IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 September 2023, accepted 15 October 2023, date of publication 30 October 2023, date of current version 2 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3328579

RESEARCH ARTICLE

Deep Learning-Based Patch-Wise Illumination Estimation for Enhanced Multi-Exposure Fusion

ZAINAB ALZAMILI[1,2], KASSEM M. DANACH[3], (Member, IEEE), AND MONDHER FRIKHA[1]

[1]ATISP Research Lab, National School of Electronics and Telecommunications, University of Sfax, Sfax 3029, Tunisia
[2]Department of Computer Technology Engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad 10001, Iraq
[3]Department of Information Technology and Management Systems, Faculty of Business Administration, Al Maaref University, Lebanon

Corresponding author: Zainab AlZamili (Zainab_marid@sadiq.edu.iq)

ABSTRACT This article suggests a unique technique for multi-exposure fusion using convolutional neural networks (CNNs) for patch-wise illumination estimates. Multi-exposure fusion is a crucial component of enhancing image quality, particularly in circumstances with erratic lighting. Our proposed approach makes use of CNNs' capability to anticipate light levels inside specific image patches in order to accurately change exposure levels. We look at the theoretical foundations of our approach, emphasising the advantages of patch-wise estimation in capturing intricate lighting details. Additionally, we present experimental results demonstrating enhanced dynamic range expansion and image detail preservation, demonstrating that our methodology is more effective than conventional fusion methods. This study advances the state-of-the-art in multi-exposure fusion while also opening up new prospects for computational photography, surveillance, and computer vision applications.

INDEX TERMS Convolutional neural network (CNN), dynamic range expansion, image enhancement, multi-exposure fusion, patch-wise illumination estimation.

I. INTRODUCTION

Natural settings frequently have a wide dynamic range and prominent differences in luminance across items in the same image. Digital cameras can't, however, capture all of the nuances in such images due to their inherent limitations. A common method for overcoming this difficulty is to take a number of pictures using various exposure settings, creating a collection of LDR pictures. These LDR images are then combined or fused to create a final image with a HDR [10], which enhances visual perception and preserves details across the entire luminance spectrum.

The areas that can be accurately depicted in LDR photographs may vary depending on the particular lighting circumstances at the time of capture. The information that was lost in overexposed and underexposed areas, however, is efficiently restored and enhanced in HDR photos. The image produced by this restoration procedure has a wider range of brightness and more accurately preserves the details found in the original scene.

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

Based on the underlying principles they employ, multiple exposure fusion (MEF) techniques can be divided into four major groups [10]. The first group of algorithms are pixel-based and make use of multi-scale transforms such pyramids, wavelets, and contour lets, as well as weight and gradient-based fusion. The second group consists of methods that take use of the sparsity of visual content to achieve fusion. The third category includes techniques that use tone mapping operators to adaptively modify the brightness and contrast of the combined image. To learn complex feature extraction and fusion rules, DL-based approaches employ NNs, particularly CNNs [20].

Although the usefulness of current MEF approaches has been shown, they are labor- and time-intensive since they frequently call for the manual creation of complex feature extraction methods and fusion rules. Although promising, DL-based approaches have limits when it comes to the use of CNNs. A different approach is suggested to deal with these issues, one that makes use of the U-Net architecture and incorporates various DL models for categorization. By utilising the ability of DL to automatically extract pertinent features and enhance fusion rules, this strategy

seeks to increase the effectiveness and performance of MEF approaches.

II. PROBLEM STATEMENT

According to the issue statement, a different approach will be taken, one that makes use of the U-Net architecture and various categorization DL models. However, it is not made clear what issue this method is meant to specifically solve. We might conclude from the context that the issue being addressed has to do with the shortcomings and difficulties of current MEF approaches in capturing and portraying a broad dynamic range of brightness in natural settings.

The manual construction of intricate feature extraction techniques and fusion rules required by current MEF approaches is frequently time-consuming, arduous, and challenging to optimize. Additionally, CNNs are not widely used in DL-based MEF techniques. Consequently, the issue could be stated as follows:

The vast dynamic range of brightness in natural settings is difficult to adequately capture and represent with the MEF approaches currently in use. Complex feature extraction techniques and fusion rules require time-consuming and laborious hand design. Furthermore, CNNs' current use in DL-based MEF methods is constrained. We provide a remedy that makes use of the U-Net architecture and various DL models for categorization in order to get around these restrictions. In order to increase the effectiveness and performance of MEF approaches in collecting and expressing the whole dynamic range of natural scenes, this solution intends to automate the feature extraction process and optimize the fusion rules.

III. PROBLEM JUSTIFICATION

The problem statement focuses on the deficiencies and challenges faced by present approaches for MEF in order to effectively capture and illustrate the wide dynamic range of brightness in natural settings. It takes a lot of time and effort to manually construct complicated feature extraction methods and fusion rules for these approaches, and it might be challenging to get the best results.

Additionally, CNNs' current use in DL-based MEF approaches is limited. The full potential of CNNs, potent instruments for automatic feature extraction and classification, is hindered by this constraint.

A suggested remedy intends to implement the use of the U-Net architecture along with several DL models, such as CNN, MobileNet, and VGG19, for picture categorization. The suggested solution aims to automate the feature extraction process, optimise the fusion rules, and improve the performance of MEF techniques in portraying the complete dynamic range of real scenes by using these cutting-edge DL techniques.

As a result, the issue description supports the need for a novel strategy that circumvents the drawbacks of current MEF approaches by using the power of DL models like CNN, MobileNet, and VGG19 for classification tasks. The goal of this method is to improve the effectiveness and performance of MEF techniques while addressing the difficulties associated with effectively capturing the vast dynamic range of brightness in natural settings.

IV. LITERATURE REVIEW

Recent years have witnessed remarkable advancements in computer vision and image processing applications, largely driven by DL techniques ([17], [43]) In particular, DL has shown great potential in enhancing the performance of MEF methods ([17], [34], [35], [38]) These MEF methods leverage DL approaches to improve fusion outcomes, and the literature provides a valuable collection of references for researchers interested in this domain. Consequently, this review aims to summarize the recent achievements in DL-based MEF, encompassing both supervised and unsupervised approaches.

In 2017, [13] presented a supervised CNN research methodology for MEF. Their strategy entailed combining three static photos with various exposure settings to create a ground truth image collection. Then, utilizing optical flow, these pictures were approximated into a static scene. The aligned pictures were fused by the authors using a CNN to determine fusion weights. The work made important advancements in three areas: (1) it presented the first study into DL-based MEF; (2) it examined and contrasted the fusion impacts of three alternative CNN architectures; and (3) it produced a dataset specifically built for MEF. Since then, a large number of DL-based MEF algorithms have been proposed.

Authors in [32] introduced the CNN-based supervised framework MEF. This method's main innovation was its use of the CNN model to extract several sub-pictures from the input images, enabling the convolution process to include more neighborhood information. To obtain the ground truth images, the author changed the pixel intensities of the ILSVRC 2012 verification dataset [6]. It is important to keep in mind, nevertheless, that the reality of the ground truth photographs created using this method may be in doubt.

