



OPEN Transferring enhanced material knowledge via image quality enhancement and feature distillation for pavement condition identification

Zejiu Wu^{1✉}, Yuxing Zou^{2,4}, Boyang Liu^{2,4}, Zhijie Li², Donghong Ji³ & Hongbin Zhang²

In the context of rapid advancements in autonomous driving technology, ensuring passengers' safety and comfort has become a priority. Obstacle or road detection systems, especially accurate pavement condition identification in unfavorable weather or time circumstances, play a crucial role in the safe operation and comfortable riding experience of autonomous vehicles. To this end, we propose a novel framework based on image quality enhancement and feature distillation (IQEFD) for detecting diverse pavement conditions during the day and night to achieve state classification. The IQEFD model first leverages ConvNeXt as its backbone to extract high-quality basic features. Then, a bidirectional fusion module embedded with a hybrid attention mechanism (HAM) is devised to effectively extract multi-scale refined features, thereby mitigating information loss during continuous upsampling and downsampling. Subsequently, the refined features are fused with the enhanced features extracted through the image enhancement network Zero-DCE to generate the fused attention features. Lastly, the enhanced features serve as the guidance online for the fused attention features through feature distillation, transferring enhanced material knowledge and achieving alignment between feature representations. Extensive experimental results on two publicly available datasets validate that IQEFD can accurately classify a variety of pavement conditions, including dry, wet, and snowy conditions, especially showing satisfactory and robust performance in noisy nighttime images. In detail, the IQEFD model achieves the accuracies of 98.04% and 98.68% on the YouTube-w-ALI and YouTube-w/o-ALI datasets, respectively, outperforming the state-of-the-art baselines. It is worth noting that IQEFD has a certain generalization ability on a classical material image dataset named MattrSet, with an average accuracy of 75.86%. This study provides a novel insight into pavement condition identification. The source code of IQEFD will be made available at <https://github.com/rainzyx/IQEFD>.

Keywords Pavement condition identification, Enhanced material knowledge, Image quality enhancement, Feature distillation, Hybrid attention mechanism

In recent years, with the rapid development of autonomous driving technology, how to ensure the safety of passengers and provide them with a relatively comfortable riding experience has become the focus of common attention inside and outside the industry. Autonomous vehicles not only need to complete driving tasks accurately in various complex traffic environments but also should respond to emergencies in a timely manner to ensure the life safety of occupants¹. Obstacle or road detection systems play a key role in this procedure, where the accurate identification of pavement conditions such as dry, slippery, icy, or snow, is directly related to the safety and driving experience of autonomous vehicles². According to the latest data³, many traffic accidents occur under non-dry pavement conditions, such as rain, snow, and fog conditions, causing vehicles to lose control while driving. Reduced road adhesion significantly increases the braking distance, thereby increasing the potential hazard⁴. Therefore, accurate pavement condition identification can not only help early braking

¹School of Science, East China Jiaotong University, Nanchang 330013, China. ²School of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China. ³Cyber Science and Engineering School, Wuhan University, Wuhan 430072, China. ⁴Yuxing Zou and Boyang Liu contributed equally to this work. ✉email: wzjhdjd@qq.com

but also provide the necessary dynamic adjustment for autonomous vehicles to ensure the safety and stability of our driving. By continuously monitoring pavement conditions, the vehicle can better adjust its running speed according to real-time data, thereby improving the safety and comfort of passengers⁵.

As we know, different pavement conditions usually have unique physical characteristics and effective detection methods are needed to monitor them in real-time. Traditional methods are often labor-intensive and inefficient, especially in some inclement weather conditions. Over the past few years, diverse methods have been proposed to investigate various conditions on different road surfaces. Many studies attempting to use visible or near-infrared images to classify pavement conditions have shown satisfactory classification accuracy^{6,7}. Jokela et al.⁸, proposed a pavement condition identification method based on the polarization variation of the light reflected from the road surface, and improved image contrast through texture analysis. However, the adaptability of this method is limited under different conditions. Jonsson et al.⁹ explored the feasibility of using three wavelengths of near-infrared bands to distinguish different pavement conditions. The images of pavement conditions captured by cameras on road sections have been used to build robust detection systems. Moreover, the effect of infrared illumination and infrared detectors on pavement detection under different viewing angles has also been studied. However, passive optical sensors, although capable of using artificial light, limited their usage in the dark condition¹⁰. Recently, with the continuous development of modern machine learning (ML) technologies, more and more researchers have begun to apply numerous classical ML classification models for pavement condition identification. For example, support vector machines (SVM) have achieved an accuracy of about 90%, but this result is still less than ideal for ensuring the safety of autonomous driving¹¹. Certainly, SVM performed better than the naive Bayes classifier in this task¹². To further enhance the robustness of the corresponding recognition model, Fauzi et al.¹³ proposed a novel method combining gray co-occurrence matrix and local binary pattern features, which focuses on low-level feature extraction for better characterizing diverse pavement conditions.

Although the above methods injected vitality into this field, current identification performance is relatively low to some degree. The application of well-known advanced technologies such as deep learning provides a new perspective for pavement condition identification. These technologies can identify the nuances of different pavement conditions by analyzing a large amount of data and improve the identification accuracy and response speed of detection systems¹⁴. High-quality data is important for the successful usage of these intelligent systems, while low-quality media may degrade the performance of these systems. Many studies^{15–18} have focused on image quality modeling to address this issue well, aiming to improve the robustness and accuracy of these systems.

As we know, as a classical representative, convolutional neural networks (CNNs) have greatly advanced the research of computer vision (CV)¹⁹. CNN adaptively detects features or patterns in images, which is well suited for pavement condition identification because CNN detects spatial patterns of surface differences²⁰. Hence, deep learning-based methods began to play an important role in pavement condition identification, promoting the development of the field of autonomous driving. Using pre-trained CNN backbones, such as VGG-16²¹, SqueezeNet²², ResNet50²³, DenseNet121²⁴, and ConvNeXt²⁵, to build classifier has been widely studied. These methods performed well on the classification tasks. For example, Cheng et al.²⁶ proposed a new CNN for pavement condition identification, which creates a new activation function based on the rectified linear unit (ReLU), significantly enhancing the classification metrics. Garcea et al.²⁷ proposed a semi-supervised LSTM-based model that leverages contrastive self-supervised pretraining and temporal consistency augmentation to effectively enhance wet pavement condition identification performance. Zhang et al.²⁸ proposed a multi-supervised bidirectional fusion network (MBFN) for detecting pavement conditions. The model uses the classical backbone ConvNeXt for feature extraction, which can improve classification efficiency. Although these methods demonstrate that it is possible to classify the pavement conditions ahead using camera images, challenges including adverse weather effects and inadequate lighting remain problematic²⁹. And they also lack a certain model interpretability, affecting models' practicality.

