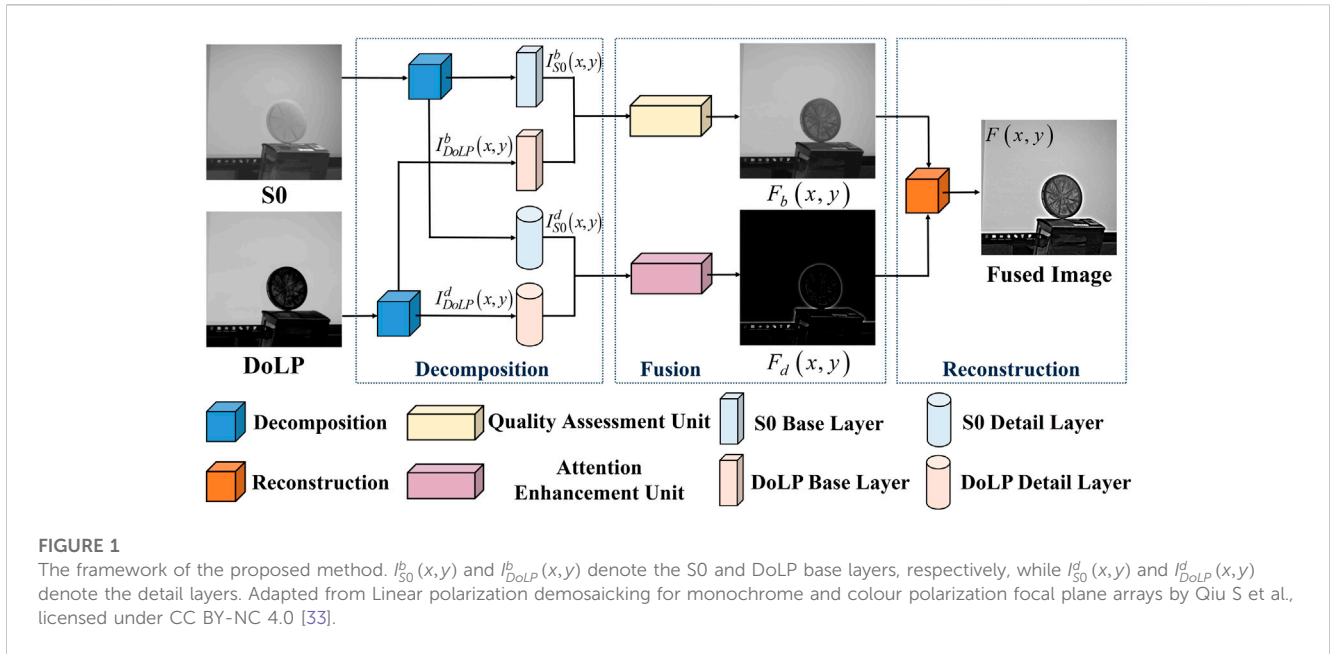
$$F(x, y) = F_b(x, y) + F_d(x, y) \qquad (1)$$

where $F_b(x, y)$ and $F_d(x, y)$ denote the fused base layer and the detail layer, respectively; and $F(x, y)$ denotes the fused image.

3.1 QAU for base layer fusion

In the fusion process of the base layers, it is necessary to consider how to reasonably retain the advantageous information of different source images. Therefore, we designed a quality assessment unit (QAU) consisting of information entropy (EN) [28], no-reference image quality assessment (NR-IQA) [29], and local energy (LE) [30]. The QAU is shown in Figure 2.

Among them, EN [31] can both reflect the amount of information in the image and serve as an evaluation index of the fused image, as shown in Eq. 2. In general, the more information the image contains, the larger the EN value. Since DoLP has more noise compared to S0, the EN value of DoLP will be higher. Therefore, it is not accurate enough if only EN is used to evaluate the image quality. Image quality assessment (IQA) can evaluate the quality of the information contained in the source image, but since high-quality original images are more difficult to obtain, we use the no-reference model (NR) instead of the full-reference model. Image quality is judged by the NR-IQA

FIGURE 1
The framework of the proposed method. $I_{S0}^b(x,y)$ and $I_{DoLP}^b(x,y)$ denote the S0 and DoLP base layers, respectively, while $I_{S0}^d(x,y)$ and $I_{DoLP}^d(x,y)$ denote the detail layers. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

FIGURE 2
The procedure of the fusion strategy for base layers. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

simulation of the visual system, which measures the degree of distortion caused by block effects, noise, compression, etc., in each semantic region of the base layers. Since S0 consists of specular and diffuse reflections, which represent the sum of the intensities of two orthogonal polarization directions, NR-IQA will calculate a higher value. Also, because DoLP has a larger microscopic surface difference, NR-IQA will judge it as a low-quality image. If only NR-IQA is used to assess image quality, the balance of information relationships cannot be accurately calculated. Therefore, we combine NR-IQA with EN to be able to estimate the weight relationship of the source image substrate more reasonably and retain more information. In addition, since LE can reflect the degree of uniformity of image gray distribution, it is used as an adjustment factor to ensure that the fused base layer has a natural visual perception. The formulas are described as Eq. 3, Eq. 4, respectively.

$$EN = -\sum_{l=0}^{L-1} P_l \log_2 P_l \qquad (2)$$

$$E_k(x, y) = \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{n=-(N-1)/2}^{(N-1)/2} w(m, n) \times [D_k(x + m, y + n)]^2 \quad (3)$$

$$H_k(x, y) = \frac{E_k(x, y)}{\sum_{k=1}^n E_k(x, y)} \qquad (4)$$

where L is the number of gray levels, which is set to 256, and P_l denotes the probability of each level. $n = 2$, denotes S0 and DoLP, respectively; $M \times N$ and $w(m, n)$ are the window area and coefficients, respectively; (x, y) denotes the pixel centroid; and $D_k(x, y)$ denotes the coefficient value of the source image at that point.

Therefore, by combining the above three quality assessment methods, the optimal weight map can be obtained, which is defined as $M_k(x, y)$. Where $EN(\cdot)$ represents the calculation of information entropy, $NR\text{-}IQA(\cdot)$ is the image quality assessment without reference, and $H_k(\cdot)$ is the adjustment factor obtained from LE. The formula is as follows.