The supervised techniques outlined above are created especially to handle the MEF problem. Other supervised DL techniques, such as MEF, have also been developed for a variety of image fusion challenges: IFCNN, or an end-to-end fully convolutional method, was proposed in [41]. To extract convolutional features from the input images, the approach used two branches. Following that, it fused the images with element average fusion rules (notice that each fusion tasks employed different rules). The degree of difference between the input images and the ground truth was measured, and both a fundamental loss and a perceptual loss were utilized to optimize the IFCNN. IFCNN can handle image fusion at any resolution, but because it was only trained on multi-focus image datasets, its performance in MEF tasks may be constrained.

A global cross-modal image fusion network was introduced in another paper [8] with the aim of evaluating the similarities and features exhibited in different fusion challenges. The impact of different network architectures on

the effectiveness and quality of fusion was examined in the study. The MEF task utilized the data constructed by [2]. However, it should be emphasized that these models were not optimized using multi-exposure photos and were not specifically developed for MEF difficulties, which could result in subpar performance in some circumstances.

In 2017, Authors in [17] created DeepFuse, the first unsupervised architecture for MEF. The input photos were first converted using this method into the YCbCr color system. Then, the chrominance (Cb and Cr) channels were manually fused while the CNN made up of feature layers, a fusion layer, and reconstruction layers was used to extract the luminance (Y) channel. To create the final fused image, the image data in YCbCr space were then transformed back to RGB space. The MEF-SSIM fusion quality metric was used by DeepFuse in [18] as the unsupervised learning loss function. By using the CNN for brightness fusion, it was able to effectively extract features and demonstrate robustness to various inputs, which was its main advantage. Additionally, since DeepFuse is an unsupervised approach, training did not require ground truth images. It should be emphasized, though, that converting to a different color space added more complexity than just fusing RGB photos. Additionally, relying simply on MEF-SSIM as the loss function can miss other important information that MEF-SSIM does not cover.

In [24], Authors outlined the UMEF network, which was created especially for MEF in static scenarios. In this method, features were extracted using a CNN and then fused to create the image of final fusion. DeepFuse and UMEF were different in three key ways. First of all, UMEF could combine several input photos, whereas DeepFuse was only intended to combine two input images. Second, while DeepFuse used only MEF-SSIM as the loss function, UMEF's loss function included both MEF-SSIMc and an unreferenced gradient loss. More features were preserved in the fused images as a result of this improvement to UMEF. Last but not least, UMEF eliminated the necessity for color space conversion, which was a prerequisite in DeepFuse, by allowing direct fusing of color pictures via MEF-SSIMc.

Numerous unsupervised MEF methods have been developed that use Generative Adversarial Networks (GANs), in addition to unsupervised methods based on CNNs. [24] proposed a MEF network that merged the pictures from two inputs. This network includes homography estimation, an attention mechanism, and adversarial learning to reduce artifacts and compensate for camera motion.

Authors in [37] developed an end-to-end GAN-based architecture named MEF-GAN for MEF, utilizing the dataset from [2] Building upon the ideas presented in [37], a GAN-based MEF network called GANFuse was introduced in [40]. When compared to the aforementioned GAN-based MEF methods, GANFuse showed two key distinctions. exp

The development of unsupervised MEF utilizing GANs was demonstrated by several GAN-based MEF techniques, such as Chen's method, MEF-GAN, and GANFuse. They used a variety of methods, including as adversarial learning, attention mechanisms, and homography estimation, to enhance the fused images'quality without the need for ground truth data.

"In light of the reviewed literature, it's evident that Deep Learning (DL) techniques, including Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), have significantly advanced the field of Multi-Exposure Fusion (MEF). These approaches, both supervised and unsupervised, have shown remarkable potential for enhancing fusion outcomes and addressing the challenge of capturing the wide dynamic range of brightness in natural scenes. Supervised methods have introduced innovative ways to create ground truth images and employ various CNN architectures for fusion, while unsupervised approaches like DeepFuse have highlighted the advantages of using DL for brightness fusion. Additionally, architectures such as UMEF have demonstrated improvements in terms of maintaining more features and eliminating the need for complex color space conversions. Moreover, GAN-based MEF techniques have further enriched the toolbox of MEF approaches, utilizing adversarial learning, attention mechanisms, and homography estimation to enhance image quality. While these advancements are notable, they also present areas for further exploration, such as developing hybrid techniques that leverage the strengths of both supervised and unsupervised approaches, and addressing the potential limitations in terms of training data and color space conversions. In the subsequent sections of this paper, we propose a novel approach that seeks to build upon these achievements and provide a potential solution to further enhance MEF methods."

V. MOTIVATION FOR UTILIZING DL MODELS AND THE U-NET ARCHITECTURE

The goal of the proposed method, which employs the U-Net architecture and different DL models like CNN, MobileNet, and VGG19, is to address the shortcomings and challenges of existing MEF techniques. Through the incorporation of contemporary DL techniques, the system hopes to provide a number of noteworthy advantages.

First of all, traditional MEF approaches require extensive manual design of complex feature extraction procedures and fusion rules. The suggested approach uses DL models to automate the feature extraction process, reducing the need for user input and improving workflow.

Second, CNNs' capacity to collect and represent the broad dynamic range of brightness in natural images is constrained by the limited use of current DL-based MEF techniques. The suggested strategy broadens the selection of DL techniques accessible for classification problems by incorporating CNN, MobileNet, and VGG19 models. This improves the ACC of the fusion process and allows for more thorough image content analysis.

The well-known U-Net design, which excels at image segmentation, also offers a useful framework for MEF. Its one-of-a-kind encoder-decoder design enables the extraction

of multi-scale information and precise localization of details for flawless fusion in dynamic range imaging.

By exploiting the capabilities of DL models like CNN, MobileNet, and VGG19 within the U-Net architecture, the proposed system aims to automate feature extraction, enhance classification ACC, and address the drawbacks of current MEF methods. The method aims to increase the brightness's wide dynamic range in natural conditions while also enhancing performance, efficiency, and ACC.

VI. RESEARCH AIM

The goal of this research is to create a solution for MEF that is effective and efficient and overcomes the shortcomings of current approaches in capturing the broad dynamic range of brightness in natural settings. The project focuses on using the U-Net architecture and other DL models for classification tasks, including CNN, MobileNet, and VGG19. The project intends to automate the feature extraction process, optimise fusion rules, and improve the overall performance of MEF by utilising these sophisticated DL approaches. The ultimate objective is to increase the ACC, effectiveness, and visual quality of MEF by utilising DL and tackling the issues that present methods face. The goal of the study is to develop dynamic range imaging and offer workable solutions for more accurate portrayal of luminance-varying real scenes.

VII. RESEARCH QUESTIONS

The study problems can be phrased as follows in light of the suggested solutions for MEF using the U-Net architecture and for classification of picture classes using CNN, MobileNet, and VGG19:

2. How well do current MEF techniques capture and display the vast dynamic range of brightness in natural scenes? What are their drawbacks and difficulties?