On the other side, the pavement conditions images at night usually have relatively poorer quality due to low brightness or contrast. They may be disturbed by the noise caused by various weather conditions. A common CNN is not good at processing these low-quality images and may lose some significant context information, which limits the generalization ability of the identification model and affects the final identification performance. To address the above problems, we propose a novel idea of transferring enhanced material knowledge by devising a novel framework named image quality enhancement and feature distillation (IQEFD), which is robust to diverse pavement conditions during the day and night. The IQEFD model first adopts the mainstream ConvNeXt as its backbone to extract high-quality basic features. The features extracted by ConvNeXt do not contain sufficient semantics information. Then, a novel bidirectional fusion module embedded with a hybrid attention mechanism (HAM) is designed to extract sufficient multi-scale refined features, thereby reducing the negative effect of continuous up-sampling and down-sampling operations. Our HAM focuses on modelling the importance of each channel through the channel attention module as well as the importance of each location through the spatial attention module to accurately identify the critical regions in the image, enabling the network to extract more robust features. Subsequently, we perform image quality enhancement on the original image using the image enhancement network called Zero-DCE³⁰. Zero-DCE predicts a higher-order curve by training a lightweight network named DCE-NET that adjusts the image by that curve to obtain the enhanced image. The refined features are fused with the enhanced features extracted from the enhanced image to generate the fused attention features. Additionally, the enhanced features serve as the teacher online for the fused attention features, facilitating the transfer of enhanced material knowledge and obtaining alignment between feature representations, which contributes to promoting robustness as well as identification performance. The main contributions are shown as follows:

- (1) We propose a novel idea of transferring enhanced material knowledge by designing an efficient but robust framework called IQEFD for effectively and robustly detecting weather-induced pavement conditions, es-

pecially for the path of automatic vehicles at night. To the best of our knowledge, this is an early work that absorbs enhanced material knowledge for pavement condition identification.

- (2) We bridge the gap between the image enhancement network named Zero-DCE and the fused attention features through online feature distillation. The enhanced features generated by Zero-DCE are regarded as the “teacher” to optimize the feature learning procedure of our model, which contributes to learning more complex feature patterns for effective and robust pavement condition identification.
- (3) We devise a novel bidirectional fusion module embedded with a hybrid attention mechanism (HAM), which not only reduces the negative effect of continuous upsampling and downsampling but also focuses on the focal road regions, significantly improving the model’s accuracy.
- (4) Experimental results on two benchmark datasets show that IQEFD reduces the negative impact of noises and obtains satisfactory identification performance on those night images, which outperforms the most advanced baselines. Additionally, the IQEFD model shows satisfactory generalization ability on a classical coarse-grained material image dataset. We also validate the robustness of IQEFD using both black-box testing with noise effects and gradient-based white-box attack testing. Furthermore, we provide sufficient model interpretability of IQEFD by implementing extensive visualization analysis.
- (5) The rest of the article is organized as follows. The second section describes the relevant literature and the motivation of our study. The third section describes the proposed IQEFD model. Then, the fourth section presents the relevant benchmark datasets used in this study. The fifth section discusses the experimental results on three benchmark datasets. Some qualitative results of the model are shown in the sixth section. Finally, the seventh section presents our conclusions and future work.

Related works

Material image identification

Material image identification is a fundamental problem in CV. The complexity of visual material appearance as observed in the huge variation under different viewing and illumination conditions makes material image identification a highly challenging task. Traditional works for material image identification usually use two distinct approaches. One approach assesses material identity using reflectance as an intrinsic property of the surface^{31,32}. Another approach identifies material labels using the appearance of the surface within the real-world scene^{33–35}. The success of deep learning methods in object identification also transferred into the material image identification field. For example, Bell et al. achieved per-pixel material category labeling by retraining the state-of-the-art object recognition network on a large-scale dataset of material appearance³⁶. Zhang et al.³⁷ introduced a deep texture encoding network (Deep-TEN) that absorbs the dictionary learning and feature pooling approaches into the classical CNN pipeline to learn an encoding for an orderless texture representation. Recently, Zhang et al.³⁸ proposed the gene selection XGBoost algorithm that combines feature selection and boosting strategy to generate image features with stronger discriminative and generalization abilities, which demonstrate satisfactory performance of material image identification. Asheghi et al.³⁹ proposed a novel two-level network called detail-aware salient object detection (DASOD), which addresses the challenges posed by complex backgrounds, low contrast, transparent objects, and occluded objects in images.

Pavement condition identification

We have seen significant advances in autonomous driving technology, and more and more people are relying on the safety and utility of this technology. However, in dynamic and varied road environments, the driving strategy of an intelligent vehicle will be mostly determined by the pavement conditions. Scholars have proposed lots of methods for this problem, such as the utilization of traditional ML models, such as SVM, decision trees, and random forests⁴⁰. Kim et al.⁴¹ used weather station data to predict rainy road conditions. Their method required considerable human resources. Deviations in human measurements often result in unexpected uncertainties. Zhao et al.¹¹ and Omer et al.⁴² used an SVM model to classify pavement conditions, obtaining an accuracy of approximately 90%, which was still unsatisfactory considering the safety required in self-driving. Smolyakov et al.⁴³ devised the model of the environment and the temperature of roads to predict icy-road conditions, which combines a physical model for predicting pavement conditions based on site measurements with an ML model to detect incorrect data.

Nevertheless, these traditional ML approaches only performed well in a few specific situations and poorly in a wider range of situations or special cases. With the rapid development of high-performance computing devices, deep learning methods offer new solutions to this problem. For example, Roychowdhury et al.⁴⁴ achieved an accuracy of 97% using SqueezeNet. Fink et al.⁴⁵ went one step further by using SqueezeNet to reduce the computational complexity without significantly affecting the accuracy. Guo et al.⁴⁶ proposed an improved YOLOv5 model with attention mechanisms to enhance accuracy and robustness in detecting pavement distress, demonstrating its effectiveness in smart transportation. Jiang et al.⁴⁷ proposed a novel pavement condition identification system by integrating the whale optimization algorithm-enhanced back-propagation neural network (WOA-BP) with multi-sensor data, which demonstrates superior performance in terms of detection accuracy and model stability compared to traditional ML methods. Karunasekera et al.⁴⁸ improved the pavement condition identification performance by fusing multiple information sources from temperature sensors and other image regions, and the pavement condition identification performance was significantly improved by fusing the additional information sources. Zhang et al.²⁸ proposed the MBFN model for effectively detecting pavement conditions, which obtains relatively better identification performance. Previous studies focused mainly on recognizing pavement conditions during the daytime while ignoring pavement conditions at night. As we know, low brightness or low contrast at nighttime usually leads to relatively poorer image quality and limits the generalization ability of existing methods to obtain sufficient discriminative information, thus posing a great challenge to the identification of nighttime pavement conditions. Moreover, wet pavement creates reflections

that blur the texture and color of the underlying surface. While dry pavement typically provides sharper details, but the colors vary depending on the material and weather conditions, making it more challenging to generalize across environments. Furthermore, different pavement materials have different visual characteristics. They also reflect light differently, which adds additional difficulty to recognizing pavement conditions under varying nighttime conditions.