FIGURE 3
The procedure of the fusion strategy for detail layers.
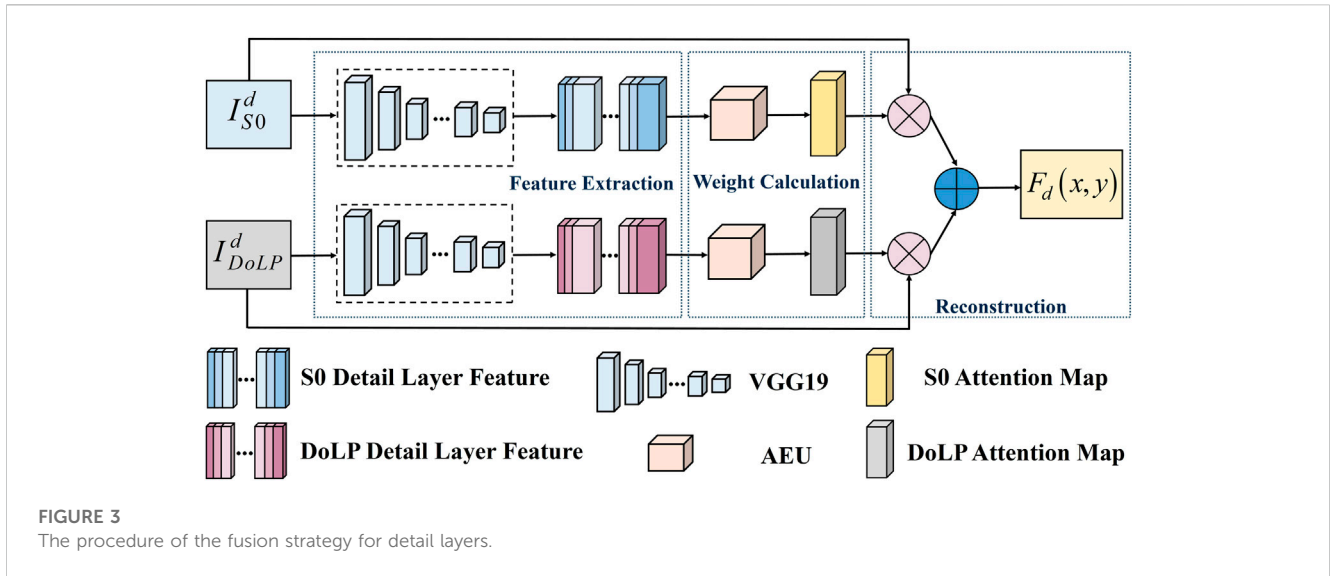
$$M_k(x, y) = H_k\big(I_k^b(x, y)\big) \cdot \big(EN\big(I_k^b(x, y)\big) + NR\text{-}IQA\big(I_k^b(x, y)\big)\big) \quad (5)$$

The base layers of S0 and DoLP are input to QAU to obtain the corresponding weight maps, and the weighting is calculated to obtain the fused base layer with the following equation.

$$F_b(x, y) = \sum_{k=1}^{n} M_k(x, y) \cdot I_k^b(x, y) \quad (6)$$

3.2 AEU for detail layer fusion

As shown in Figure 3, the detail layer fusion process consists of three main parts: feature extraction, weight calculation, and reconstruction. First, we extract the detail layer depth features of S0 and DoLP using the pre-trained VGG19. Second, the weight map is obtained using the attention enhancement unit. And then, the weight map is scaled to the source image size. Among them, the reasons for using the pre-trained VGG19 are analysed in Section 4.3.2. The weight map is obtained from the activation level map by calculating the soft-max operator, defined as Eq. 7.

$$W_k^i(x, y) = \frac{C_k^i(x, y)}{\sum_{k=1}^{n} C_k^i(x, y)} \quad (7)$$

Finally, the weighting calculation is performed to obtain the fused detail layer, which can be formulated as in Eq. 8.

$$F_d(x, y) = \sum_{k=1}^{n} W_k^i(x, y) \cdot I_k^d(x, y) \quad (8)$$

where $i \in \{1, 2, 3, 4\}$, represent $relu\_1\_1$, $relu\_2\_1$, $relu\_3\_1$ and $relu\_4\_1$, respectively.

As shown in Figure 4, the attention enhancement unit (AEU) consists of spatial attention (SA) and channel attention (CA), which aim to enhance the semantic targets and texture contours in the detail layers of the source image. The AEU can extract the feature distributions in the source image that complement each other and can generate different weights for spatial and channel features, while

the global information of the source image can be enhanced, which in turn improves the feature representation of the fused image. Among them, the SA focuses more on information such as high-frequency regions, which can enhance the details of the fused image, while the CA focuses on different channel features with completely different weighting information.

The SA consists of L1-norm and soft-max, and the formula is formulated as follows.

$$\alpha_k^i(x, y) = \frac{\big\| F_k^i(x, y) \big\|_1}{\sum_{k=1}^{n} \big\| F_k^i(x, y) \big\|} \quad (9)$$

$$\phi_k^i(x, y) = \alpha_k^i(x, y) \times F_k^i(x, y) \quad (10)$$

where $F_k^i(x, y)$ denotes the feature vector and $\| \cdot \|_1$ denotes the L1 parametric calculation.

The CA consists of a global average pooling operator $(P(\cdot))$ and soft-max, and the formula is as follows.

$$\eta_k^i(x, y) = P\big(F_k^i(x, y)\big) \quad (11)$$

$$\beta_k^i(x, y) = \frac{\eta_k^i(x, y)}{\sum_{k=1}^{n} \eta_k^i(x, y)} \quad (12)$$

$$\psi_k^i(x, y) = \beta_k^i(x, y) \times F_k^i(x, y) \quad (13)$$

The detail layer depth features of S0 and DoLP are fed into SA and CA, and the corresponding weight maps are obtained and then summed, as defined in the following equation.