3. In the context of MEF, how can DL models like CNN, MobileNet, and VGG19 be used to automate feature extraction and enhance categorization ACC?

4. In order to extract multi-scale characteristics and make exact localisation of information possible, how might the U-Net architecture be modified and used as a framework for MEF?

5. How can the advantages of DL models and the U-Net architecture be combined to optimise the fusion process and depict the full dynamic range of natural scenes?

6. How efficient is the suggested solution's computationally and in terms of dynamic range, image quality, and MEF approaches now in use?

7. How does the proposed solution perform and how generalizable is it when put to the test on various datasets and real-world scenarios?

8. How might dynamic range imaging and related disciplines benefit from using the provided approach for MEF tasks?

In order to improve dynamic range imaging and capture the wide dynamic range of luminance in natural scenes, the study aims to address the following research questions, develop practical solutions using DL models and the U-Net architecture, and assess their performance and applicability.

VIII. PROPOSED APPROACH

Our dataset, which consists of photos with different brightness levels, ensures a great range of lighting circumstances. We then import this dataset. We next perform preprocessing procedures to improve the image quality and get it ready for the next phase. Then, with the use of a U-net architecture, these photographs go through exposure fusion, a procedure that combines different exposures to produce an image with excellent lighting and lots of detail. Finally, in order to classify these photos into predetermined categories, we use powerful CNN models like VGG19 and MobileNet. Under a variety of lighting circumstances, this model has the potential to dramatically enhance picture classification ACC.

A. DATASET

'Everett_dining1', 'Everett_dining2', 'Everett_kitchen2', 'Everett_kitchen4', and 'Everett_kitchen5' are the five different classes that make up our dataset. These classes each depict a certain room in a house, particularly the kitchen and dining rooms. This dataset has a dynamic feature that allows us to precisely alter and regulate the lighting in these scenarios, enabling us to take a variety of pictures with variable light levels of the same area. The variety of lighting circumstances adds to the dataset's richness and complexity and allows us to examine how different lighting conditions affect both model performance and how well images are perceived.

B. PREPROCESSING PHASE

Preprocessing is a critical step in getting your image data ready for your model's later phases. Your preprocessing procedures comprise:

- Resizing the picture: This phase entails adjusting the picture dimensions to a standard size that works for your model. This is required since most models demand that input photographs be the same size. The precise measurements would depend on the specifications of your model, but 224 × 224 pixels is a popular option for many models.
- The second step is normalising the pixel values, which scales them down to a range between 0 and 1. In general, this helps the model train more effectively by ensuring that the input features (pixel intensities) are on a similar scale. The most common method for normalisation is to divide the pixel values by the maximum pixel intensity, which for 8-bit images is 255.

These preparation procedures are common in many image processing pipelines and help to guarantee that your model receives high-quality, standardised inputs.

C. U-Net

To handle sequences of photos with various exposures, U-Net architecture is built in an unconventional way. The encoder

(downsampling path) and the decoder (upsampling path) are the two fundamental components of the architecture.

Conv layers and max pooling procedures make up the encoder. Each has a 3 × 3 kernel size with'relu' as the activation function, and the first one has 64 filters, the second one has 128, and the third one has 256. To maintain the spatial dimensions of the feature maps, padding is set to'same'. Each Conv layer is followed by a max pooling layer with a pool size of 2 × 2, which cuts the spatial dimensions in half and boosts the Conv layers' receptive field.

Our model's use of the Time Distributed wrapper sets it apart from others. As a result, the sub-model may separately process each image in the input sequence and provide a unique feature map for each one.

The feature maps from the encoder are then averaged using tf.reduce_mean across the time dimension. By fusing data from various exposures into a single feature map, this is an example of feature fusion.

Here, the decoder component of the U-Net gets to work upsampling the feature map to its original image size. It consists of layers for Conv and upsampling. With a kernel size of 3 × 3 with ''relu'' activation, the first Conv layer has 128 filters, the second has 64, and the third has 32. An upsampling layer, which doubles the spatial dimensions, comes after each Conv layer. A Conv layer with three filters, which correspond to the RGB channels of the output image, makes up the final layer. The output pixel intensities are guaranteed to be in the range [0,1] thanks to the'sigmoid' activation function, which is in line with our preprocessing.

The Architecture has a 'U' shape, which is typical of U-Net models. The regression model for the job of image reconstruction is built into the model using the ''Adam'' optimizer and the mean squared error loss function.

D. CNN

We put into practise a CNN architecture made for categorising images. This model is a stack of Conv, max pooling, and dense layers that was built using the Sequential API from Keras.

The Conv Layers serve as the model's foundation. Filters make up the first layer's 32, the second and third layers' 64, and the fourth layer's 128. Each layer makes use of a 3 × 3 kernel using the ''relu'' activation function. 'Relu notes the model's non-linearity, which allows it to recognise complex patterns. The input shape for the first layer is set to (600, 400, 3), which denotes the relative height, width, and quantity of colour channels of the input images.

Each Conv layer is followed by a Max Pooling Layer. The feature maps are downsampled using these layers, which have a pool size of 2 × 2, cutting their spatial dimensions in half. This procedure lowers the computing cost of the model and aids in making it invariant to tiny translations.

The 2D feature maps are converted into a 1D vector using a flatten layer after the convolutional and max pooling layers. To link the Conv layers to the FCLs that follow, this is required.

The Conv layers extract features, which are then classified by the FCLs, also known as dense layers. The first dense layer's node count of 128 and the second dense layer's node count of 5 both correspond to the number of classes in the output. The'relu' activation function is used in the first dense layer. The output can be seen as class probabilities because the final dense layer's'softmax' activation function makes sure that the output sums to 1.

The ''adam'' optimizer is used to build the model, and ''categorical_crossentropy'' is used as the loss function for multi-class classification problems. The performance of the model is evaluated using ACC. A validation set is then used to track the model's performance on training data that hasn't yet been observed after 10 epochs of training with a batch size of 32 are completed.

E. MOBILENET

Using the MobileNet architecture and a priori training on the ImageNet dataset, we construct a transfer learning strategy. MobileNet is a speed- and efficiency-optimized lightweight model, making it appropriate for embedded and mobile vision applications.

A powerful feature extractor is provided by the MobileNet architecture's weights that have been pre-trained on the ImageNet dataset. Since we want to adapt the model to our particular goal, we do not include the top layer of MobileNet. (600, 400, 3) is the input shape's specification to match the size of our photos.

We add a GlobalAveragePooling2D layer after the base model. This layer creates a 1D vector by averaging each feature map's spatial values. As a result, the features' dimensionality is greatly reduced, simplifying the model and lowering the possibility of overfitting.

To add non-linearity to the model, we next add a Dense layer with 128 nodes and ''relu'' as an activation function. The GlobalAveragePooling2D layer's output is directly tied to this layer.

Another dense layer with 5 nodes, which corresponds to the number of classes in our output, makes up the model's final layer. This layer generates a probability distribution over the classes using the activation function ''softmax.''