Our motivation

Hence, pavement condition identification is still challenging to some degree, especially in unfavorable weather or time circumstances. In this study, we attempt to devise an effective but robust model to better address the above problems. Unlike the above methods, we combine image quality enhancement and online feature distillation ideas seamlessly to build our model. We hope the model is effective and robust for unfavorable weather or time conditions. Moreover, we will offer sufficient model interpretability.

Related technologies

Low-light image enhancement

Low-light image enhancement (LLIE) has been studied extensively in previous literature. LLIE aims at improving the perception or interpretability of an image captured in an environment with poor illumination. Traditional LLIE methods contain two classical categories, namely Histogram Equalization-based methods^{49–52} and Retinex model-based methods^{53–58}. Recent advances in this area are dominated by deep learning-based solutions. They usually employ specific learning strategies, network structures, loss functions, etc.

The original deep learning-based LLIE method called low-light net (LLNet)⁵⁹ uses a variant of stacked sparse denoising autoencoder to simultaneously brighten and denoise low-light images. Lv et al.⁶⁰ proposed an end-to-end multi-branch low-light enhancement network (MBLLEN), the key idea of MBLLEN is to extract rich features up to different levels, so that we can apply enhancement via multiple subnets. Unlike previous methods that rely on pairwise training data, EnlighthenGAN⁶¹ was the first unsupervised learning work that successfully introduced unpaired training data into LLIE. Zhu et al.⁶² proposed a three-branch CNN, called robust retinex decomposition network (RRDNet), for underexposed image restoration. The RRDNet decomposes an input image into illumination, reflectance, and noise via iteratively minimizing specially designed loss functions. Liu et al.⁶³ proposed a Retinex-inspired unrolling method for LLIE, in which the cooperative architecture search was used to discover lightweight prior architectures of basic blocks. Zhai et al.⁶⁴ provided a comprehensive survey of perceptual image quality assessment, including both traditional and emerging approaches, providing a valuable reference for image quality enhancement directions.

Our motivation

Evidently, the above LLIE methods can enhance the quality of those images with complex illumination. However, few works have considered this positive factor in the field of pavement condition identification, which contributes to boosting the effectiveness and robustness of the identification model. As an effective but efficient LLIE network, Zero-DCE³⁰ can help us achieve this goal.

Feature distillation

Knowledge distillation (KD) transfers the implicit but valuable knowledge from a teacher network to a student network with the goal of greatly improving the performance of the student network. For instance, Hinton et al.⁶⁵ utilized pre-trained teacher-generated logits as an additional goal for students. Motivated by KD, various logit-based approaches have been proposed for performance improvement. For example, Zhang et al.⁶⁶ proposed a deep mutual learning (DML) model, which replaces a pre-trained teacher with a set of students so that the distillation mechanism needs to train a large network of students in advance. Mirzadeh et al.⁶⁷ proposed a teacher-assistant knowledge distillation (TAKD) model, in which better teachers distill poorer students due to the large performance gap between them. Therefore, a similarity-based distillation method was proposed that is different from the traditional logits-based method, which attempts to explore the potential relationship between samples in the feature space. Tung et al.⁶⁸ proposed a similarity-preserving knowledge distillation (SPKD) method to allow pairs of input samples with similar activations in the teacher network to produce the same activations in the student network, thus guiding the learning procedure of the student network. Unlike logit-based method, feature distillation aims to make the mid-layer features of student and teacher as similar as possible. To make it easier to transfer knowledge from the teacher network, Kim et al.⁶⁹ introduced so-called “factors” as a more understandable form of intermediate representations. To match the semantics gap between teacher and student, Chen et al.⁷⁰ proposed a kind of cross-layer KD method, which adaptively assigns the appropriate teacher layer to each student layer through attention allocation. Furthermore, contrastive representation distillation (CRD)⁷¹ and Softmax regression representation learning (SRRL)⁷² show that the last-layer feature representations are more suitable for KD. One potential reason is that the last-layer feature representation is relatively closer to the final classifier. This study also draws on this idea to use the final layer for feature distillation.

Our motivation

More importantly, few works have ever employed the KD method to transfer material knowledge from enhanced features generated using the LLIE network to the student for promoting the effectiveness of feature learning procedure in the task of pavement condition identification. The gap between them needs to be broken while a significant relationship between them should be built up.

Attention mechanisms

The method of shifting specific attention to the most important regions of an image while ignoring those irrelevant parts is called the attention mechanism. The attention mechanism can capture the most related semantics from images. For instance, different channels in different feature maps represent different objects. Hu et al.⁷³ proposed the channel attention and squeeze-and-excitation net (SENet) for this purpose. SENet includes a SE block, which collects global information, capturing channel-wise relationships and improving representation ability. GSoP-Net⁷⁴ attempted to improve the squeeze module by changing the way they were compressed, and ECANet⁷⁵ reduced the complexity of SENet by improving the excitation module. Unlike them, SRM⁷⁶ improved both the squeeze and excitation modules. Mnihet al.⁷⁷ proposed the recurrent attention model (RAM) using RNNs and reinforcement learning to enable the network to learn where to pay attention. Jaderberg et al.⁷⁸ proposed spatial transformer network (STN) that explicitly learns the invariance to translation, scaling, rotation, and other general warps, making the STN pay more attention to the most relevant regions.

Inspired by ResNet, Wang et al.⁷⁹ proposed the residual attention network (RAN) by combining the attention mechanism with residual connections. Park et al.⁸⁰ proposed the bottleneck attention module (BAM), aiming to efficiently improve the representational capability of networks. As an important innovation, BAM used dilated convolution to enlarge the receptive field of the spatial attention sub-module and built a bottleneck structure to save computational costs. Moreover, channel attention and spatial attention were computed independently, ignoring the relationships between the two domains. Motivated by spatial attention, Misra et al.⁸¹ proposed novel triplet attention, a lightweight but effective attention mechanism that can capture cross-domain interaction.

Our motivation

Attention mechanism is also important for this study. On the one hand, it helps refine the extracted features to improve their discriminative abilities. On the other hand, it contributes to focusing on the focal pavement regions, significantly improving the model's accuracy and interpretability.

Whole design motivation

The overall design idea of this study is to improve the road condition detection accuracy and robustness under more complex nighttime environments by introducing a variety of innovative techniques while ensuring the model's generalization ability and interpretability to some degree. Most of all, we try to bridge the gap between Zero-DCE and the fused features through online feature distillation. By using the enhanced features generated by Zero-DCE as “teachers”, we can optimize the feature learning procedure and capture more complex feature patterns, which could significantly improve the accuracy and robustness of the pavement condition identification task. Moreover, we intend to build an attention module to effectively combines Channel Attention and Spatial Attention, allowing our model to better focus on the key road regions in each image.

