FDD-MEF: Feature-Decomposition-Based Deep Multi-Exposure Fusion

JONG-HAN KIM1, JE-HO RYU2, AND JONG-OK KIM3, (Member, IEEE)
School of Electrical Engineering, Korea University, Seoul 02841, South Korea
Corresponding author: Jong-Ok Kim (jokim@korea.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant by the Korean Government through the Ministry of Science and ICT (MSIT) under Grant 2019R1A2C1005834, and in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program supervised by the Institute of Information Communications Technology Planning Evaluation (IITP) under Grant IITP-2021-2020-0-01749.

ABSTRACT Multi-exposure image fusion is an effective algorithm for fusing differently exposed low dynamic range (LDR) images to a high dynamic range (HDR) images. In this study, a novel network architecture for multi-exposure image fusion (MEF) based on feature decomposition is proposed. The conventional MEF methods are weak for restoring detail and color, and they suffer from visual artifacts. To overcome these challenges, a feature of each LDR image is decomposed to the common and residual components at a feature level. Then, fusion is performed on the residual domain. It was found through diverse experiments that the proposed network could improve the MEF performance in three aspects; detail restoration in bright and dark regions, reduction of halo artifacts, and natural color restoration. In addition, an attempt was made to find the underlying principles of feature-decomposition-based MEF by visualizing the features through RGB channels.

INDEX TERMS Deep multi-exposure image fusion, feature decomposition, detail restoration, color restoration, halo artifact reduction.

I. INTRODUCTION
The dynamic range one can see with naked eyes is much wider than that of commercial camera sensors. For natural scenes, the images taken with a single exposure level frequently do not have satisfying visual quality in terms of the dynamic range [1]–[3]. The low dynamic range of imaging sensors makes the visibility of scenes low in terms of image details and contrast. To solve this dynamic range issue, multi-exposure image fusion (MEF) has been widely explored by Mertens et al. [4]. MEF is a high dynamic-range (HDR) imaging technique for merging multiple low dynamic-range (LDR) images (typically more than two LDR images) with different exposure levels into a high-quality image [3], [4]. In an LDR image, visibility is significantly affected by the non-uniform illumination environment and camera exposure level. For example, detailed information in bright regions is lost with over-exposure, whereas that in dark regions is lost with under-exposure [5]–[7]. To solve these problems, many non-deep-learning-based studies have been conducted and MEF performance has been improved dramatically. Despite of the proven improvement, these methods create serious visual unnaturalness (detail or color distortion) under some bad conditions (too bright or too dark). To overcome these weaknesses, the recent researches of MEF have been conducted based on deep convolutional neural networks (CNNs) [8]–[11]. In this work, the aim was to develop a novel deep MEF method with three desirable advantages.

• Detail restoration: The detail loss caused by low contrast should be further improved sufficiently.
• Color restoration: Color distortion occurs frequently in the fused images. Natural color restoration is also required.
• Halo artifact reduction: The halo artifact is caused by the weighted fusion in particular for input images with a large brightness difference between.

These three challenges have been studied through a variety of experiments using the previous methods. In this study, a novel feature-decomposition-based deep multi-exposure image fusion (FDD-MEF) method was developed. First, in order to restore image details with high contrast, a novel deep MEF method based on the idea of common
and residual feature decomposition is proposed. It can be naturally assumed that multi-exposure images for the same scene share a common component. The other residual component (i.e., the difference between the original image and the common component) is distinct according to the exposure level. In other words, each MEF image can be decomposed into common and residual components. The common component is characterized by structural information, such as edges of objects, and the residual feature tends to include unique color and luminance information of each input source image. In this method, each MEF image is decomposed into the two components at a feature level within a deep network. Ma et al. [12] previously proposed a component based MEF method. Inspired by this, we propose component decomposition based MEF via deep learning for the first time. Unlike the conventional approach in which all images are fused, the residual features are fused, and then the fused result is added to the enhanced common component at a feature level. This is a key difference between the proposed and previous methods.

Second, for vivid color restoration, the proposed FDD-MEF is trained on the red, green, and blue (RGB) domain. In previous deep MEF methods such as [8] and [9], the input exposure image stack is first converted into the Y’CbCr space. A CNN is used to fuse the only luminance channels of the input images. This is because image structural details are present in luminance (Y) rather than chrominance (Cb and Cr). Y channels are fused through a deep network, but chrominance channels are simply fused by a weighted sum. This is one of the reasons for incorrect color restoration in a fused image.