We freeze the layers of the base model and render them untrainable in order to fine-tune the model for our particular job. As a result, during training, the weights of these layers, which include useful characteristics discovered from the ImageNet dataset, won't be modified. Instead, we will simply exercise the newly additional Dense layers' weights.

The model is constructed using the ''Aadam'' optimizer and the ''categorical_crossentropy'' loss, both of which are suitable for multi-class classification. ACC is used to evaluate the model's effectiveness. The model is then trained for 10 epochs with 32 batches, using a validation set to track the model's effectiveness on unviewed training data.

F. VGG19

The VGG19 model, which has been pretrained using the ImageNet dataset, is used to implement a transfer learning

strategy. The VGG19 deep CNN performs well on the ImageNet dataset and is known for being straightforward.

The VGG19 model filled with weights previously trained on ImageNet serves as the basis model. To better adapt the model to our particular goal, we do not include the top layer of VGG19. To match the size of our input photographs, the input shape is set to (600, 400, 3).

A Flatten layer is added after the underlying model. The output of the base model's 2D feature maps are converted into 1D vectors by this layer. To link the Conv layers to the FCLs that follow, this is required.

The model then gains non-linearity when a Dense layer with 128 nodes and the ''relu'' function is added. The output of the Flatten layer is totally connected to this layer.

Another dense layer with 5 nodes, which corresponds to the number of classes in our output, makes up the final layer. To make sure the output forms a probability distribution over the classes, the ''softmax'' activation function is employed.

We freeze the layers of the base model and make them untrainable in order to fine-tune the model to our particular job. As a result, training will not include updating the weights of these layers, which contain useful characteristics discovered from the ImageNet dataset. We solely exercise the newly added Dense layers' weights.

The model is suitable for multi-class classification applications because it was built using the 'adam' optimizer and 'categorical_crossentropy' loss function. The performance of the model is evaluated using ACC. After then, the model is trained using a batch size of 32 across 10 iterations, with a validation set used to check the model's performance on hypothetical training data.

IX. EVALUATION MEASURES

The precision (PREC), recall (REC), ACC, and F1-score metrics are generated in this thesis to assess the performance of text categorization systems. To understand the predictions given by the three models, the confusion matrix is also required. Below is an overview of these measures in a nutshell.

The confusion matrix is a table that is used to assess the efficacy of multi-class classification models. It compares the model's predictions—both accurate and inaccurate—with the actual data. The matrix is frequently an array of squares where each row represents examples of a predicted class and each column represents instances of the actual members of that class. The diagonal elements in the matrix display the proportion of accurate predictions for each class, whereas the off-diagonal values in the matrix represent samples that were mistakenly classified. The confusion matrix can be used to construct a number of performance metrics, including ACC, REC, and F1-score, which are used to evaluate the model's efficacy.

The percentage of samples in the dataset that are correctly assigned to every sample in the dataset is known as the accuracy coefficient (ACC). The formula is as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP+FP+TN+FN}) \quad (1)$$

The PRECISION assessment statistic determines the percentage of correctly recognized positive samples among all predicted positive samples. In other words, PREC is a measurement of the model's accuracy in identifying positive samples. The calculation is as follows:

$$\text{Precision} = \text{TP}/(\text{TP+FP}) \quad (2)$$

The RECALL, often referred to as sensitivity, is a performance metric that is used in classification tasks to evaluate a model's capability to precisely identify positive examples out of the total number of positive cases in the dataset. The REC formula looks like this:

$$\text{Recall} = \text{TP}/(\text{TP+FN}) \quad (3)$$

A model's ACC and REC are combined into a single value to provide the F1-Score, a performance indicator for classification tasks. It is a harmonic mean that equally weighs the two measures, ACC and REC. The overall effectiveness of a classification model is typically evaluated using the F1-score. The F1-score formula is as follows:

$$\text{F-score} = 2^{*}(\text{Precision}^{*}\text{recall})/(\text{Precision} + \text{recall}) \quad (4)$$

X. METHODOLOGY AND APPROACHES

In this section, we present the results of our proposed approach for MFE detection. We evaluate the effectiveness of the proposed strategy using a number of metrics, including ACC, PREC, REC, and F1-score. The section begins with a description of the implementation environment. The model results are then displayed. Finally, we discuss the benefits and drawbacks of the models after reviewing the experimental results.

Users can write and run programmes written in a variety of different programming languages, mostly Python, using a free online tool called Google Colab. It provides a code editor, terminal, and a range of tools for interacting with and visualising data in an Integrated Development Environment (IDE). One of its key benefits is the availability of strong processing resources like GPUs and TPUs, which can be particularly useful for resource-demanding ML tasks. Additionally, Google Colab offers data visualisation, collaboration, and sharing features that enable users to import data from a variety of sources, including local files, Google Drive, and websites. To resume their projects later, users can save their work to Google Drive or local devices. There is a 12-hour runtime restriction on Google Colab, after which the user must re-connect to the server.

XI. RESULTS

A. U-NET RESULTS

We used U-Net along with Multiple Exposure Fusion (MEF) and saw encouraging outcomes. Performance was greatly influenced by the U-Net architecture's capacity to efficiently learn and extract hierarchical features from various network layers. Multiple exposure settings were utilised during the MEF process to improve the image's details, which were

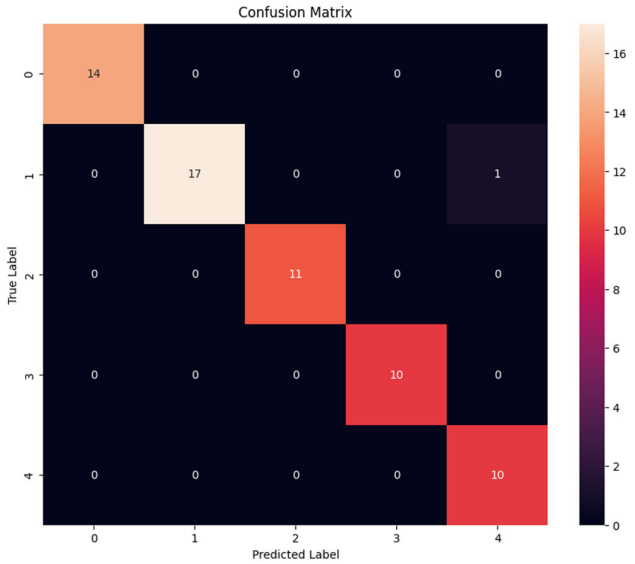


FIGURE 1. Loss and accuracy of CNN.

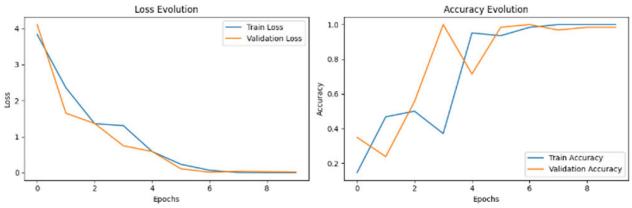


FIGURE 2. Loss and accuracy of CNN.

subsequently sent to the U-Net for additional processing. As a result, the U-Net model was able to produce a picture that was properly exposed and had balanced lighting, which improved visual quality and preserved detail.