The IQEFD model

The proposed IQEFD model is shown in Fig. 1. We describe the model using the following step-by-step mode.

Step 1: Data preprocessing. Before training the proposed model, the resolution of each input image is normalized and all images are resized to 224×224 with each color space considered. In addition, histogram-based image equalization is applied to each image for preliminary contrast enhancement.

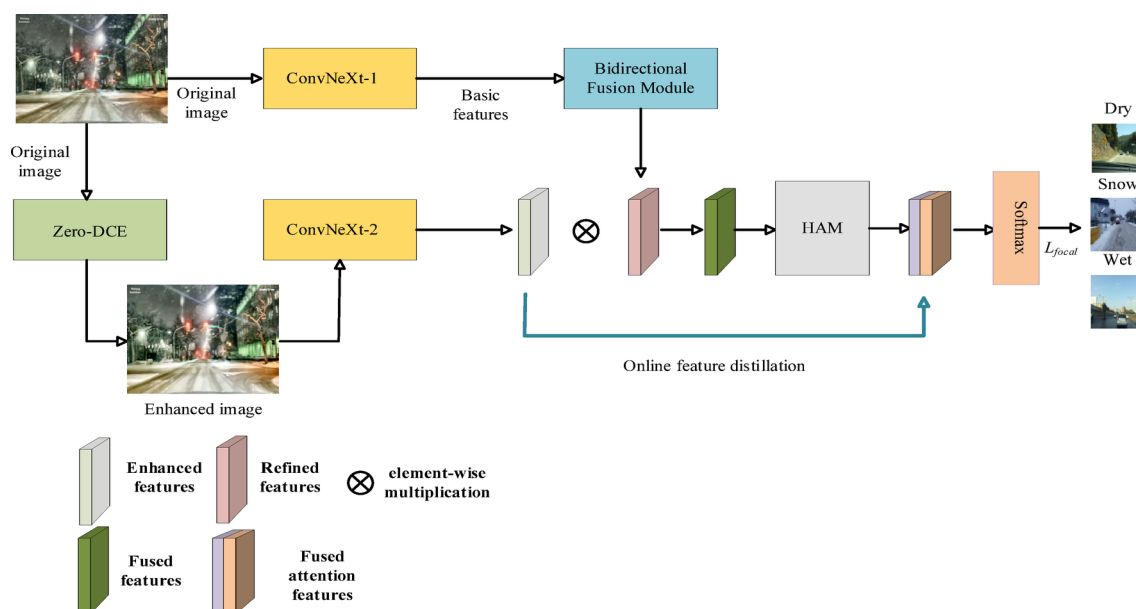


Fig. 1. The network architecture of IQEFD.

Step 2: Image quality enhancement. The effective but efficient LLEI network Zero-DCE is utilized to enhance the original pavement condition images, which builds a firm foundation for the subsequent material knowledge transfer.

Step 3: Feature extraction. Each original pavement condition image is input into ConvNeXt-1 for extracting basic features. Similarly, each enhanced pavement condition image is input into ConvNeXt-2 for extracting enhanced features. ConvNeXt-1 and ConvNeXt-2 employ the same structure of ConvNeXt but are used for different goals.

Step 4: Feature refinement. The refined features with sufficient multi-scale semantics information are obtained by throwing the basic features into the bidirectional fusion module.

Step 5: Feature fusion. We get the fused features by fusing the refined features with the enhanced features through element-wise multiplication, thereby capturing more valuable semantics information from the pavement condition images. Subsequently, the fused features are dynamically weighed using HAM to generate the fused attention features.

Step 6: Material knowledge transfer. The enhanced features guide the fused attention features online to enrich many more visual details through feature distillation. The semantics gap between feature learning and image quality enhancement is broken.

Step 7: Pavement condition identification. We input the fused attention features containing implicit material knowledge into a fully connected layer. Then a Softmax function using focal loss is employed to complete pavement condition identification.

Zero-DCE network

As analyzed above, few work leverage image enhancement strategies to boost the final performance of pavement condition identification. Most previous models are usually poor at handling dark conditions. To address this issue, on the one hand, we employ Zero-DCE to enhance the quality of the original pavement condition images, which provides many more reliable image samples for the field of nighttime pavement condition identification. On the other hand, we regard the features extracted from the enhanced image (we also call them enhanced features) as teachers, which transfers the implicit pavement knowledge online to the student (the fused attention features). All these contribute to improving the final recognition performance.

Zero-DCE uses a simple curve called a brightening curve to map the pixels of different brightness onto a new brightness, producing an image with just the right brightness. Zero-DCE uses neural networks to learn the mapping between low-light images and their optimal curve parameter mappings, and then generates a brightening image according to the curve and the original image. The brightening curve is formulated as below:

$$LE(I(\mathbf{x}); \alpha) = I(\mathbf{x}) + \alpha I(\mathbf{x}) (1 - I(\mathbf{x})) \quad (1)$$

Where \mathbf{x} is the pixel coordinate, $\alpha \in [-1, 1]$ is the learnable parameter, $LE(I(\mathbf{x}); \alpha)$ is an enhanced image of the original $I(\mathbf{x})$. In order to handle more complex low light situations, we iterated a brightening curve to learn more appropriate parameters, as shown in formula 2:

$$LE_n(\mathbf{x}) = LE_{n-1}(\mathbf{x}) + A_n(\mathbf{x}) LE_{n-1}(1 - LE_{n-1}(\mathbf{x})) \quad (2)$$

Here, n represents the number of iterations, A is a parameter map with the same dimensions as the image, where each pixel corresponds to an optimal adjustment parameter $A(\mathbf{x})$. This setup preserves the monotonic relationship of neighboring pixels during the enhancement process, ensuring continuous brightness distribution in local regions.

The enhancement process for low-brightness pavement condition images is illustrated in Fig. 2. The computation process of the brightening curve is differentiable, making it easy to optimize via gradient descent in neural networks. Here, DEC-Net is used, which has a total of 7 layers (6 hidden layers and 1 output layer). All layers are standard 3×3 convolutional layers with a stride of 1. To preserve the relationship between neighboring pixels, batch normalization is not applied after the convolutional layers. The activation function for each hidden layer is ReLU, and since the output lies in the range $[-1, 1]$, the activation function for the output layer is tanh. Notably, the 6 hidden layers use symmetric skip connections similar to U-Net⁸². After the output layer, each pixel has 24 channels, including 3 color channels (red, green, and blue), with each channel containing 8 parameters. The trained brightening curve automatically adjusts brightness and contrast. Consequently, Zero-DCE is relatively lightweight for image quality enhancement, laying a solid foundation for our feature learning. Enhanced brightness in nighttime pavement condition images improves the visibility of critical features, aiding more accurate pavement condition identification.