$$C_k^i(x, y) = SA\big(F_k^i(x, y)\big) + CA\big(F_k^i(x, y)\big) \quad (14)$$

4 Results and analysis

4.1 Experiment settings

We compare nine algorithms on two publicly available [32, 33] datasets and our own. Among them, the public datasets are from the University of Tokyo and King Abdullah University of Science and Technology, respectively, and both contain 40 sets of polarization
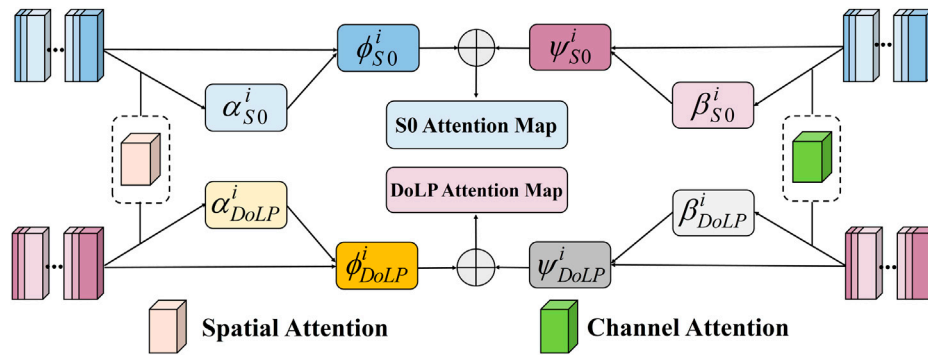
FIGURE 4
The framework of the AEU. α_k^i and β_k^i are obtained from SA and CA, respectively, while ϕ_k^i and ψ_k^i denote the enhanced features, $k = S0, DoLP$.
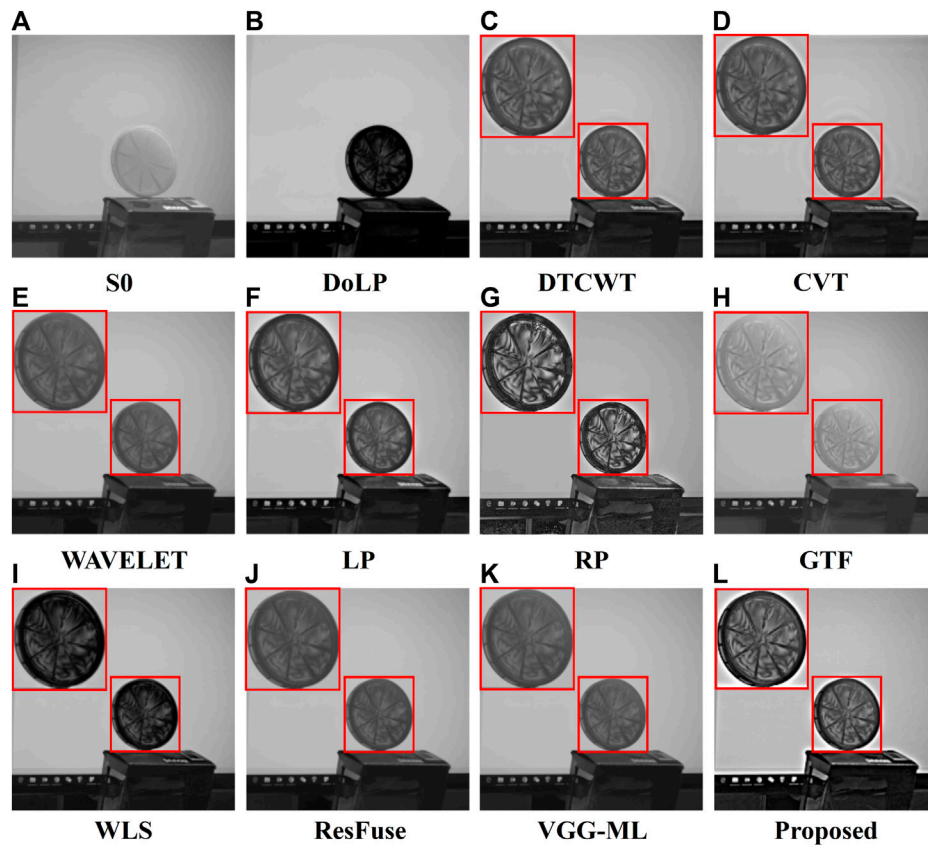
FIGURE 5
Qualitative fusion result of scene 1 on the public dataset. **(A)** S0; **(B)** DoLP; **(C)** DTCWT; **(D)** CVT; **(E)** WAVELET; **(F)** LP; **(G)** RP; **(H)** GTF; **(I)** WLS; **(J)** ResFuse; **(K)** VGG-ML; **(L)** Proposed. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

images. The experimental device is an Intel (R) Core (TM) i7-6700 CPU with 16 GB of RAM. The comparison algorithms include Dual-tree Complex Wavelet (DTCWT) [34], Curvelet Transform (CVT) [35], Wavelet [36], Laplacian Pyramid (LP) [37], Ration of Low-Pass Pyramid (RP) [38], Gradient Transfer Fusion (GTF) [39], WLS [40], ResFuse [11], and VGG-ML [16].

The evaluation metrics include information entropy (EN), spatial frequency (SF), standard deviation (SD), average gradient (AG), sum of difference correlation (SCD), and mutual information (MI). Among them, EN can reflect the information content of the fused image, SF reflects the rate of change of image grayscale, and both have a role in measuring
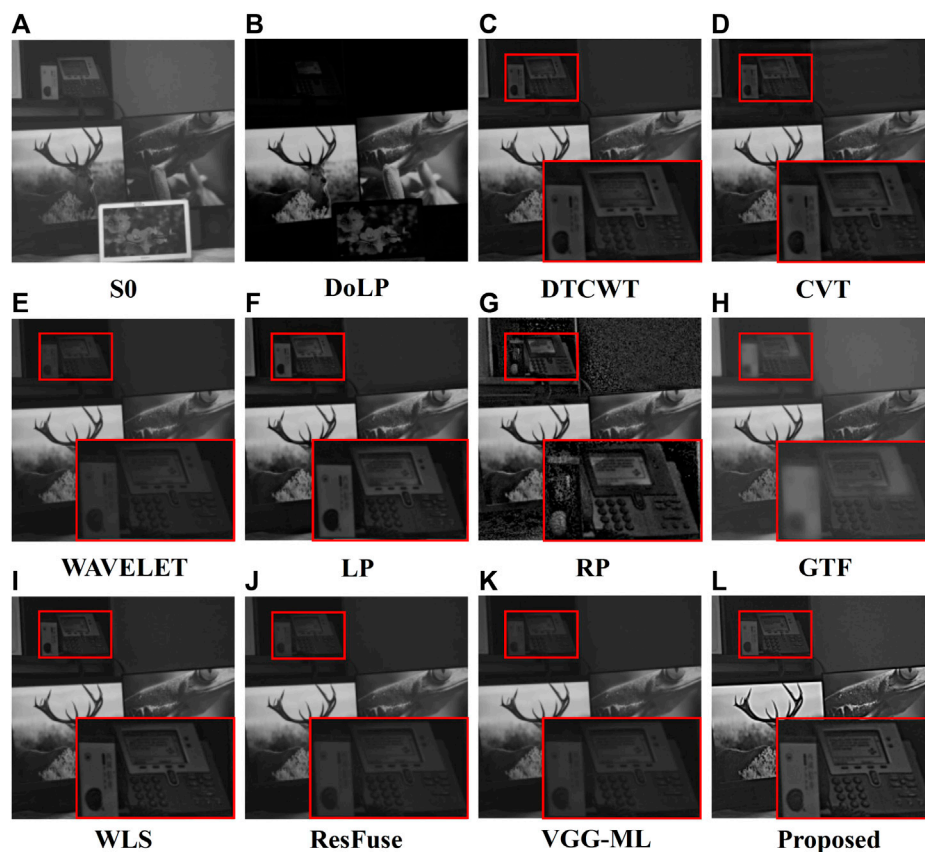
FIGURE 6
Qualitative fusion result of scene 2 on the public dataset. **(A)** S0; **(B)** DoLP; **(C)** DTCWT; **(D)** CVT; **(E)** WAVELET; **(F)** LP; **(G)** RP; **(H)** GTF; **(I)** WLS; **(J)** ResFuse; **(K)** VGG-ML; **(L)** Proposed. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