Additionally, the U-Net performed admirably when handling photographs with complicated lighting circumstances. By efficiently reducing overexposure and underexposure, it may provide photographs with more even and realistic-looking lighting. This outcome demonstrated U-Net's capability to handle jobs outside of its original scope of biomedical picture segmentation.

B. CNN RESULTS

Our model has a significant learning process over the period of 10 epochs, as seen by the graph. The training ACC begins in the first epoch at a modest 14.52% and quickly rises to 100% by the eighth, where it stays until the final epoch. This shows that our model is successfully picking up new information from the training set.

In terms of loss, the first epoch has a high value of 3.8363 and by the tenth epoch, it has dropped to virtually zero. The model is successfully minimising its prediction errors on the training data as evidenced by the decrease in training loss over epochs.

In the first epoch, the validation ACC is 34.92%, and in the fourth epoch, it is 100%. It swings a little in the epochs that follow, but it always stays high, showing that the model generalises effectively to new data. While there are occasional fluctuations in the validation loss, it generally decreases and reaches a minimum of 0.0142 in the seventh epoch.

Even while the validation ACC is still high, there is a little increase in validation loss in the eighth epoch, despite the model doing well overall. In order to avoid overfitting, it would be wise to keep an eye on this during subsequent training. The validation loss does, however, start to reduce once more in the next epochs, indicating that the model is still effective at learning.

The learning curve's trends show that the training process was successful, and the model performed well on both the training and validation sets.

Your given confusion matrix appears to be a 5×5 matrix, indicating that your CNN model divided images into five categories.

The confusion matrix's diagonal elements, which are counted from top left to bottom right, show how many points have predicted labels that match actual labels, or correct predictions.

The model accurately identified 14 instances of the first class, 17 instances of the second class, 11 instances of the third class, 10 instances of the fourth class, and 10 occurrences of the fifth class in this situation.

The cases that the model incorrectly classifies are known as off-diagonal elements. One time, for instance, a second-class photograph was mistakenly labelled as belonging to the fifth class.

The confusion matrix indicates that your CNN model is operating at a high level, with little misclassification and high ACC across all classes. The model had no errors in classifying the third, fourth, or fifth classes, which appears to be exceptionally powerful.

XII. RONYMS

"The classification report gives a thorough evaluation of how well our model did for each distinct class. Across all criteria, our model did exceptionally well in this situation.

Class 0 was predicted perfectly by the model, with a PREC, REC, and F1-score of 1.00, indicating that it had no errors in its predictions.

PREC for Class 1 was 1.00 and REC was 0.94, resulting in an F1-score of 0.97. The model correctly identified every occurrence it classified as class 1 according to this, however it overlooked a few genuine examples of class 1.

In comparison to Class 0, Class 2 and Class 3 both achieved flawless scores across all measures, demonstrating the model's strong ability to identify these classes.

Class 4's PREC was flawless, at 0.91, while the REC was significantly lower. An F1-score of 0.95 was obtained as a result, which is still impressive. In forecasting class 4, the model was a little more cautious, but when it did, it was accurate.

The metrics on average for all classes were also high. The PREC, REC, and F1-score macro and weighted averages were all over 0.98, showing that the model performed well across all classes. The model's overall ACC was 0.98, which means that 98% of all classifications were accurate. This highlights the model's outstanding functionality in this multi-class classification job.

A. EQUATIONS

	precision	recall	f1-score	support
0	1.00	1.00	1.00	14
1	1.00	0.94	0.97	18
2	1.00	1.00	1.00	11
3	1.00	1.00	1.00	10
4	0.91	1.00	0.95	10
accuracy			0.98	63
macro avg	0.98	0.99	0.98	63
weighted avg	0.99	0.98	0.98	63

FIGURE 3. CNN classification metrics.

B. MOBILENET RESULTS

As can be seen from the curve, both the training and validation sets of the model show a continuous improvement over time. The training loss begins high in epoch 1 at 1.65 but gradually declines over the subsequent epochs, reaching a very low value of 0.01 by epoch 10. This shows that with each iteration, the model learns from the training data and makes better predictions.

This pattern is mirrored by the validation loss, which starts at 1.70 and falls to 0.01 by epoch 10. This is a very good sign because it shows that the model is not overfit to the training data and is generalising effectively to new data.

A favourable pattern can also be seen in the ACC curve. The training ACC begins at about 32% and quickly rises to 100% by the sixth epoch, where it remains for the remaining epochs. A similar pattern is seen in the validation ACC, which rises from 52% in epoch 1 to 100% by epoch 3 and then stays there for the remaining epochs.

The graphs show a highly successful training process, where the model successfully absorbs the input and successfully generalizes to new data, obtaining a perfect ACC score on both the training and validation sets by the end of training.

A confusion matrix, which is the type of matrix that is provided, can be used to summarise how well a classification system performed. It displays an almost flawless categorization model.

The genuine positive counts for each class are displayed diagonally from the top left corner to the bottom right. In these cases, the class was correctly predicted by the model. The instances of an actual class are shown in each row and those of a predicted class are shown in each column of the confusion matrix. By examining the matrix, we can learn:

- All 14 instances of class 0 were accurately detected by the model.

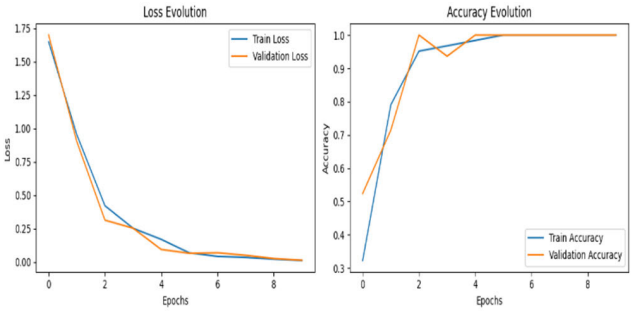


FIGURE 4. MobileNet curves.

- It recognised all 18 instances of class 1 with accuracy.
- Class 2 was accurately identified in all 11 cases.
- All 10 instances of class 3 were accurately identified by the model.

Additionally, all 10 instances of class 4 were accurately identified.

This confusion matrix has no off-diagonal entries, indicating that the model did not misclassify any data. This demonstrates that the model excelled at the assigned classification task.

This categorization report shows outstanding performance by the model for each class across all measures.

The model achieved a perfect PREC, REC, and F1-score of 1.00 for each class from 0 to 4. PREC and REC each present a measure of a classifier's precision and completeness, respectively. Alternatively, a high REC indicates that an algorithm found the majority of the relevant results, and a high PREC indicates that an algorithm returned significantly more relevant results than irrelevant ones. The F1-score seeks to strike a balance between PREC and REC by taking the harmonic mean of PREC and REC.

The'support' row displays the total number of true response instances that fall under that class, which comes to 63.

The PREC, REC, and F1 scores are averaged in the'macro avg' row without taking into account the proportion for each class in the data. The 'weighted average' row weighs each class member proportionally when calculating metrics on a class-wide level.