Bidirectional fusion module

As a well-known framework, a feature pyramid network (FPN) provides high-resolution details by integrating high-level global semantics into low-level feature maps⁸³, while retaining sufficient semantics differentiation. This multi-scale feature extraction capability enables the FPN to be more robust when processing objects of different sizes. However, simple stepwise fusion does not consider the varying importance of features at different scales. This may result in some unimportant high-level information being incorporated into low-level features, while crucial fine-grained information is not effectively retained. Additionally, although continuous downsampling captures more global semantics information, it severely loses spatial details, leading to insufficient model performance in detecting small objects or detailed features.

To address this issue, we propose a bidirectional fusion module with an embedded HAM. Specifically, we upsample the final layer of ConvNeXt-1 to match the scale of the features from the downsampling stage, allowing

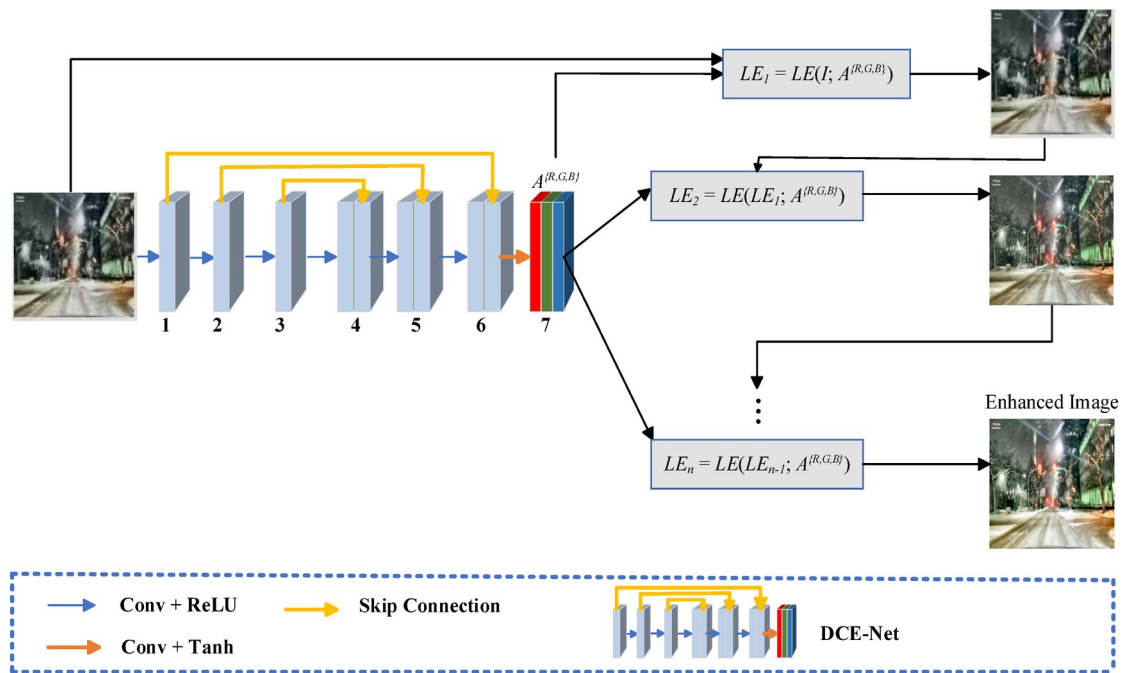


Fig. 2. The network structure of zero-DCE.

finer spatial details to be better preserved during fusion. This helps our model retain more discriminative information when integrating multi-scale features, improving its ability to detect small objects and fine-grained details. The features at different stages of the upsampling and downsampling operations are fused so that the fused features have sufficient semantics information. This operation allows the model to efficiently learn semantically complementary but robust features at different scales to mitigate the loss of semantics information during continuous down-sampling. We then concatenate these features from different stages to generate a new feature with sufficient semantics information. Additionally, during the concatenation of ConvNeXt-1 features with the bidirectional fusion module, as well as within the fusion module itself, we employ the proposed HAM that selectively focuses on the key information across different scales, enabling the model to concentrate on useful fine-grained and semantics information during fusion and enhancing the representation of the fused features.

Figure 3 shows the details of the bidirectional fusion module. The left side of this module performs an upsampling strategy. The multi-scale features extracted from the ConvNeXt-1 are incorporated, which are further refined by the HAM module. The right side of this module employs a top-down mode to perform downsampling. At each stage, the features of the same scale obtained in the intermediate upsampling are fused together to generate the final refined features. Notably, our model can learn much more multi-scale material knowledge for pavement condition identification.

Hybrid attention mechanism (HAM)

In pavement condition identification, the target object usually appears under different weather circumstances. To better adapt to the variations of the environment and enhance the robustness ability of target detection, a hybrid attention mechanism (HAM) is designed and embedded into the bidirectional fusion module (please see Fig. 3). The HAM module combines channel attention and spatial attention together to focus on those local areas in a pavement condition image. Compared with simple feature fusion or splicing operations, HAM can capture much more important semantics information. First, channel attention helps extract the semantics related to the target by selectively emphasizing channel information in different feature maps. In contrast, spatial attention focuses on identifying the significant regions in the spatial dimension. As expected, HAM is able to integrate multi-layer features extracted from ConvNeXt, thereby improving the final performance under very complex pavement conditions, especially for those conditions influenced by adverse weather. The structure of the proposed HAM module is shown in Fig. 4.

As shown in Fig. 4, the input feature F is passed through a channel attention module, where it goes through two parallel layers: Global Max Pooling and Global Avg Pooling. These produce the two feature maps F_{avg}^c and F_{max}^c , both with dimensions $1 \times 1 \times C$. These two feature maps are then fed into a two-layer multi-layer perceptron (MLP) with the shared weights, represented by W_0 and W_1 . The two output features generated by the MLP are added together and processed by a Sigmoid activation function, resulting in a channel-wise weight vector $M_c(F)$. The detailed computation is shown in Eq. 3. Next, $M_c(F)$ is multiplied with the original input feature F to obtain the channel attention feature F' .

Subsequently, a spatial attention module is applied to F' , as shown in Eq. 4. First, Max Pooling and Avg Pooling are performed along the channel dimension of F' , resulting in two feature maps, F_{avg}^s and F_{max}^s ,