image quality with SD; AG can reflect the sharpness of the image; SCD can measure the information correlation between the fused image and the source image; and MI indicates the degree of correlation between images. Therefore, we use these metrics to comprehensively evaluate the fused image quality of different algorithms, and then verify the advantages of the proposed method.

4.2 Experimental results of the fusion algorithm

4.2.1 Results on the public dataset

The fusion results are shown in Figures 5, 6, where the key areas are marked using red boxes and enlarged. DTCWT and CVT are enhanced, but the texture details are less clear. Wavelet retains more balanced information, but the characterization effect for details is not sufficient. LP can retain the information from the source image, but the details are not clear enough. The fusion effect of RP is more outstanding, but there is a serious distortion problem, and the overall quality of the fused image is not ideal. The contrast of GTF is improved, but the focus is on retaining the information of S0, and the fused image is distorted and blurred. WLS has a fusion effect

closer to the feature distribution of DoLP but retains little information from S0, and the overall contrast of the fused image is low. The fusion effect of both ResFuse and VGG-ML is relatively clear and natural, but the effect of these two algorithms on texture detail and overall contrast enhancement is still lacking. The comparison shows that our method can better balance the information of the source image, i.e., while retaining the high contrast and clear details of S0, it can also fully combine the polarization characteristics of DoLP.

The average calculation of the dataset images using the six metrics mentioned above and the experimental results are shown in Table 1. Our method achieves the best values in five metrics, EN, SF, SD, AG, and SCD, and the index values of SF and AG are improved by 12.963% and 40.152%, respectively, compared to the maximum values in the comparison algorithm. The best values of EN indicate that the fusion results of our method can obtain more information; the best values of SF and SD indicate that the fused images have higher quality; the metric values of AG are improved substantially, which represents a clearer, more detailed texture; and the best value of SCD can indicate the better fusion performance of the proposed network. In addition, the maximum value of MI in the metrics is obtained by VGG-ML. Although the MI value of our method is not optimal, the target fused image needs to have both

TABLE 1 Quantitative comparisons of the six metrics, i.e., EN, SF, SD, AG, SCD, and MI, on the public dataset. The best results are highlighted in bold.

Methods	EN	SF	SD	AG	SCD	MI
DTCWT	5.93668	4.05239	9.67117	2.50937	1.05129	3.32856
CVT	6.03409	4.05189	9.69706	2.56217	1.05134	3.00594
WAVELET	5.60282	2.48074	9.50487	1.49367	1.05394	4.25609
LP	5.96029	3.13571	8.73245	2.05446	0.74964	2.98191
RP	6.19134	6.65251	8.74362	2.78993	0.68997	2.42422
GTF	5.75619	2.85640	10.76809	1.75516	0.80165	4.06385
WLS	5.74227	4.24776	8.63091	2.68236	0.85460	3.56277
ResFuse	6.01416	2.59245	9.59885	1.57764	1.04945	3.70814
VGG-ML	5.64989	2.64256	9.53368	1.61344	1.06090	**4.29472**
Proposed	**6.55868**	**7.51489**	**11.06179**	**3.91015**	**1.17152**	3.84981

FIGURE 7
Our polarization dataset and fusion results. Row 1 contains five S0 images, Row 2 contains five DoLP images, and Row 3 contains five fused images obtained by the proposed method.

high contrast and texture, while other algorithms do not have these features at the same time. In a comprehensive comparison, our method can highlight the edge contours of the target more effectively and is more advantageous in enhancing the image contrast and details.

4.2.2 Results on our dataset

We use a focal plane polarization camera with a Sony IMX250MZR CMOS to photograph the campus scene and construct our dataset. Partial images and fusion results are shown in Figure 7. Our dataset mainly includes buildings, trees, etc. In outdoor scenes, which is quite different from the public dataset.

Two scenes in our polarization dataset are selected and compared with the above nine algorithms. As shown in Figures 8, 9, our method has a more obvious enhancement effect for both man-made and natural objects. The green and red boxes in Figure 8 show the fusion effect on the near and far views of the building, and our method has sharper details and higher contrast than the other methods. The green and red boxes in Figure 9 show the fusion effect on natural plants, and our method also has a more natural and visual perception.