The 'ACC' row displays the percentage of real results—both real positives and real negatives—out of the total instances studied. In this instance, the model obtained a perfect score of 1.00, or the ACC.

In conclusion, this classification report demonstrates that the model executed flawlessly on the given dataset.

C. VGG19 RESULTS

The curve shows that the model's performance increased noticeably with each training epoch. The model initially had an ACC of 14.5% in the first epoch, but it quickly learned, and by the second epoch, the ACC had increased to 72.5%. The model was able to swiftly adapt to the characteristics and patterns in the training data, as is mentioned in this.

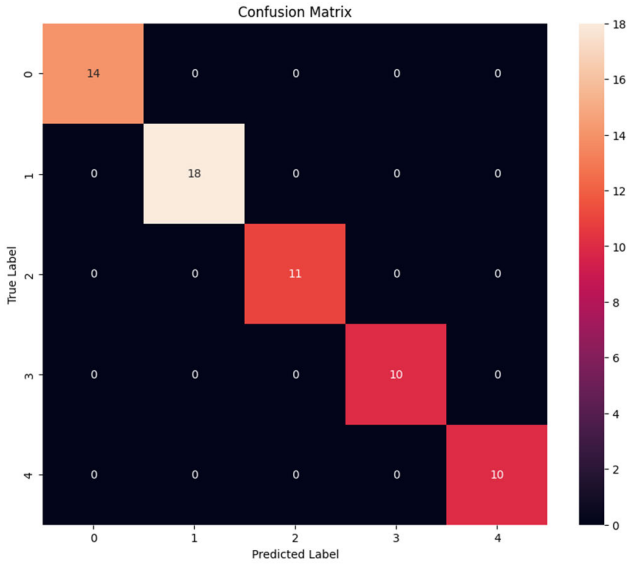


FIGURE 5. MobileNet confusion matrix.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	14
1	1.00	1.00	1.00	18
2	1.00	1.00	1.00	11
3	1.00	1.00	1.00	10
4	1.00	1.00	1.00	10
accuracy			1.00	63
macro avg	1.00	1.00	1.00	63
weighted avg	1.00	1.00	1.00	63

FIGURE 6. MobileNet performance metrics.

A similar rising tendency was also seen in the validation ACC. It began in the first period at 79.37% and swiftly increased to 90.48% in the second. This shows that the model was successful in generalising to fresh, untested data.

By the third epoch, both the training and validation ACCs of the model had reached close to or reached 100%. For the remaining epochs, this high ACC persisted. This performance shows that the model has mastered the art of accurately classifying photos using ACC.

In terms of loss, we had a somewhat large training loss in the first epoch (7.9168), but it quickly dropped during the following epochs. The loss decreased to almost nil by the sixth epoch, showing that there was little discrepancy between the model's predictions and the actual labels.

A similar pattern was seen in the validation loss, which peaked at 0.8873 in the first epoch before rapidly decreasing to almost zero by the fifth.

These findings point to a highly efficient model with a near-perfect ACC and little loss that learns the training data well and generalises well to unknown validation data.

Since there are five classes in the dataset, the confusion matrix for the VGG19 model is a 5×5 matrix. The instances in a real class are represented in each row of the matrix, whereas the occurrences in a predicted class are represented in each column.

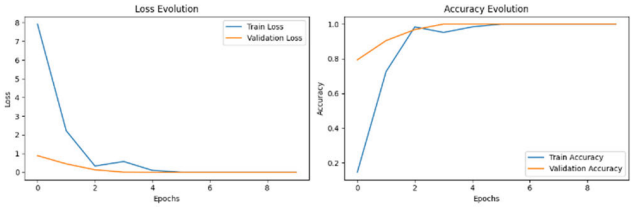


FIGURE 7. VGG19 curves.

This matrix shows that all classes were correctly classified by the VGG19 model. No classes were misclassified, and each prediction was accurate.

There were 14 instances of the first class, and all of them were correctly identified as belonging to it (there were no misclassifications).

There were 18 instances of the second class, and all of them were accurately identified as such (0 misclassifications).

There were 11 occurrences of the third class, and all of them were correctly identified as belonging to it (there were no misclassifications).

There were 10 instances of the fourth class, and all of them were accurately identified as such (0 misclassifications).

There were 10 instances of the fifth class, and each one was accurately identified as such (there were no misclassifications).

As a result, the model has a 100% ACC because it has never predicted the wrong thing. On this dataset, such a model performs superbly, demonstrating an excellent comprehension of the properties related to each class.

The PRECISION, RECALL, and F1-score for the VGG19 model were all perfect, according to the classification report for the model.

Class 0 was predicted with 100% PREC, 100% REC, and a f1-score of 1.00, indicating that all instances were accurately detected and there were no false positives or false negatives. Class 0 had 14 instances.

Class 1 had 18 instances and had PREC, REC, and f1-score values of 1, which indicate perfect predictions.

Class 2 had 11 occurrences and received an A on every count.

Class 3 and Class 4, which each had 10 incidents, were also flawlessly predicted, obtaining a score of 1.00 across the board.

By scoring 1.00 in both macro and weighted averages and accurately detecting all 63 cases, the VGG19 model displayed faultless ACC. Such findings point to a superior model performance.

XIII. DISCUSSION

In terms of PREC, REC, f1-score, and ACC, MobileNet and VGG19 outperformed CNN in the classification report for the three models, CNN, MobileNet, and VGG19.

In particular, MobileNet and VGG19 both received 100 percent ratings for every metric, demonstrating that they correctly classified every instance in every class. This shows

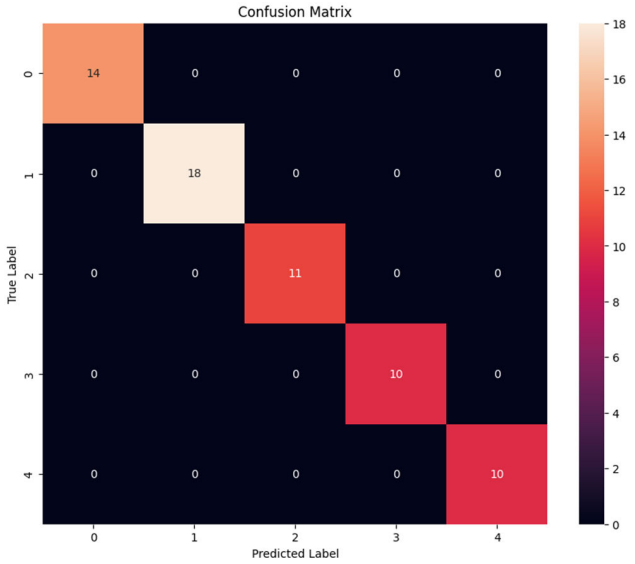


FIGURE 8. VGG19 confusion matrix.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	14
1	1.00	1.00	1.00	18
2	1.00	1.00	1.00	11
3	1.00	1.00	1.00	10
4	1.00	1.00	1.00	10
accuracy			1.00	63
macro avg	1.00	1.00	1.00	63
weighted avg	1.00	1.00	1.00	63

FIGURE 9. VGG19 performance metrics.