Third, to avoid a halo artifact, the residuals of multiple input images are combined by a weighted sum, where elaborate weight maps are learned in the network. The learned weight map can contribute to natural image fusion, leading to the reduction of the halo artifact. In summary, we will evaluate and describe the performance of MEF by focusing on detail restoration, color restoration, and halo artifact reduction.

The remainder of this article is organized as follows. Section II provides a brief description of previous MEF algorithms and feature decomposition. In Section III, the proposed FDD-MEF is presented with the details of the architecture and the composition of loss functions. The experimental results are presented in Section IV. Finally, the article concludes with an insightful discussion in Section V.

II. RELATED WORKS

In this section, previous studies of MEF are described briefly, and a novel FDD-MEF idea is proposed.

A. EXISTING MEF ALGORITHMS

Here, previous multi-exposure image fusion algorithms are reviewed briefly. Many conventional algorithms rely on the weighted sum of input source images for fusion. Their main
FIGURE 2. A visual description of the flow of input multi-exposure images ($I_i$) in FDD-MEF. The subscript ‘vis’ attached to each annotation means that it has been converted from feature to RGB channels through visual mapping block for visualization.

concepts are basically similar. Weight is computed for each image patch or pixel, depending on its importance to the HDR image quality. The fused image is obtained by a weighted sum as follows.

$$X_f(i) = \sum_{k=1}^{K} W_k(i) \odot X_k(i)$$  \hspace{1cm} (1)$$

where $\odot$ denotes the Hadamard product. $W_k$ and $X_k$ represent the $k$-th weight map and the corresponding exposure image, respectively, $X_f$ is the fused image, $K$ is the number of exposure input images, and $i$ means input sequence index. Many algorithms have been proposed to improve MEF quality over the decades [4], [10], [13]–[19] since Mertens et al. [4] introduced the concept of MEF. The Laplacian pyramid method was proposed by Burt and Adelson [20]. It has made considerable contributions to many kinds of image fusion study [21]–[23], including MEF. In one study [4], the first pixel-wise exposure fusion method to combine Gaussian and Laplacian pyramids was introduced. It decomposes input source sequences into contrast, saturation, and well exposedness components for fusion. This method is resistant to the formation of halo artifacts, but it has difficulty reconstructing low- or high-light regions. In other research [24], guided image filtering was employed. In another study [18], Gaussian and Laplacian pyramids were utilized to generate a weight map for each input source sequence. However, they could not avoid halo artifacts. Ma et al. [12] attempted to decompose a set of color image patches extracted at the same spatial location of inputs into three different components. Although, this algorithm shows robust performance in handling ghosting artifacts caused by moving objects in the input sequences, it suffers from a serious artifact in restoring high-light regions. Ma et al. later improved the algorithm by optimizing a structural similarity index [25]. This algorithm further improved the restoration quality of high-light regions, but unexpected artifacts, including halos, were still observed. Meanwhile, Prabhakar et al. suggested a multi-exposure image fusion algorithm based on a CNN to obtain more conservative fused images with fewer artifacts [8]. Elsewhere [9], a multi-scale weight map was proposed for multi-exposure image fusion. It works with the Y’CbCr format and evaluates the luma component(Y) only with the MEF structural similarity index measure (SSIM) [26]. In summary, the goal of MEF algorithms is to calculate a weight $W_k$ that indicates the importance of each input $X_k$ to reconstruct the best result image $X_f$ in Eq. (1).
B. DECOMPOSITION APPROACH ON MEF

In conventional non-deep-learning works, Ma et al. [12] decomposed multi-exposure images into a couple of components, and their fusion was performed on a component basis.

\[ \hat{x} = \hat{c} \cdot \hat{s} + \hat{l}. \]  

They decomposed an input image \( \hat{x} \) into uniquely defined vectors \( \hat{c}, \hat{s}, \) and \( \hat{l} \), which represents signal strength, signal structure, and mean intensity components of \( \hat{x} \), respectively. They constructed a fused patch with the components. This approach was quite effective, and this inspired us to propose decomposition-based fusion with deep-networks for the first time. Multi-exposure images are taken on the same scene with only different exposure times. This was a key motivation idea. From this perspective, it is expected that some information about each multi-exposure image would be commonly shared. Motivated by this observation, an attempt is made to decompose each image into the common and residual components. Figure 2 presents how the input images are decomposed and flows. That can be realized with a deep-learning network, which was originally inspired by the concept of feature decomposition studies in other tasks [12], [23], [27]–[30].