As shown in Figure 10 fused images were selected in our dataset, and line graphs were drawn using the metrics mentioned above. The abscissa of Figures 10A–F represents the image sequence, and the ordinate represents the specific value of each metric. It can be found
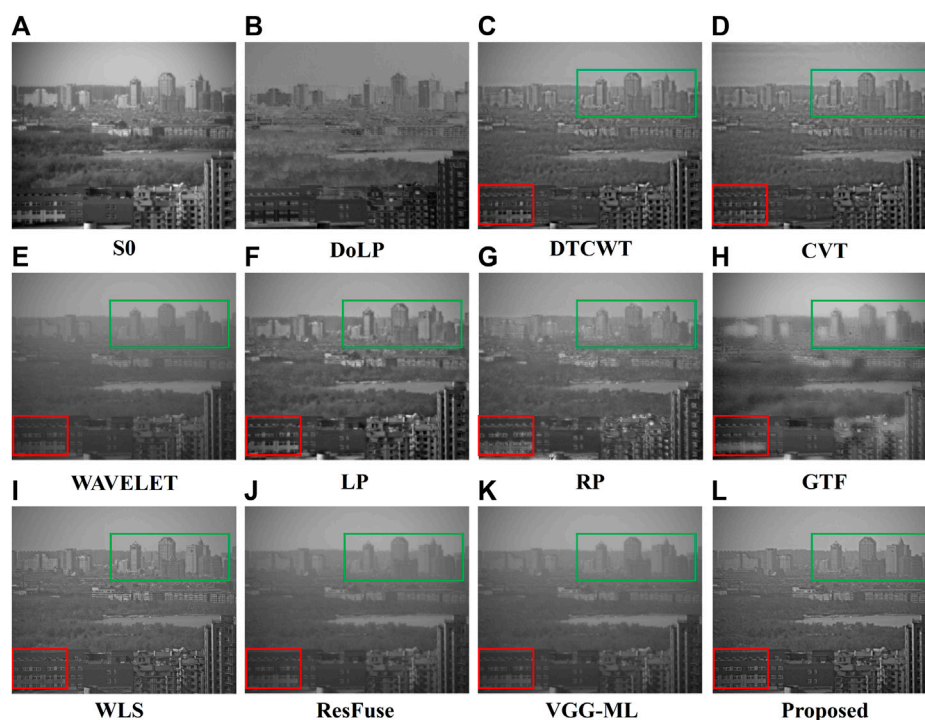
FIGURE 8
Qualitative fusion result of scene 1 on our dataset. **(A)** S0; **(B)** DoLP; **(C)** DTCWT; **(D)** CVT; **(E)** WAVELET; **(F)** LP; **(G)** RP; **(H)** GTF; **(I)** WLS; **(J)** ResFuse; **(K)** VGG-ML; **(L)** Proposed.

that our method achieves the best values for EN, SF, SD, and AG, thus verifying the outstanding advantages of our proposed method over other algorithms.

4.3 Ablation experiments

4.3.1 Ablation experiment of the QAU

Different fusion strategies are applied to the base layers without the use of an attention enhancement unit. First, the two commonly used strategies are compared; then, the fusion effect of the local energy weighting strategy is verified; and finally, EN and NR-IQA are added for experiments. Specifically, it includes: E1: average weighting strategy; E2: absolute maximum selection strategy; E3: local energy weighting strategy; and E4: QAU.

Qualitative comparisons are shown in Figure 11, where the focus areas are marked and enlarged using green and red boxes. The fusion effect of the average weighting strategy is more balanced, but the retention effect of details is not sufficient. The fused image obtained by using the absolute maximum selection strategy can retain the features of S0 better, but the information retention effect of DoLP is not ideal and does not balance the source image information reasonably. The fused images obtained by the local energy weighting strategy have more details, but the effect is still not outstanding. The fused

images obtained by using the QAU can more fully reflect the different advantageous features of S0 and DoLP, while the overall effect is more natural.

As shown in Table 2, the maximum value of EN is obtained by the local energy weighting strategy, and the optimal value of SCD is obtained by the average weighting strategy. Although our method does not achieve optimal values for these two metrics, it achieves maximum values for SF, SD, AG, and MI. These metrics can objectively reflect the polarization image fusion effect and then verify the advantages of QAU compared with other fusion rules.

4.3.2 Ablation experiment of the VGG19

Experiments were conducted using ResNet50, ResNet101, ResNet152, VGG16, and VGG19 pre-trained on the MSCOCO dataset [41], respectively, while keeping other conditions constant, and quantitative metrics were calculated.