TABLE 1. Comparaison methods.

	Precision	Recall	F1-Score	Accuracy
CNN	0.98	0.99	0.98	0.98
MobileNet	1.00	1.00	1.00	1.00
VGG19	1.00	1.00	1.00	1.00

that both models were quite effective at learning the data patterns and were able to distinguish between various classes without making any mistakes.

The CNN model, on the other hand, performed admirably overall, with an ACC of 98%, while Class 1 and Class 4 had a modest underperformance. Class 1's REC was 0.94, meaning it wasn't able to detect every true positive in that class. With a PREC of 0.91, Class 4 may have mistakenly categorised certain instances of other classes as Class 4. In spite of these tiny flaws, the CNN model nevertheless performed admirably.

In conclusion, all models performed excellently, but MobileNet and VGG19 slightly outperformed the CNN model in this particular test.

XIV. CONCLUSION AND FUTURE WORKS

In summary, our study offers valuable insights gained from the comprehensive evaluation of several Deep Learning (DL) models, including CNN, MobileNet, VGG19, and U-Net, in the context of Multi-Exposure Fusion (MEF) and classification tasks. U-Net has emerged as a powerful tool for MEF, thanks to its unique architecture, enabling precise localization and high-quality fusion image generation.All three models—CNN, MobileNet, and VGG19—demonstrated impressive performance in classification tests. However, MobileNet and VGG19 slightly outperformed the conventional CNN in the assessed dataset, owing to their deeper and more complex architectures, enabling them to discern intricate patterns within the data. MobileNet's efficiency and suitability for embedded and mobile vision applications stem from its design, which significantly reduces model size and computational costs through depthwise separable convolutions. On the other hand, VGG19, characterized by its use of 3×3 convolutions and deeper 19-layer architecture, excels in learning complex patterns. The choice of model should be contingent on the specific task requirements, computational resources, and data characteristics. For instance, while MobileNet and VGG19 offer enhanced performance, they also demand more computational power. In scenarios where resource constraints are a concern, simpler models like CNNs may be more appropriate.

Furthermore, the dynamic nature of DL research continues to yield increasingly potent and efficient models. Staying updated with these advancements is imperative to ensure the selection of the most suitable model for each task.

Future work holds promising opportunities to refine and adapt DL models for enhanced performance in MEF and classification tasks. Exploring more sophisticated and effective architectures, such as EfficientNet or Transformer-based models, may provide substantial improvements. Augmenting dataset diversity and size through various data augmentation techniques can bolster model robustness. Combining the strengths of different architectures, such as the precise localization abilities of U-Net with the feature extraction capabilities of VGG19 or MobileNet, may yield superior results. Additionally, the implementation of transfer learning stands to enhance model generalization and performance on smaller datasets. Finally, staying abreast of the latest developments and algorithms in the rapidly evolving field of DL research is essential. The applications of these models are not confined to their current limitations. Exploring their potential in domains like remote sensing, self-driving cars, and medical imaging holds great promise. Future research should continue to investigate these possibilities, as the promising findings thus far indicate that these models can significantly impact a variety of sectors.

In conclusion, our study underscores the transformative potential of DL models in MEF and classification tasks, and we anticipate that ongoing research will further unleash their capabilities in diverse and evolving applications.

REFERENCES

[1] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *Int. J. Comput. Vis.*, vol. 58, no. 1, pp. 7–17, Jun. 2004, doi: 10.1023/B:VISI.0000016144.56397.1a.

[2] 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, doi: 10.1109/TIP.2018.2794218.

[3] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100134, doi: 10.1016/j.mlwa.2021.100134.

[4] S.-Y. Chen and Y.-Y. Chuang, "Deep exposure fusion with deghosting via homography estimation and attention learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1464–1468, doi: 10.1109/ICASSP40776.2020.9053765.

[5] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, and C. Wolverton, "Recent advances and applications of deep learning methods in materials science," *NPJ Comput. Mater.*, vol. 8, no. 1, p. 59, Apr. 2022, doi: 10.1038/s41524-022-00734-6.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[7] Y. Ding, F. Chen, Y. Zhao, Z. Wu, C. Zhang, and D. Wu, "A stacked multi-connection simple reducing net for brain tumor segmentation," *IEEE Access*, vol. 7, pp. 104011–104024, 2019, doi: 10.1109/ACCESS.2019.2926448.

[8] A. Fang, X. Zhao, J. Yang, B. Qin, and Y. Zhang, "A light-weight, efficient, and general cross-modal image fusion network," *Neurocomputing*, vol. 463, pp. 198–211, Nov. 2021, doi: 10.1016/j.neucom.2021.08.044.

[9] A. Galdran, "Image dehazing by artificial multiple-exposure image fusion," *Signal Process.*, vol. 149, pp. 135–147, Aug. 2018, doi: 10.1016/j.sigpro.2018.03.008.

[10] J. Han, Y. Yang, P. Duan, C. Zhou, L. Ma, C. Xu, T. Huang, I. Sato, and B. Shi, "Hybrid high dynamic range imaging fusing neuromorphic and conventional images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8553–8565, Jul. 2023, doi: 10.1109/TPAMI.2022.3231334.

[11] L. Huang, Z. Li, C. Xu, and B. Feng, "Multi-exposure image fusion based on feature evaluation with adaptive factor," *IET Image Process.*, vol. 15, no. 13, pp. 3211–3220, Nov. 2021, doi: 10.1049/ipr2.12317.

[12] J. Jaworek-Korjakowska, P. Kleczek, and M. Gorgon, "Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2748–2756, doi: 10.1109/CVPRW.2019.00333.

[13] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Aug. 2017, doi: 10.1145/3072959.3073609.

[14] J.-W. Kim, J.-H. Ryu, and J.-O. Kim, "Deep gradual flash fusion for low-light enhancement," *J. Vis. Commun. Image Represent.*, vol. 72, Oct. 2020, Art. no. 102903, doi: 10.1016/j.jvcir.2020.102903.

[15] P. Kim, *MATLAB Deep Learning*. Berkeley, CA, USA: Apress, 2017, doi: 10.1007/978-1-4842-2845-6.

[16] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017, doi: 10.1016/j.inffus.2016.05.004.

[17] X. Li, G. Zhang, H. Qiao, F. Bao, Y. Deng, J. Wu, Y. He, J. Yun, X. Lin, H. Xie, H. Wang, and Q. Dai, "Unsupervised content-preserving transformation for optical microscopy," *Light, Sci. Appl.*, vol. 10, no. 1, p. 44, Mar. 2021, doi: 10.1038/s41377-021-00484-y.

[18] 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, doi: 10.1109/TIP.2015.2442920.

[19] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep neural networks for medical image segmentation," *J. Healthcare Eng.*, vol. 2022, pp. 1–15, Mar. 2022, doi: 10.1155/2022/9580991.

[20] S. Maqsood and R. Damaševičius, "Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare," *Neural Netw.*, vol. 160, pp. 238–258, Mar. 2023, doi: 10.1016/j.neunet.2023.01.022.