III. PROPOSED METHOD

In this section, the proposed FDD-MEF method is described. It consists of four stages: feature decomposition, fusion, reconstruction [31], and visualizing (which maps the learned feature to three RGB channels for visualization, as shown in Figure 1). These stages are described in detail in the following subsections, along with a loss function.

A. FEATURE DECOMPOSITION STAGE

FDD-MEF is fed with multiple exposed images. In the first stage of the network, the feature of each input is partitioned into two components (common and residual) within the network. Multi-exposure images are acquired from the same scene with different exposure levels. Depending on exposure time, they are typically distinct in terms of brightness, contrast, and visibility. However, it is expected that there is a common component shared by multi-exposed images. For example, the common component probably includes the overall structure information such as edge component about a scene. Based on this observation, an attempt was made to decompose multi-exposed images into the common and residual components for visually high-quality image fusion. The decomposition is a challenging task, and it is learned using a deep-network. In other words, it is performed at the feature level of the network as illustrated in Figure 3.

For the proposed MEF method, the input images are first decomposed after feature extraction and then, they are fused for the residual component. The common components of all the inputs are almost identical to each other, so they do not have to be fused elaborately. They are fused through several convolution layers only for strengthening structural information (e.g., edge) in the common fusion block. Through residual weight map fusion, however, the fused image quality can be significantly improved in various aspects such as detail restoration, and halo artifacts. To ensure the accurate decomposition between common and residual components, a constraint is placed on this stage. It is an equalizing loss \( L_{cc} \) in Eq. (3) so that the common components from all the inputs should be identical.

In addition, visualized residual \( (R_{n, \text{vis}}) \) and common \( (C_{n, \text{vis}}) \) features are shown in Figure 2. They are originally in a latent feature level \( (C_n, R_n) \). However, they can be visualized in the visual mapping block on the RGB channels. They are also used to calculate the visualizing loss \( (L_{vis}) \) for the stability of the network training in Eq. (4).

B. VISUAL MAPPING STAGE

The decomposition is done at the feature level. Features extracted from the input image are decomposed into common and residual features, respectively. Thus, it is quite difficult to confirm the decomposed components visually. The common and residual features are mapped to RGB channels at the visual mapping stage. This visual mapping is also needed for stably training the feature decomposition block (FDB) in the previous subsection. Each component passes through the visual mapping block, respectively. The weights of the convolution layers in the block are shared among all features as shown in Figure 1. The visualized common and residual image components are combined into a single image, which should be identical to the input image. The reconstruction of the decomposed components is measured by the mean squared error, and this is reflected to the visualizing loss as in Eq. (3). Figure 4 is a schematized flow of the feature decomposition and visualizing mapping processes and corresponding loss functions. In this step, the equalizing loss \( (L_{cc}) \), and visualizing loss \( (L_{vis}) \) are equal to Eqs. (3) and (4), respectively.
FIGURE 4. Feature visualization and its losses for two inputs. The equalization loss ($L_{cc}$) is measured between $C_n$ at the feature level.

C. FUSION STAGE
After image decomposition, the common and residual features from all the input images are fused together in the fusion stage. A weighted-sum learned from spatial attention architecture [32] is adopted for the residual feature fusion. The contribution of each input image is different, depending on usefulness of its information for the HDR reconstruction, and its weight is learned on a pixel basis through the training process. A weight map for each input residual is learned separately for high-quality fusion. It indicates the extent of contribution to the fused image. As confirmed in the weight maps ($W_n$) of Figure 5, we can see that the learned weight maps from the attention architecture [32] are reasonable when they are compared among multiple inputs. In this process, we can avoid critical halo artifact in our FDD-MEF. However, common features are constrained to be identical to each other in the feature decomposition stage, and, consequently, they are trained to be very similar to each other. Thus the common features from all the inputs are fused by concatenation with a few of convolution blocks.