The experimental results are shown in Table 3. The VGG series networks have better fusion results than the residual series networks in our method, and VGG19 has higher metric values than VGG16 in all metrics. Among them, VGG19 achieves the highest metric values in EN, SF, SD, and AG, while SCD and MI have the best metric values in ResNet50. Therefore, after comparing the metric values, we selected VGG19 as the feature extraction network for the detail layers.
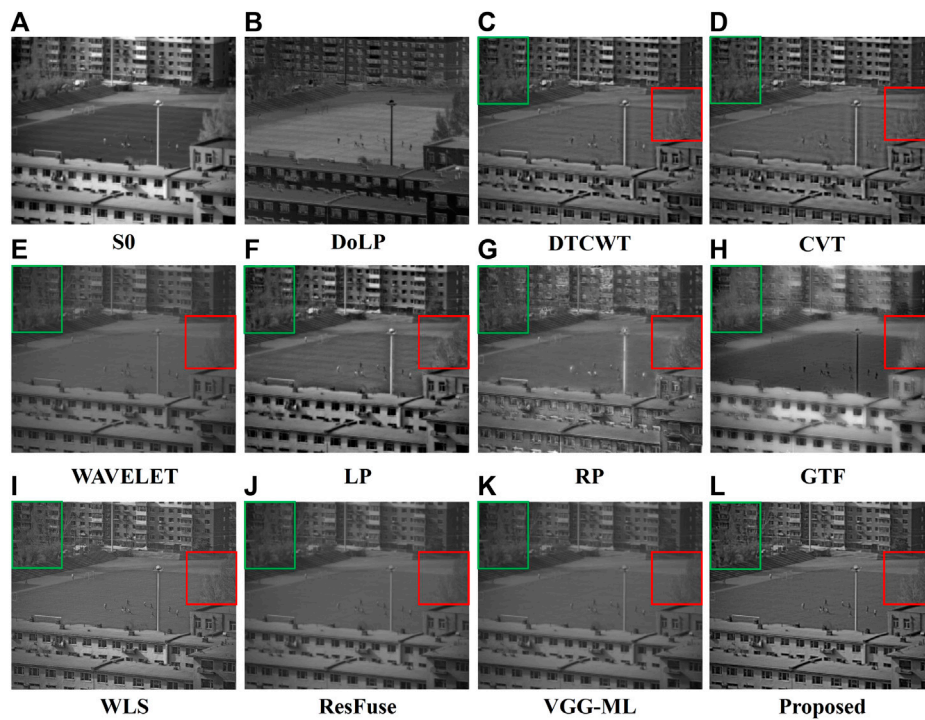
FIGURE 9
Qualitative fusion result of scene 2 on our dataset. **(A)** S0; **(B)** DoLP; **(C)** DTCWT; **(D)** CVT; **(E)** WAVELET; **(F)** LP; **(G)** RP; **(H)** GTF; **(I)** WLS; **(J)** ResFuse; **(K)** VGG-ML; **(L)** Proposed.
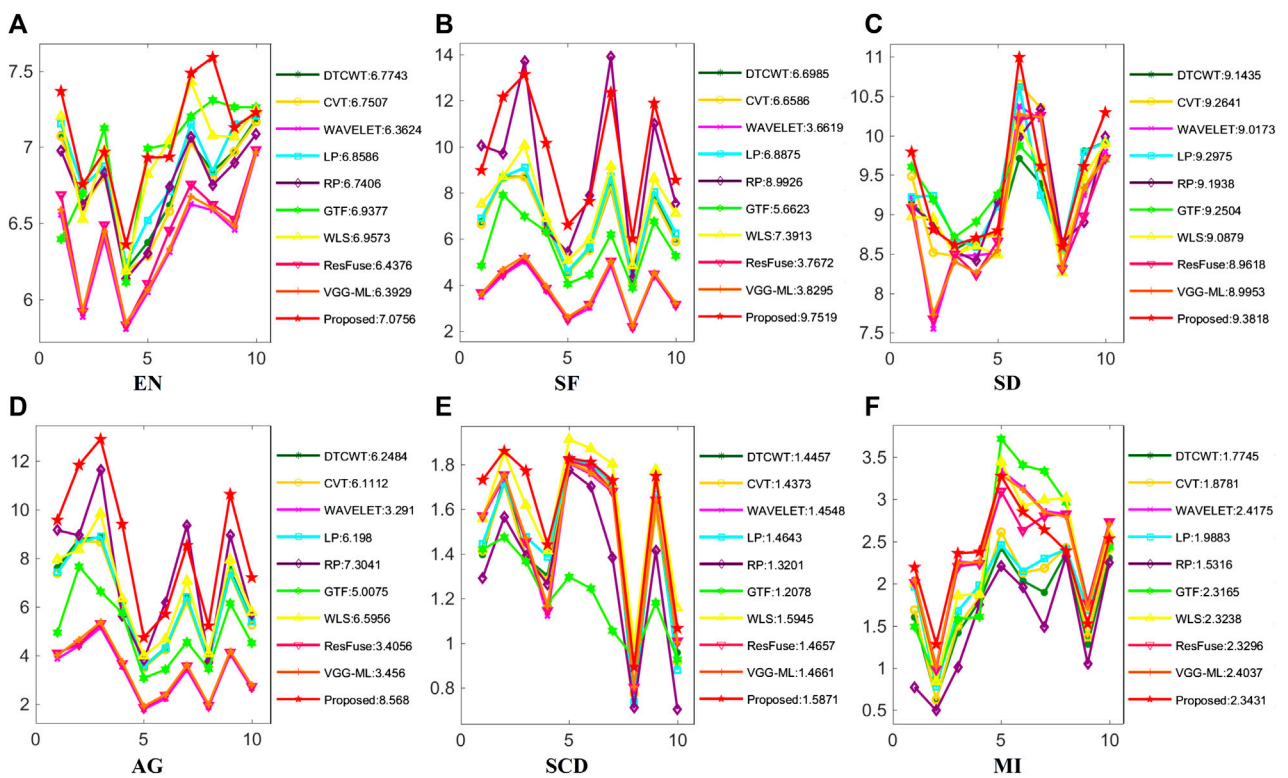
FIGURE 10
Quantitative comparison of six metrics in our dataset. **(A)** EN; **(B)** SF; **(C)** SD; **(D)** AG; **(E)** SCD; **(F)** MI.
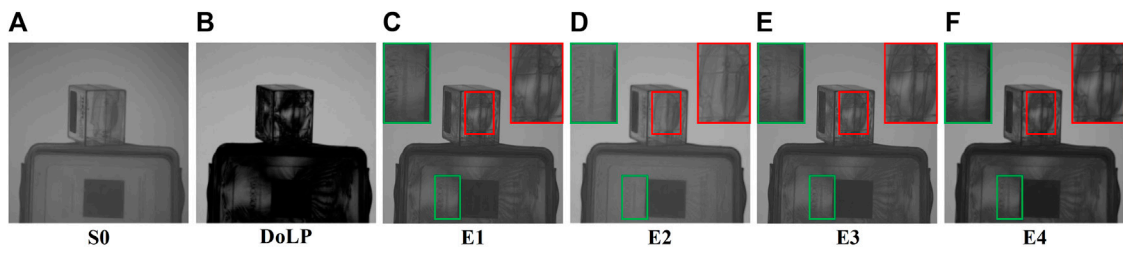
FIGURE 11
Qualitative fusion result of the QAU ablation experiment. **(A)** S0; **(B)** DoLP; **(C)** E1; **(D)** E2; **(E)** E3; **(F)** E4. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

TABLE 2 Quantitative comparison of the QAU ablation experiment. The best results are highlighted in bold.

Methods	EN	SF	SD	AG	SCD	MI
E1	5.63052	2.50456	9.51918	1.54976	**1.05928**	4.27311
E2	5.69372	2.86766	10.71943	1.76710	0.85805	4.28513
E3	**5.83124**	2.69043	9.63691	1.64626	0.94455	4.21881
E4	5.82460	**2.89027**	**10.90021**	**1.78798**	1.03466	**4.29613**

TABLE 3 Quantitative comparison of fusion results from different CNNs. The best results are highlighted in bold.

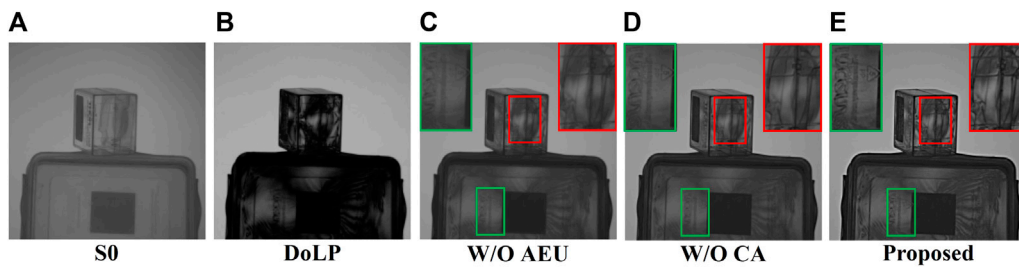| Methods	EN	SF	SD	AG	SCD	MI
ResNet50	6.15019	6.30726	11.05501	3.81852	**1.20510**	**3.57606**
ResNet101	6.14998	6.31661	11.05548	3.82214	1.20248	3.57515
ResNet152	6.15012	6.31411	11.05543	3.82139	1.20322	3.57527
VGG16	6.15518	6.49264	11.06048	3.89827	1.17482	3.55340
VGG19	**6.15587**	**6.51489**	**11.06179**	**3.91016**	1.17752	3.55981

FIGURE 12
Qualitative fusion result of the AEU ablation experiment. **(A)** S0; **(B)** DoLP; **(C)** without the AEU; **(D)** without the CA; **(E)** Proposed. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY-NC 4.0 [33].