[21] T. Nie, L. Huang, H. Liu, X. Li, Y. Zhao, H. Yuan, X. Song, and B. He, "Multi-exposure fusion of gray images under low illumination based on low-rank decomposition," *Remote Sens.*, vol. 13, no. 2, p. 204, Jan. 2021, doi: 10.3390/rs13020204.

[22] 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.

[23] R. Qayyum, V. Akre, T. Hafeez, H. A. Khattak, A. Nawaz, S. Ahmed, P. Mohindru, D. Khan, and K. U. Rahman, "Android based emotion detection using convolutions neural networks," in *Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE)*, Mar. 2021, pp. 360–365, doi: 10.1109/ICCIKE51210.2021.9410768.

[24] Y. Qi, S. Zhou, Z. Zhang, S. Luo, X. Lin, L. Wang, and B. Qiang, "Deep unsupervised learning based on color un-referenced loss functions for multi-exposure image fusion," *Inf. Fusion*, vol. 66, pp. 18–39, Feb. 2021, doi: 10.1016/j.inffus.2020.08.012.

[25] N. Rai, Y. Zhang, B. G. Ram, L. Schumacher, R. K. Yellavajjala, S. Bajwa, and X. Sun, "Applications of deep learning in precision weed management: A review," *Comput. Electron. Agricult.*, vol. 206, Mar. 2023, Art. no. 107698, doi: 10.1016/j.compag.2023.107698.

[26] J. Raja, P. Shanmugam, and R. Pitchai, "An automated early detection of glaucoma using support vector machine based visual geometry group 19 (VGG-19) convolutional neural network," *Wireless Pers. Commun.*, vol. 118, no. 1, pp. 523–534, May 2021, doi: 10.1007/s11277-020-08029-z.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[28] R. Shen, I. Cheng, and A. Basu, "QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2469–2478, Jun. 2013, doi: 10.1109/TIP.2012.2236346.

[29] R. M. Schmidt, "Recurrent neural networks (RNNs): A gentle introduction and overview," 2019, *arXiv:1912.05911*.

[30] N. Siddique, S. Paheding, A. A. R. Angulo, M. Z. Alom, and V. K. Devabhaktuni, "Fractal, recurrent, and dense U-Net architectures with EfficientNet encoder for medical image segmentation," *Proc. SPIE*, vol. 9, no. 6, Dec. 2022, Art. no. 064004, doi: 10.1117/1.JMI.9.6.064004.

[31] J. Tumblin, A. Agrawal, and R. Raskar, "Why I want a gradient camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 103–110, doi: 10.1109/CVPR.2005.374.

[32] J. Wang, W. Wang, G. Xu, and H. Liu, "End-to-end exposure fusion using convolutional neural network," *IEICE Trans. Inf. Syst.*, vol. E101.D, no. 2, pp. 560–563, 2018, doi: 10.1587/transinf.2017EDL8173.

[33] J. Wang, X. Li, and H. Liu, "Exposure fusion using a relative generative adversarial network," *IEICE Trans. Inf. Syst.*, vol. E104.D, no. 7, 2021, Art. no. EDP7028, doi: 10.1587/transinf.2021EDP7028.

[34] X. Wang, Z. Sun, Q. Zhang, Y. Fang, L. Ma, S. Wang, and S. Kwong, "Multi-exposure decomposition-fusion model for high dynamic range image saliency detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4409–4420, Dec. 2020, doi: 10.1109/TCSVT.2020.2985427.

[35] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 117–132.

[36] F. Xu, J. Liu, Y. Song, H. Sun, and X. Wang, "Multi-exposure image fusion techniques: A comprehensive review," *Remote Sens.*, vol. 14, no. 3, p. 771, Feb. 2022, doi: 10.3390/rs14030771.

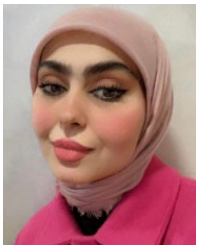
[37] H. Xu, J. Ma, and X.-P. Zhang, "MEF-GAN: Multi-exposure image fusion via generative adversarial networks," *IEEE Trans. Image Process.*, vol. 29, pp. 7203–7216, 2020, doi: 10.1109/TIP.2020.2999855.

[38] Q. Yan, D. Gong, P. Zhang, Q. Shi, J. Sun, I. Reid, and Y. Zhang, "Multi-scale dense networks for deep high dynamic range imaging," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 41–50, doi: 10.1109/WACV.2019.00012.

[39] Q. Yan, "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 1751–1760. [Online]. Available: https://donggong1.github.io/ahdr

[40] Z. Yang, Y. Chen, Z. Le, and Y. Ma, "GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 6133–6145, Jun. 2021, doi: 10.1007/s00521-020-05387-4.

[41] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020, doi: 10.1016/j.inffus.2019.07.011.
[42] Y. Zhang, T. Liu, M. Singh, E. Çetintaş, Y. Luo, Y. Rivenson, K. V. Larin, and A. Ozcan, "Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data," *Light, Sci. Appl.*, vol. 10, no. 1, p. 155, Jul. 2021, doi: 10.1038/s41377-021-00594-7.
[43] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021.



ZAINAB ALZAMILI was born in Iraq, in 1990. She received the Graduate degree from the Department of Computer Science, University of Baghdad, in 2014, and the master's degree from MUBS. She is currently pursuing the Ph.D. degree with the University of Sfax, Tunisia. She began working in many universities and held many positions, including the Rapporteur of the Department of Engineering and Information Technology, Sumer University. She worked for a long time in research and academic work. She completed her scientific career. She has published several research articles in international journals.

KASSEM M. DANACH (Member, IEEE) was born in Lebanon, in 1985. He received the Doctor of Philosophy (Ph.D.) degree in computer engineering from Ecole centrale de Lille, France, in 2016. He embarked on his academic journey with a fervent passion for computer engineering. Throughout his career, he has amassed a wealth of experience. He has applied his expertise in data analytics, artificial intelligence, machine learning, deep learning, educational technology, and business analytics to various military and professional roles. His illustrious career includes roles, such as the Chairperson of the Information Technology and Management Systems Department, Faculty of Business Administration, Al Maaref University, Lebanon. He has made substantial contributions to the field of operational research and artificial intelligence, earning him prestigious awards, including the Best Paper Award at LCIS, in 2017, and AISD, in 2019, and active participation as a Team Member in the EU-funded SOCORRO Project, from 2019 to 2021.

MONDHER FRIKHA received the master's degree in applied science and in electrical engineering from the University of Ottawa, Canada, in 1991, and the Ph.D. degree from the National School of Engineering of Sfax, Tunisia, in 2007. He is currently a Full Professor with the National School of Electronics and Telecommunications, University of Sfax, Tunisia. He is also the Director of the Advanced Technologies of Image and Signal Processing Research Laboratory. Then, he was a Project Head with the Industrial Land Agency, Tunisia. His research interests include digital signal and image processing, speech and audio processing, pattern recognition, and IA applications.