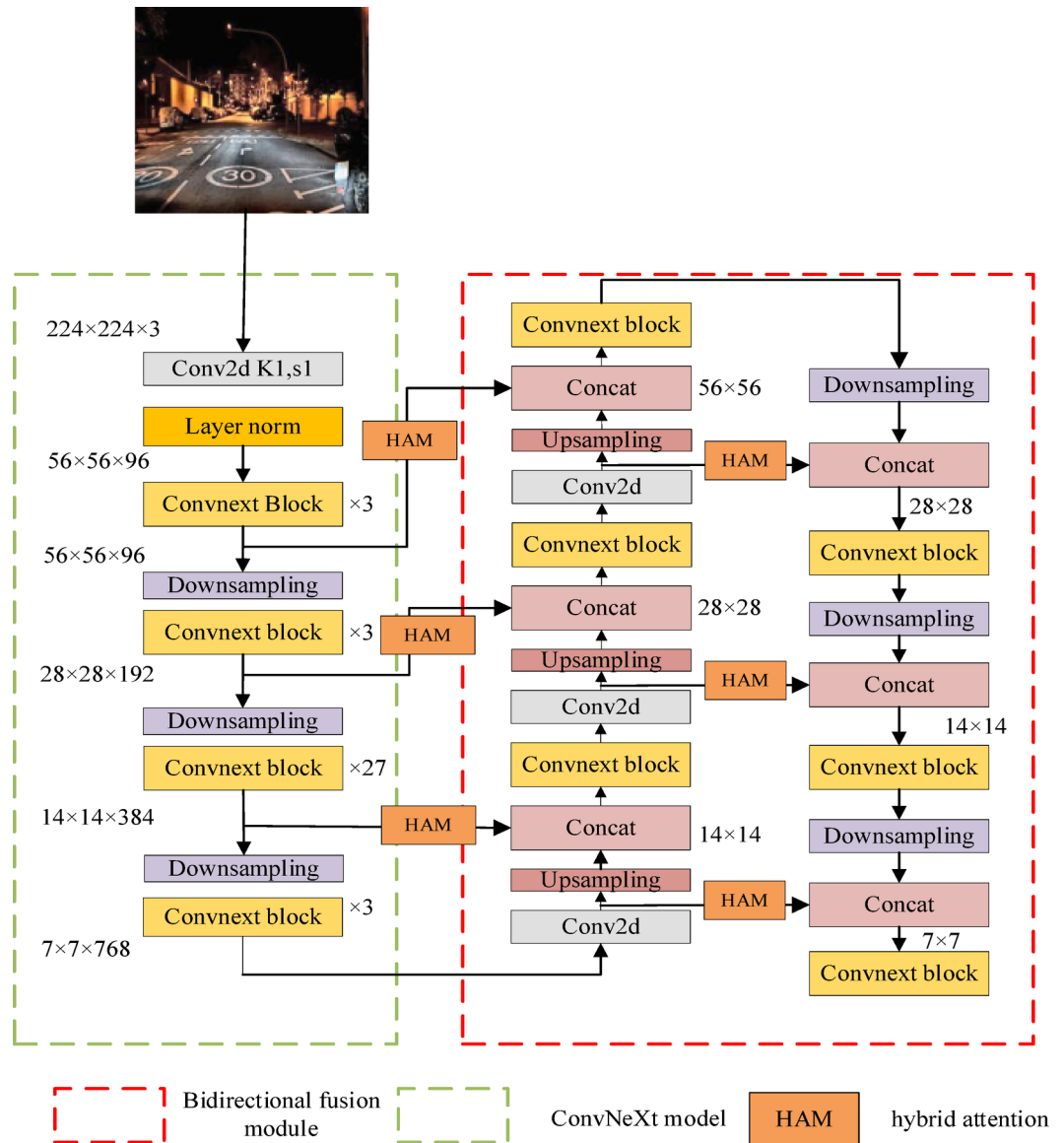


Fig. 3. The bidirectional fusion module with an embedded HAM.

both with dimensions $H \times W \times 1$. These two features are then concatenated, followed by a 7×7 convolution and a Sigmoid function to generate the spatial attention weight vector $M_s(F')$.

$$\begin{aligned}
 M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\
 &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))
 \end{aligned} \quad (3)$$

$$\begin{aligned}
 M_s(F') &= \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) \\
 &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s]))
 \end{aligned} \quad (4)$$

$$\begin{aligned}
 F' &= M_c(F) \otimes F \\
 F'' &= M_s(F') \otimes F'
 \end{aligned} \quad (5)$$

Where the symbol σ stands for the Sigmoid activation function. $f^{7 \times 7}$ represents a convolution operation with a convolution kernel size of 7×7 . The final output is shown as Eq. 5. F' is the result of multiplying the input feature F by the channel attention weight map $M_c(F)$. F'' is the result of multiplying F' by the spatial attention weight map $M_s(F')$.

In our model, on the one hand, we embed HAM into the bidirectional fusion module to refine feature at different scales, which can enrich much more discriminative information for pavement condition identification. The features extracted by ConvNeXt do not contain sufficient semantics information and need to go through a

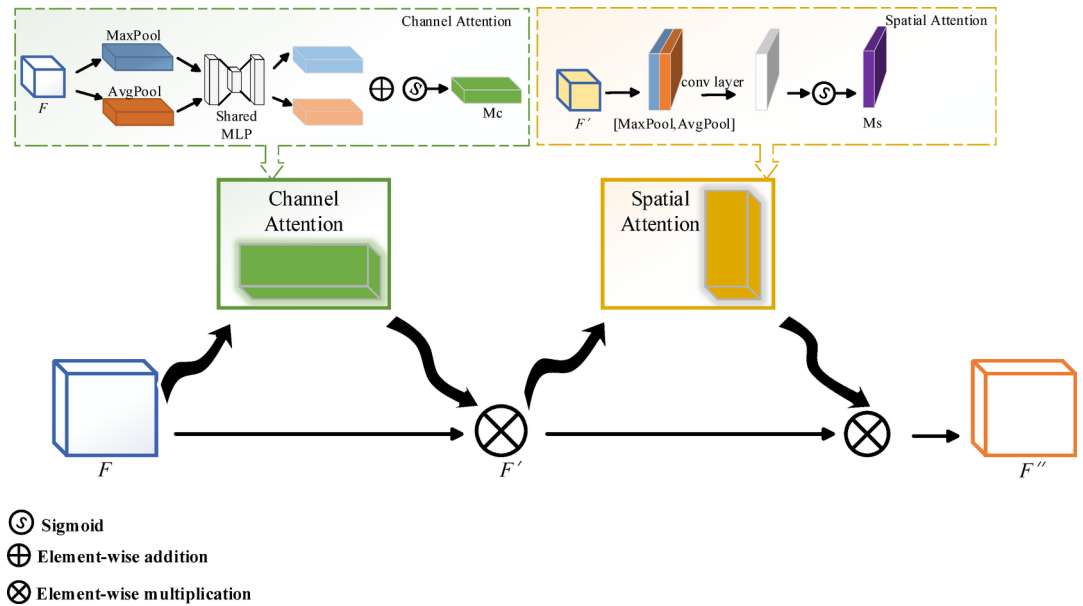


Fig. 4. The network structure of HAM.

bidirectional fusion network to extract more refined features for pavement condition identification. Hence we design the HAM module and embed it into the bidirectional fusion network so as to enhance the model's ability for focusing on the key regions in the pavement image, which helps eliminate the background interference and improve the final performance. Additionally, we use the HAM module on the fused features and get the fused attention features, further extracting crucial information for pavement condition identification.