4.3.3 Ablation experiment of the AEU

To verify the effectiveness of the AEU, the following experiments were conducted separately. First, without AEU; then, CA is added; finally, both CA and SA are used, i.e., the proposed method. Qualitative comparisons are shown in Figure 12, and some areas are marked and enlarged using green and red boxes for better observation of the experimental effect.

When without AEU, the detail information of the fused image is not prominent; when only CA is used, the texture detail and overall contrast are somewhat improved, indicating
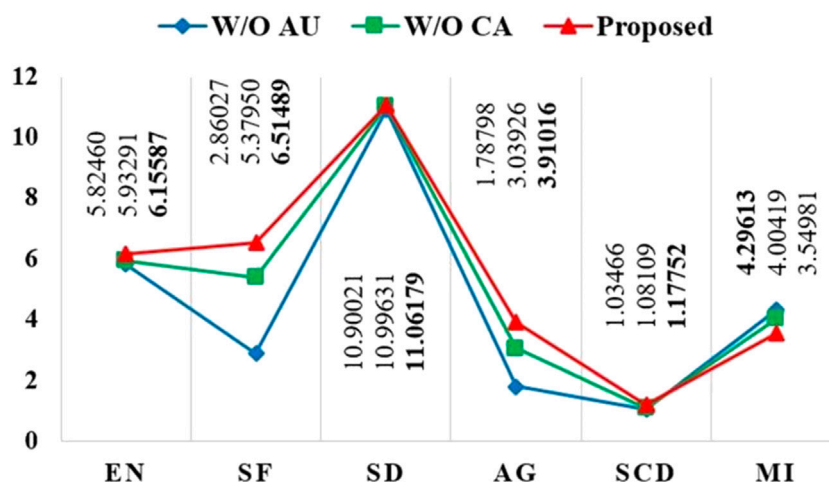
FIGURE 13
Quantitative comparison of the AEU ablation experiment. The proposed method is highlighted in red. Adapted from Linear polarization demosaicking for monochrome and colour polarization focal plane arrays by Qiu S et al., licensed under CC BY–NC 4.0 [33].

that the network can combine more information from S0 and DoLP at this time; and for the fusion result of using both CA and SA, it has a clearer texture and a more natural visual effect.

The quantitative comparison is shown in Figure 13. Compared with the fused images without AEU and with SA removed, our method has significant improvements in multiple metrics, which shows that AEU is beneficial to obtain higher-quality fused images.

5 Conclusion

This paper presents a dual-weighted polarization image fusion method that fuses S0 and DoLP to obtain a fused image of the target with high contrast and clear details at the same time. The source images are decomposed into base layers and detail layers, and the corresponding fusion strategies are designed based on quality assessment and attention mechanisms, respectively. A quality assessment unit is constructed for the fusion process of the base layers to ensure the high contrast of the fused image; a pre-trained VGG19 is used to extract the depth features of the detail layers, and a combined spatial and channel attention enhancement unit is constructed to achieve fuller preservation of scene information and texture contours to ensure the clear details and global information of the fused image. Experimental results on both public and our datasets show that the proposed method has more obvious enhancement effects in terms of contrast and detail texture for both small scene targets and complex outdoor environments, with better subjective visual perception and higher objective evaluation metrics compared to other algorithms. In future research work, we will explore how to reduce the complexity of the model while maintaining high

fusion performance and combine the angle of polarization (AoP) to achieve better fusion results.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JD and HZ responsible for paper scheme design, experiment and paper writing. JD and JL guide the paper scheme design and revision. MG, CC, and GC guide to do experiments and write papers. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

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

References

1. Liu Y, Yang X, Zhang R, Albertini MK, Celik T, Jeon G. Entropy-based image fusion with joint sparse representation and rolling guidance filter. *Entropy* (2020) 22:118. doi:10.3390/e22010118

2. Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey. *Inf Fusion* (2019) 45:153–78. doi:10.1016/j.inffus.2018.02.004

3. Chen J, Li X, Luo L, Mei X, Ma J. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition. *Inf Sci* (2020) 508:64–78. doi:10.1016/j.ins.2019.08.066

4. Li G, Lin Y, Qu X. An infrared and visible image fusion method based on multi-scale transformation and norm optimization[J]. *Inf Fusion* (2021) 71:109–129. doi:10.1016/j.inffus.2021.02.008

5. Li Y, Sun Y, Huang X, Qi G, Zheng M, Zhu Z. An image fusion method based on sparse representation and sum Modified-Laplacian in NSCT domain. *Entropy* (2018) 20:522. doi:10.3390/e20070522

6. Wang C, Wu Y, Yi Y, Zhao J. Joint patch clustering-based adaptive dictionary and sparse representation for multi-modality image fusion. *Machine Vis Appl* (2022) 33:69. doi:10.1007/s00138-022-01322-w

7. Gai D, Shen X, Chen H, Xie Z, Su P. Medical image fusion using the PCNN based on IQPSO in NSST domain. *IET Image Process* (2020) 14:1870–80. doi:10.1049/iet-ipr.2020.0040

8. Panigrahy C, Seal A, Mahato N. Parameter adaptive unit-linking dual-channel PCNN based infrared and visible image fusion. *Neurocomputing* (2022) 514:21–38. doi:10.1016/j.neucom.2022.09.157

9. Zhang J, Shao J, Chen J, Yang D, Liang B. Polarization image fusion with self-learned fusion strategy. *Pattern Recognition* (2021) 118:108045. doi:10.1016/j.patcog.2021.108045

10. Liu J, Duan J, Hao Y, Chen G, Zhang H. Semantic-guided polarization image fusion method based on a dual-discriminator GAN. *Opt Express* (2022) 24:43601–21. doi:10.1364/oe.472214

11. Li H, Wu X, Durrani T. Infrared and visible image fusion with ResNet and zero-phase component analysis. *Infrared Phys Technol* (2019) 102:103039. doi:10.1016/j.infrared.2019.103039