D. LOSS FUNCTION
In this subsection, loss functions that are used to train FDD-MEF are explained in detail. Specifically, three different losses are designed for stable and accurate training, and they are the common equalizing loss, visualizing loss, and reconstruction loss.

First, the consistency of the common components is regularized with a common equalizing loss term $L_{cc}$ at the decomposition stage. All the common features from multiple inputs should be equal to each other, and this is measured by

$$L_{cc} = \frac{1}{n} \sum_{i,j=1,i\neq j}^{n} (C_i - C_j)^2, \quad (3)$$

where $C_n$ represents the common feature component of the input image ($I_n$), as illustrated in Figures 1 and 2. In Eq. (3), $n$ denotes the number of input multi-exposure images. Although it is not perfect, this loss function ($L_{cc}$) induces common features to be similar by minimizing the differences among them.

Second, the visual mapping stage takes a feature visualizing loss term $L_{vis}$. It measures the reconstruction error of decomposed components. The common and residual components are combined into an RGB image, which should be equal to the original input. This loss term increases the stability of feature decomposition during the training process. The feature visualizing loss ($L_{vis}$) in Figures 1 and 4 is defined by

$$L_{vis} = \frac{1}{n} \sum_{k=1}^{n} (\tilde{I}_{n,vis} - I_k)^2, \quad (4)$$

where $\tilde{I}_{n,vis} = C_{n,vis} + R_{n,vis}$. (5)

Finally, the reconstruction loss term $L_{rec}$ works in the last reconstruction stage of the network. SSIM is used between the fused output and ground truth images as a reconstruction loss [33] as follows.

$$L_{rec} = 1 - SSIM(M_{out}, GT), \quad (6)$$

where $M_{out}$ is the fused output of FDD-MEF. Thus, the training loss $L$ is defined as the sum of the three losses as follows.

$$L = \alpha_1 L_{cc} + \alpha_2 L_{vis} + \alpha_3 L_{rec}. \quad (7)$$

VOLUME 9, 2021 164555
IV. EXPERIMENTAL RESULTS
To evaluate the performance of FDD-MEF, a variety of experiments and comparisons against state-of-the-art algorithms were conducted qualitatively for various natural images. Three types of metric for evaluation. Furthermore, the capability of the proposed feature decomposition was examined by visualizing the features separately. For a fair comparison, FDD-MEF was evaluated under the same conditions as the conventional algorithms. For example, the same number of input sequences was used for the training or inference process. Moreover, the proposed method was trained with an equal number of epochs and equal patch size to those of each algorithm compared. For all the comparison algorithms, code and settings officially provided by the original authors were used. DeepFuse [8] originally accepted two input source images (under-exposed and over-exposed sets), and its code was extended to accept four inputs for a fair comparison.

A. DATASET AND IMPLEMENTATION
To validate the performance of the proposed FDD-MEF, 270 large-scale multi-exposure scenes that are available publicly [34]–[36]. The images were taken with a digital camera with a tripod to avoid pixel movement caused by camera motions. Each scene consists of four LDR sequences with different exposure levels. These scenes include a variety of environments, such as indoor, and outdoor, day, and night, with the corresponding high-quality ground-truth images. The resolution of the sequences was reduced by $1/4$ to $1/8$ because the original sequences had too large scales to be utilized for training. For training, 227 scenes were used, and the other 43 scenes were used for testing. In the training
phase, 227 scenes (908 sequences) of multi-exposure images were employed, and the corresponding ground truth images were cropped into more than 10000 patches.

To train FDD-MEF, sequences were used that consist of four multi-exposure images and the corresponding ground truth image with a batch size 12 and patch size of $160 \times 160$. The framework was implemented with the PyTorch. The PC components used for implementation were Intel Core i7-8700 and NVIDIA TITAN RTX.

B. PERFORMANCE COMPARISON

For the evaluation, universal image sequences were chosen to cover various image characteristics. FDD-MEF was compared with seven existing MEF algorithms—Mertens09 [4], Li13 [24], GGIF [18], SPD-MEF [12], MEF-Opt [25], DeepFuse [8], and MEF-Net [9] with various visual quality metrics, such as peak signal-to-noise ratio (PSNR), SSIM, and MEF-SSIM [26]. In particular, from a qualitative comparison perspective, the focus was on detail restoration in dark and bright regions, natural color restoration, and the halo artifact occurrence caused by extreme difference in brightness between sequences. The input source images, and experimental results are shown in Figures 6 - 8.