Online features distillation

As analyzed above, few works have attempted to transfer material knowledge from an LLIE network to a feature learning network for pavement condition identification. The implicit knowledge embedded in the LLIE network is beneficial in improving the representation ability of the classification model. To this end, we explore the above-mentioned relationship through online feature distillation to further optimize the classification performance of our model. We define the pavement conditions dataset as $X = \{x_i\}_{i=1}^N$, where N is the total number of samples. The image category is $y_i \in \{1, 2, \dots, M\}$. In our model, the fused attention features serve as the student while the enhanced features are regarded as the teacher that can guide the learning procedure of the student online. We design a feature distillation loss called L_{fd} to capture and align the two feature representations:

$$L_{fd} = \|F_s - F_t\|_2^2 \quad (6)$$

Where F_s and F_t represent the features extracted from the student and teacher, respectively. The fused attention features obtained after weighting by the HAM module is denoted as F_s , whereas the feature of the front layer of the ConvNeXt-2 classification header is denoted as F_t . In addition, feature dimension alignment is achieved in advance using a 1×1 convolution kernel.

As the student, the fused attention features benefit from the implicit material knowledge of the teacher (the ConvNeXt-2 network using brightened images), effectively reducing the semantics distance between them. Through this online feature distillation, the student can extract more comprehensive pavement material knowledge, enabling more accurate pavement condition identification. This design is especially critical under nighttime and complex lighting conditions, allowing the proposed model to maintain robust classification performance across variable road scenes. By leveraging the enhanced features provided by the LLIE network, the student gradually adapts to pavement semantics features affected by complex lighting variations, providing powerful features for pavement condition identification and enhancing the overall classification accuracy.

Losses

In the IQEFD model, we employ focal loss to address the training instability problem caused by a slight sample imbalance. Focal loss incorporates a weighting factor into the loss function from the perspective of sample distribution, enhancing the model's focus on hard-to-classify samples by increasing their loss weight. The corresponding formula is shown in Eq. 7:

$$L_{fl} = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (7)$$

Where p_t represents the model's predicted probability or confidence for a given class, defined as $p_t = \text{sigmoid}(z_t)$, with z_t as the model's prediction score. The parameter α_t balances the number of positive and negative samples, decreasing with larger sample quantities and increasing otherwise. The parameter γ



Fig. 5. Some images in YouTube-w-ALI and YouTube-w/o-ALI.

| Dataset | Training set | | | Test set | | |
|-----------------|--------------|------|------|----------|------|------|
| | Dry | Wet | Snow | Dry | Wet | Snow |
| YouTube-w-ALI | 3219 | 3464 | 3510 | 3722 | 2837 | 4222 |
| YouTube-w/o-ALI | 3722 | 3830 | 3601 | 6081 | 2809 | 4183 |

Table 1. Data distribution of the YouTube-W-ALI and YouTube-w/o-ALI datasets.

adjusts for sample imbalance by reducing the loss for easy-to-classify samples. Typically, $\gamma \geq 1$ is chosen to enhance the model’s focus on difficult samples.

On this basis, we construct the loss function L_s for the bidirectional fusion network, which includes the focal loss L_{f11} and the feature distillation loss L_{fd} , as shown in Eq. 8. L_{f11} and L_{f12} represent the loss functions used for training the student and teacher networks, respectively. Their basic expressions are the same, as shown in Eq. 7, but the training samples used are different. The feature distillation loss L_{fd} guides the student online in learning the implicit pavement knowledge from the teacher through a weighted approach, with the weight values following Zhang’s research⁸⁴. Additionally, we define the teacher’s loss function L_t using the focal loss L_{f12} to serve as a benchmark for guiding the student’s learning. As shown in Eq. 9, only focal loss is required for the teacher’s loss.

$$L_s = L_{f11} + 0.000001 \times L_{fd} \tag{8}$$

$$L_t = L_{f12} \tag{9}$$

In summary, the IQEFD model combines focal loss and feature distillation loss, effectively addressing sample imbalance and hard-to-classify issues. Leveraging knowledge guidance from the teacher, we enhance the feature representation capacity of the proposed IQEFD model, providing robust and precise feature representation for pavement condition identification.

Dataset preparation

Compared with daytime pavement condition identification, it is more challenging to detect pavement conditions at night. There are relatively few works on nighttime data collection and pavement condition identification. To this end, Zhang et al.⁸⁵ collected data from YouTube videos to form new datasets. Each image was extracted from the video frame at an interval of at least 1 s between the two frames to ensure the difference between the images within the dataset, and the extracted images were all manually labeled. The corresponding video can be found at <https://doi.org/10.6084/m9.figshare.22775078>. As we know, the characteristics of the reflected light vary depending on the lighting conditions, and videos shot with ambient lighting differ from videos shot without it. The videos shot with ambient lighting are usually captured from urban areas while the videos shot without ambient lighting are usually shot from the countryside or highways. The images taken under ambient light illumination and other images taken without ambient light illumination have a large difference. Therefore, the datasets were collected separately under two lighting conditions, forming two different datasets respectively, namely YouTube-W-ALI and YouTube-w/o-ALI. The two datasets after necessary pre-processing can be found at <https://doi.org/10.6084/m9.figshare.22761149.v1>. Each dataset includes snow, wet, and dry conditions as shown in Fig. 5. Obviously, the wet pavement has more reflected light, whereas the dry pavement has less reflected light. Furthermore, it is more challenging to accurately detect pavement conditions on the YouTube-w/o-ALI dataset due to the low-light environment. Please refer to the right section of Fig. 5 to learn more details.

Zhang et al.⁸⁵ also collected validation datasets from other video sources. Each image has an original resolution of either $1,920 \times 1,080$ or $1,280 \times 720$, with the RGB color space. Prior to the training model, all images were resized to 224×224 , considering each color space. For the YouTube-w-ALI dataset, 10,193 images were finally chosen for training while the remaining 10,781 images were chosen for testing. For the YouTube-w/o-ALI dataset, 11,153 images were used for training with the remaining 13,073 images allocated for prediction. Table 1 provides more details of the two datasets.

Road surface is a kind of specific material. Hence, pavement condition identification belongs to a branch of the traditional task of material image classification. Therefore, to further verify the generalization ability of

| Category | Bag_pu | Bag_canvas | Bag_nylon |
|----------|---------------|------------|-------------|
| Number | 1982 | 1948 | 1764 |
| Category | Bag_polyester | Shoe_pu | Shoe_canvas |
| Number | 1715 | 1757 | 1855 |

Table 2. Data distribution of the MattrSet dataset.