12. Li Y, Wang J, Miao Z, Wang J. Unsupervised densely attention network for infrared and visible image fusion. *Multimedia Tools Appl* (2020) 79:34685–96. doi:10.1007/s11042-020-09301-x

13. Wang S, Meng J, Zhou Y, Hu Q, Wang Z, Lyu J. Polarization image fusion algorithm using NSCT and CNN. *J Russ Laser Res* (2021) 42:443–52. doi:10.1007/s10946-021-09981-2

14. 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) 432:516–29. doi:10.1016/j.ins.2017.09.010

15. Zhu Z, Zheng M, Qi G, Wang D, Xiang Y. A phase congruency and local laplacian energy based multi-modality medical image fusion method in NSCT domain. *IEEE Access* (2019) 7:20811–24. doi:10.1109/access.2019.2898111

16. Li H, Li X, Kittler J. Infrared and visible image fusion using a deep learning framework. *2018 24th international conference on pattern recognition*. Beijing, China: ICPR (2018). pp. 2705–10. doi:10.1109/ICPR.2018.8546006

17. Liu X, Wang L. Infrared polarization and intensity image fusion method based on multi-decomposition LatLRR. *Infrared Phys Technol* (2022) 123:104129. doi:10.1016/j.infrared.2022.104129

18. Hu P, Wang C, Li D, Zhao X. An improved hybrid multiscale fusion algorithm based on NSST for infrared–visible images. *Vis Comput* (2023) 1–15. doi:10.1007/s00371-023-02844-8

19. Zou D, Yang B, Li Y, Zhang X, Pang L. Visible and NIR image fusion based on multiscale gradient guided edge-smoothing model and local gradient weight. *IEEE Sensors J* (2023) 23:2783–93. doi:10.1109/jsen.2022.3232150

20. Gao L, Guo Z, Zhang H, Xu X, Shen H. Video captioning with attention-based lstm and semantic consistency. *IEEE Trans Multimedia* (2017) 19:2045–55. doi:10.1109/tmm.2017.2729019

21. Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning. *IEEE Trans Knowledge Data Eng* (2021) 35:3279–98. doi:10.1109/tkde.2021.3126456

22. Wang X, Hua Z, Li J. Cross-UNet: dual-branch infrared and visible image fusion framework based on cross-convolution and attention mechanism. *Vis Comput* (2022) 1–18. doi:10.1007/s00371-022-02628-6

23. Li J, Li B, Jiang Y, Cai W. MSAt-GAN: A generative adversarial network based on multi-scale and deep attention mechanism for infrared and visible light image fusion. *Complex Intell Syst* (2022) 8:4753–81. doi:10.1007/s40747-022-00722-9

24. Liu Y, Zhou D, Nie R, Ding Z, Guo Y, Ruan X, et al. TSE_Fuse: Two stage enhancement method using attention mechanism and feature-linking model for infrared and visible image fusion. *Digital Signal Process.* (2022) 123:103387. doi:10.1016/j.dsp.2022.103387

25. Zhang S, Huang F, Liu B, Li G, Chen Y, Chen Y, et al. A multi-modal image fusion framework based on guided filter and sparse representation. *Opt Lasers Eng* (2021) 137:106354. doi:10.1016/j.optlaseng.2020.106354

26. Cao H, Luo X, Peng Y, Xie T. MANet: A network architecture for remote sensing spatiotemporal fusion based on multiscale and attention mechanisms. *Remote Sensing* (2022) 14:4600. doi:10.3390/rs14184600

27. Wang J, Xi X, Li D, Li F, Zhang G. GRPAFusion: A gradient residual and pyramid attention-based multiscale network for multimodal image fusion. *Entropy* (2023) 25:169. doi:10.3390/e25010169

28. Liu L, Liu B, Huang H, Bovik A. No-reference image quality assessment based on spatial and spectral entropies. *Signal Process. Image Commun* (2014) 29:856–63. doi:10.1016/j.image.2014.06.006

29. Bosse S, Maniry D, Müller K, Wiegand T, Samek W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans Image Process* (2018) 27:206–19. doi:10.1109/tip.2017.2760518

30. Hu S, Meng J, Zhang P, Miao L, Wang S. A polarization image fusion approach using local energy and MDLatLRR algorithm. *J Russ Laser Res* (2022) 43:715–24. doi:10.1007/s10946-022-10099-2

31. Xu H, Ma J, Le Z, Jiang J, Guo X. FusionDN: a unified densely connected network for image fusion. in Proceedings of the AAAI Conference on Artificial Intelligence, (2020). pp. 12484–91. doi:10.1609/aaai.v34i07.6936

32. Morimatsu M, Monno Y, Tanaka M, Okutomi M. Monochrome and color polarization demosaicking using edge-aware residual interpolation, in *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, (2020) pp. 2571–75. doi:10.1109/ICIP40778.2020.9191085

33. Qiu S, Fu Q, Wang C, Heidrich W. Linear polarization demosaicking for monochrome and colour polarization focal plane arrays. *Comput Graphics Forum* (2021) 40:77–89. doi:10.1111/cgf.14204

34. Lewis J, O'Callaghan R, Nikolov S, Bull D, Canagarajah R. Pixel-and region-based image fusion with complex wavelets. *Inf Fusion* (2007) 8:119–30. doi:10.1016/j.inffus.2005.09.006

35. Nencini F, Garzelli A, Baronti S, Alparone L. Remote sensing image fusion using the curvelet transform. *Inf Fusion* (2007) 8:143–56. doi:10.1016/j.inffus.2006.02.001

36. Chipman L, Orr T, Graham L. Wavelets and image fusion. Proceedings., International Conference on Image Processing, Washington, DC, USA. (1995). pp. 248–51. doi:10.1109/ICIP.1995.537627

37. Burt P, Adelson E. The laplacian pyramid as a compact image code. *IEEE Trans Commun* (1983) 31:532–40. doi:10.1109/tcom.1983.1095851

38. Toet A. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Lett* (1989) 9:245–53. doi:10.1016/0167-8655(89)90003-2

39. Ma J, Chen C, Li C, Huang J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf Fusion* (2016) 31:100–9. doi:10.1016/j.inffus.2016.02.001

40. Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys Technol* (2017) 82:8–17. doi:10.1016/j.infrared.2017.02.005

41. Lin T, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: common objects in context, in European conference on computer vision (2014) 8693: pp. 740–755. doi:10.1007/978-3-319-10602-1_48