In Figure 6, FDD-MEF is compared with the other algorithms on the input source image “Canyon”. The yellow boxed region of in Figure 6 is generally under a low light condition. Thus, a strong weight on the over-exposure image would be preferred for MEF methods. However, Mertens09 [4], SPD-MEF [12], and MEF-Net [9] have difficulty recovering details in the dark regions because they have the weakness of a strong weight on the proper sequence (i.e., a strong weight on an over-exposed sequence for a dark region). In contrast, FDD-MEF shows its merits (detail and vivid color restoration) in both yellow boxed dark regions. As shown in Figure 6-(h), the deep canyon and leaves regions in the result demonstrate well the restoration of brightness and color saturation, enhancing the visibility of the detailed shape of the rocks and leaves.

FIGURE 7. Subjective comparisons on the scene ‘Hallway’. FDD-MEF presents the best performance in terms of detail restoration.
In Figure 7, we present input source image “Hallway”. The red boxed region of the image is a detail region under saturated by excessive light. Thus, a strong weight on the under-exposure image would be better for MEF. However, SPD-MEF [12], and MEF-Net [9] suffer from poor performance in terms of detail restoration. However, the proposed FDD-MEF shows stable and natural color and the best detail restoration capability.

FIGURE 8. Subjective comparisons using the scene ‘Objects.’

Figure 8 shows the results for the “Objects” test image. FDD-MEF performed best on restoration of detail and natural color restoration of the objects in MEF image in comparison with the state-of-the-art algorithms. Especially, in red boxed bright region of the image, one can observe the obvious quality difference between the SOTA methods and FDD-MEF. Mertens09 [4] and SPD-MEF [12] are poor at recovering color on the face of the doll by introducing over-exposed color distortion. MEF-Opt [25] also shows unnatural color distortion, although it accomplishes good detail restoration. DeepFuse [8] is more grayish than the others because of lower intensity and saturation.

Finally, Figure 9 demonstrates the superiority of the FDD-MEF method in terms of visual quality. The first column shows detail restoration. The most clearly detailed pattern on the roof is in the FDD-MEF. It is thought that the learned common feature works well. The second column shows vivid color restoration. The color of the resting cat seems the most vivid in the FDD-MEF. Unlike Deepfuse [8] and MEF-Net [9], our FDD-MEF runs on RGB channels, not on the Y channel. This results in quality differences. In the third column of Figure 9, the extent of halo artifact reduction is compared. Unlike MEF-Net [9], FDD-MEF effectively reduced halo artifacts. This advantage comes from a well-learned spatial attention based weight map in the residual fusion block. In addition, the advantage in the ablation experiment is shown.

TABLE 1. Objective comparison scores (23 images were used for the test).

| Algorithm       | PSNR  | SSIM  | MEF-SSIM |
|-----------------|-------|-------|----------|
| Mertens09 [4]   | 21.33016 | 0.896121 | 0.929754 |
| Li3 [24]        | 19.19208 | 0.902759 | 0.922055 |
| GGIF [18]       | 22.56142 | **0.92513** | 0.927846 |
| SPD-MEF [12]    | 21.27363 | 0.904469 | 0.91615 |
| MEF-Opt [25]    | 21.44097 | 0.918595 | 0.918236 |
| DeepFuse [8]    | 17.20993 | 0.760576 | 0.905207 |
| MEF-Net [9]     | 21.51808 | 0.913024 | 0.92547 |
| U2Fusion [11]   | 16.8389 | 0.77954 | 0.82775 |
| FDD-MEF         | 22.7315 | 0.913093 | **0.93445** |

Objective comparisons were calculated in PSNR, SSIM [33], and MEF-SSIM [26] scores, as shown in Table 1. FDD-MEF achieved the best performance in the PSNR, and MEF-SSIM scores. In terms of SSIM, it still achieved a high score, but GGIF [18] and MEF-Opt [25] were marginally superior to FDD-MEF. However, FDD-MEF definitely works better from the perspective of subjective quality. For example, it shows better color and detail restoration capability than GGIF, and the unnatural color of the MEF-Opt result is the biggest problem of the algorithm as we have shown in Figures 6 - 8.

In summary, FDD-MEF has merits in terms of detail restoration, natural color, and fewer halo artifacts. Moreover, it achieves excellent objective comparison scores.

C. WEIGHT MAP GENERATION

In this subsection, the impact of the learned weight map on the reduction of halo artifacts is analyzed. MEF-Net [9] also
realizes a deep-based weighted fusion, and it was compared with FDD-MEF. Figure 10 shows the resulting images and the corresponding weight maps for MEF-Net [9] and FDD-MEF (residual fusion weight maps for FDD-MEF). For MEF-Net, one can observe unnatural boundary phenomena (a kind of halo artifact) in the red box region of Figure 10-(a) unlike the FDD-MEF result. Comparing the weight maps of the two algorithms could reveal the reason. In MEF-Net, the weight distribution of four inputs is characterized by sudden variations, whereas it is much smoother for FDD-MEF. In other words, the weight maps of MEF-Net have higher contrast as shown in Figure 10-(a). The biased weight of MEF-Net to a particular input leads to unnatural boundary phenomena. However, this artifact is highly reduced in FDD-MEF by learning less biased weight maps by a spatial-attention-based residual fusion block.

**FIGURE 10.** Comparison of the proposed FDD-MEF with MEF-Net [9] in the weight maps.

### D. ABLATION EXPERIMENT

Comprehensive ablation experiments were conducted to verify the contribution of each block in FDD-MEF. First, FDD-MEF was trained without a spatial attention architecture for weight map fusion in the residual fusion block (summation fusion of residual features). Next, the effectiveness of the common feature of FDD-MEF was evaluated. Lastly, we also assessed the role of the feature decomposition process by eliminating FDB (feature decomposition block) in the decomposition stage. And we additionally present quantitative results for the ablation experiments in Table 2.

1) **SUMMATION FUSION OF RESIDUAL FEATURES — EFFECTIVENESS OF RESIDUAL FUSION**

As mentioned in Section 3.3, FDD-MEF adopts the residual fusion block. To verify an effectiveness of the block, we conduct an experiment without a spatial attention residual fusion block [32]. In other words, the residual fusion is replaced with a simple summation of residual feature map. The comparison of fused results are shown in Figure 11. Compared with original FDD-MEF, Figure 11-(b) (summation fusion of residual features instead of weight map) suffers from halo artifacts which occur in the boundary area. The results demonstrate that the weight map-based fusion is effective in reducing halo artifacts.

**FIGURE 11.** Comparison between (a) the proposed with a weight map fusion and (b) summation fusion instead of a weight map. Compared with (a), one can observe strong halo artifact in (b).

2) **FEATURE DECOMPOSITION — EFFECTIVENESS OF COMMON FEATURE**

One novelty of FDD-MEF is that feature decomposition is applied to MEF. To verify its performance merits, an experiment was conducted without common features with only residual feature fusion. In this process, it was visually confirmed that the reconstructed MEF image without decomposed common features was slightly blurred compared with the FDD-MEF result. One can see the blurring in the second column of Figure 12. This means that the common feature tends to learn structural components, such as edges of objects, from each multi-exposure image more easily. In Figure 12, a common feature is shown in the first column that was fused in the fusion stage and solely reconstructed in the reconstruction stage. A common feature can be learned that is close to the actual common component among input multi-exposure images, and this contributes to less blurring in the MEF output.

3) **ELIMINATION OF FDB—EFFECTIVENESS OF FEATURE DECOMPOSITION**

To demonstrate the effectiveness of feature decomposition process, we also conducted the elimination of FDB
(feature decomposition block) and common equalizing loss $L_{cc}$ (Eq. (3)). In this experiment, we observed that detail loss in fused image of FDB elimination method. This result proves that the feature decomposition process we applied and the fusion process suitable for each quality worked effectively.

**FIGURE 12.** Effectiveness of the learned common feature. The common feature tends to learn structural edges.

---

4) QUANTITATIVE COMPARISON FOR THE ABLATION EXPERIMENTS

We have assessed the role of the feature decomposition and the corresponding fusion process, and the results are shown in Figures 11, 12, and 13. We also present the quantitative results for each ablation experiment using PSNR, SSIM, and MEF-SSIM. As listed in Table 2, it can be seen that FDD-MEF recorded the highest score in all three image quality metrics. And these superior performances are thanks to halo artifact suppression and good detail restoration.

**TABLE 2.** Average PSNR, SSIM, MEF-SSIM [26] scores of FDD-MEF and its variants.

| FDD-MEF variants                  | PSNR  | SSIM  | MEF-SSIM |
|-----------------------------------|-------|-------|----------|
| 1) Sum fusion of residual features| 22.3841 | 0.91291 | 0.92323 |
| 2) Elimination of FDB             | 22.4677 | 0.90483 | 0.92323 |
| 3) FDD-MEF (Proposed method)      | 22.7315 | 0.91369 | 0.93445 |

**FIGURE 13.** Effectiveness of feature decomposition. FDD-MEF is superior to FDB-eliminated version in terms of detail restoration.

**E. EXECUTION TIME**

We compared execution times for deep MEF methods in Table 3.

**TABLE 3.** Average execution time (AET) comparison for deep MEF methods. 47 sets of test images were used to measure the execution time.

| Algorithm          | AET per img | AET per pix |
|--------------------|-------------|-------------|
| DeepFuse [8]       | 12.49 [ms]  | 0.034 [ms/kilo-pixels] |
| MEF-Net [9]        | 54.53 [ms]  | 0.147 [ms/kilo-pixels] |
| U2Fusion [11]      | 1201.65 [ms]| 3.247 [ms/kilo-pixels] |
| FDD-MEF            | 33.40 [ms]  | 0.090 [ms/kilo-pixels] |

Networks were tested using same multi-exposure-imageset. It is implemented using PyTorch or TensorFlow framework on a same PC with 1 NVIDIA TITAN RTX 24GB GPU and Intel Core i7-8700k 3.2GHz CPU.

**V. CONCLUSION**

A novel feature-decomposition-based multi-exposure image fusion network, FDD-MEF, was proposed. For visually high-quality MEF, each input image is decomposed into two common and residual components, and they are fused at a feature level. The processes are learned during the training process. This work was originally motivated by an idea that high-quality MEF can be achieved by fusing the residual and common components of each input sequence. Through experiments, it was found that the proposed network can improve the MEF performance for detail restoration in bright and dark regions, halo artifacts reduction, and natural color restoration. Furthermore, an attempt was made to find the underlying principles of FDD-MEF by visualizing the features using RGB channels. Experimental results show that the proposed method is superior to seven state-of-the-art MEF algorithms.

**REFERENCES**

[1] M. A. Robertson, S. Borman, and R. L. Stevenson, “Dynamic range improvement through multiple exposures,” in Proc. Int. Conf. Image Process., vol. 3, Oct. 1999, pp. 159–163.
[2] E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in Proc. ACM SIGGRAPH Classes (SIGGRAPH), 2008, pp. 1–10.
[3] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. San Mateo, CA, USA: Morgan Kaufmann, 2010.
[4] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” Comput. Graph. Forum, vol. 28, no. 1, pp. 161–171, 2009.
[5] P. J. Burt, “The pyramid as a structure for efficient computation,” in Multiresolution Image Processing and Analysis. Berlin, Germany: Springer, 1984, pp. 6–35.
[6] K. Ma, T. Zhao, K. Zeng, and Z. Wang, “Objective quality assessment for color-to-gray image conversion,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 4673–4685, Dec. 2015.
[7] S. Raman and S. Chaudhuri, “Bilateral filter based compositing for variable exposure photography,” in Eurographics (Short Papers), 2009.

[8] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, “DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 4714–4722.

[9] K. Ma, Z. Duanmu, H. Zha, Y. Fang, and Z. Wang, “Deep guided learning for fast multi-exposure image fusion,” IEEE Trans. Image Process., vol. 29, pp. 2808–2819, 2020.

[10] H. Xu, J. Ma, and X.-P. Zhang, “MEF-GAN: Multi-exposure image fusion via generative adversarial networks,” IEEE Trans. Image Process., vol. 29, pp. 7203–7216, 2020.

[11] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, “U2Fusion: A unified unsupervised image fusion network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 502–518, Jan. 2022.

[12] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, “Robust multi-exposure image fusion: A structural patch decomposition approach,” IEEE Trans. Image Process., vol. 26, no. 5, pp. 2519–2532, May 2017.

[13] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, “Probabilistic exposure fusion,” IEEE Trans. Image Process., vol. 21, no. 1, pp. 341–357, Jan. 2012.

[14] Z. G. Li, J. H. Zheng, and S. Rahardja, “Detail-enhanced exposure fusion,” IEEE Trans. Image Process., vol. 21, no. 11, pp. 4672–4676, Nov. 2012.

[15] J. Shen, Y. Zhao, S. Yan, and X. Li, “Exposure fusion using boosting Laplacian pyramid,” IEEE Trans. Cybern., vol. 44, no. 9, pp. 1579–1590, Sep. 2014.

[16] K. Ma and Z. Wang, “Multi-exposure image fusion: A patch-wise approach,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2015, pp. 1717–1721.

[17] K. R. Prabhakar and R. V. Babu, “Ghosting-free multi-exposure image fusion in gradient domain,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2016, pp. 1766–1770.

[18] F. Kou, Z. Li, C. Wen, and W. Chen, “Multi-scale exposure fusion via gradient domain guided image filtering,” in Proc. IEEE Int. Conf. Multimedia Expo. (ICME), Jul. 2017, pp. 1105–1110.

[19] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, “A novel fast single image dehazing algorithm based on artificial multiexposure image fusion,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–23, 2021.

[20] P. J. Burt and E. H. Adelson, “The Laplacian pyramid as a compact image code,” in Readings in Computer Vision. Amsterdam, The Netherlands: Elsevier, 1987, pp. 671–679.

[21] W. Wang and F. Chang, “A multi-focus image fusion method based on Laplacian pyramid,” J. Comput., vol. 6, no. 12, pp. 2559–2566, 2011.

[22] H. Jung, Y. Kim, H. Jang, N. Ha, and K. Sohn, “Unsupervised deep image fusion with structure tensor representations,” IEEE Trans. Image Process., vol. 29, pp. 3845–3858, 2020.

[23] H. Yan and Z. Li, “Infrared and visual image fusion based on multi-scale feature decomposition,” Optik, vol. 203, Feb. 2020, Art. no. 163900.

[24] S. Li, X. Kang, and J. Hu, “Image fusion with guided filtering,” IEEE Trans. Image Process., vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[25] K. Ma, Z. Duanmu, H. Yegeaneh, and Z. Wang, “Multi-exposure image fusion by optimizing a structural similarity index,” IEEE Trans. Comput. Imag., vol. 4, no. 1, pp. 60–72, Mar. 2018.

[26] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[27] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep retinex decomposition for low-light enhancement,” 2018, arXiv:1808.04560.

[28] Z. Zhu, H. Yin, Y. Chai, L. Li, and G. Qi, “A novel multi-modality image fusion method based on image decomposition and sparse representation,” Inf. Sci., vol. 432, pp. 516–529, Mar. 2018.

[29] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, “Image dehazing by an artificial image fusion method based on adaptive structure decomposition,” IEEE Sensors J., vol. 20, no. 14, pp. 8062–8072, Jul. 2020.

[30] Z. Zhu, Y. Luo, G. Qi, J. Meng, Y. Li, and N. Mazur, “Remote sensing image defogging networks based on dual self-attention boost residual octave convolution,” Remote Sens., vol. 13, no. 16, p. 3104, Aug. 2021.

[31] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[32] J. Park, S. Woo, J.-Y. Lee, and I. So Kweon, “BAM: Bottleneck attention module,” 2018, arXiv:1807.06514.

[33] M. Hassan, “Structural similarity measure for color images,” Int. J. Comput. Appl., vol. 43, no. 14, pp. 7–12, 2012.

[34] [DATASET]. Accessed: Aug. 2020. [Online]. Available: https://github.com/csjcai/sice

[35] [DATASET]. Accessed: Aug. 2020. [Online]. Available: http://ivc.uwaterloo.ca/database/MEF.html

[36] J. Cai, S. Gu, and L. Zhang, “Learning a deep single image contrast enhancer from multi-exposure images,” IEEE Trans. Image Process., vol. 27, no. 4, pp. 2049–2062, Apr. 2018.