Fig. 6. Some representative images in the MattrSet dataset.

the proposed IQEFD model, we conducted additional validation experiments on a classical material image dataset called MattrSet⁸⁴. This dataset was derived from online real goods, including bags and shoes, and it was constructed under the guidance of several experienced materials experts. Unlike other material image datasets, MattrSet is a coarse-grained material image dataset that includes four types of material, such as polyurethane (PU), canvas, nylon, and polyester. The MattrSet dataset contains 11,021 images with resolutions ranging from 123 × 123 to 4,300 × 2,867. All these will make the image material recognition task more challenging than before. On the contrary, this material image dataset is a good choice to demonstrate the generalization ability of our IQEFD model. To fulfill this goal, we completed four-fold cross-validation testing in our experiment. Table 2; Fig. 6 exhibit more details of this dataset. For example, “Bag_pu” represents the product is a bag with the material of pu while “Shoe_canvas” represents the product as a shoe with the material of canvas. The MattrSet dataset is completely open and can be obtained freely from the following link: https://drive.google.com/open?id=12xXX_MuwII8ghwXFLtT3sneEzgA4-SN.

All in all, we selected two pavement condition image datasets with different lighting conditions to validate the effectiveness and robustness of the IQEFD model. Furthermore, we demonstrated the generalization ability of IQEFD on the classical material image dataset MattrSet. All these contribute to evaluating the proposed model more comprehensively.

Experiments and analysis

Experimental settings

We used PyTorch on a computer server with four NVIDIA GeForce GTX2080Ti GPUs with 94 GB of RAM. We used the Adam optimizer with a weight decay and set the initial learning rate to 8e-5. The regularized weight decay was 5e-4, and the learning rate decay was 0.1. The batch size was set to 8, and the proposed model was trained for 30 epochs.

Baselines and evaluation metrics

To better demonstrate the effectiveness, robustness, and generalization ability of the proposed IQEFD model, we compared our model with the following mainstream baselines. (1) Fine-tuned deep learning-based networks including CNNs (CNN-1, CNN-2, and CNN-3 represent CNN using one convolutional layer, two convolutional layers, three convolutional layers, respectively), VGG-16²¹, VGG-19²¹, SqueezeNet²², ResNet50²³, Vision Transformer (ViT)⁸⁶, DenseNet121²⁴, ConvNeXt²⁵, ResNeXt⁸⁷, EfficientNet⁸⁸, FocalNet⁸⁹. (2) Two classical material image recognition models: SENet correlation gene selection (SECGS)⁸⁴ and hierarchical multi-feature fusion (HMF²)⁹⁰. (3) Previous state-of-the-art model MBFN²⁸. Given the nature of the datasets used in this field, we compared, to the extent possible and all our effort, all baseline models that were compatible with our datasets and consistent with the problem we were addressing. We used Accuracy (Acc), recall (Rec), and specificity (Spec) to evaluate our model.

| Model | Acc. | Rec. | Spec. |
|-----------------|--------------|--------------|--------------|
| CNN-3 | 90.08 | 87.36 | 87.03 |
| CNN-2 | 90.72 | 86.29 | 86.47 |
| CNN-1 | 90.89 | 89.34 | 89.65 |
| SqueezeNet | 89.14 | 85.65 | 85.36 |
| VGG16 | 90.65 | 88.46 | 88.21 |
| VGG19 | 90.17 | 89.17 | 89.01 |
| ResNet50 | 92.54 | 90.34 | 90.35 |
| ResNeXt50 | 92.78 | 92.87 | 96.48 |
| EfficientNet-B3 | 92.74 | 92.88 | 96.42 |
| FocalNet | 93.22 | 93.19 | 96.65 |
| DenseNet121 | 94.08 | 93.27 | 93.36 |
| MBFN | 97.28 | 97.06 | 97.36 |
| IQEFD (Our) | 98.04 | 98.01 | 98.06 |

Table 3. Performance comparisons with baselines on the YouTube-w-ALI dataset, unit: %. The best value of each metric is bold and italic.

| Model | Acc. | Rec. | Spec. |
|-----------------|--------------|--------------|--------------|
| CNN-3 | 90.96 | 88.21 | 88.65 |
| CNN-2 | 90.16 | 88.98 | 89.21 |
| CNN-1 | 89.96 | 85.29 | 85.78 |
| SqueezeNet | 93.59 | 92.36 | 92.48 |
| VGG16 | 91.65 | 90.58 | 90.87 |
| VGG19 | 91.79 | 91.44 | 91.88 |
| ResNet50 | 92.17 | 92.30 | 92.51 |
| ResNeXt50 | 93.36 | 94.06 | 97.08 |
| EfficientNet-B3 | 94.58 | 94.76 | 97.23 |
| FocalNet | 95.16 | 95.06 | 97.85 |
| DenseNet121 | 95.46 | 96.02 | 96.25 |
| MBFN | 97.71 | 97.82 | 97.34 |
| IQEFD (Our) | 98.68 | 98.90 | 98.65 |

Table 4. Performance comparisons with baselines on the YouTube-w/o-ALI dataset, unit: %. The best value of each metric is bold and italic.

Quantitative comparisons

We compared the IQEFD model with a group of baselines. The corresponding experimental results are presented in Table 3 and 4. It can be observed that the proposed IQEFD model performs well on both datasets. And it outperforms all the baselines with evident performance margins. To get more valuable conclusions, we conducted an in-depth analysis from the following perspectives.

First, we compared our model with those fine-tuned deep learning networks. On the YouTube-w-ALI dataset, compared with the most competitive DenseNet121, the accuracy, recall, and specificity of IQEFD improved by 3.96%, 4.74%, and 4.7%, respectively. On the YouTube-w/o-ALI dataset, the corresponding performance improvements are 3.22%, 2.88%, and 2.4%, respectively. As mentioned above, deep learning-based networks are very complex (e.g., ResNet50, and VGG19 use very deep feature layers), and the layers of these networks rarely interact with each other. Moreover, the extracted features usually focus on relatively small independent regions of the image, ignoring the overall context information (you can see Fig. 10). All these could cause the loss of the key semantics information. However, the IQEFD model handles this issue well. On the one hand, the bidirectional fusion module implements feature fusion at different scales to retain the key semantics information. On the other hand, the Zero-DCE network provides high-quality material knowledge to guide the procedure of the final feature learning, which continuously enhances the discriminative ability of our features, thus contributing to the pavement condition identification.

Second, compared with the previous MBFN model, the accuracy, recall, and specificity of our model improved by 0.76%, 0.95%, and 0.7% on the YouTube-w-ALI dataset and improved by 0.97%, 1.08%, and 1.31% on the YouTube-w/o-ALI dataset, respectively. To our surprise, more evident performance improvements are observed on the YouTube-w/o-ALI dataset. We guess Zero-DCE offers positive “bias” to the YouTube-w/o-ALI dataset, which means that image quality enhancement plays a more positive role in this dataset. This also firmly validates the robustness of our method. As analyzed above, the enhanced features generated through Zero-DCE

| Dataset | 1st fold | 2nd fold | 3rd fold | 4th fold | Avg | Std |
|-----------------|----------|----------|----------|----------|-------|-------|
| YouTube-w-ALI | 98.21 | 97.65 | 97.86 | 98.12 | 97.96 | 0.220 |
| YouTube-w/o-ALI | 98.42 | 98.81 | 97.96 | 98.25 | 98.36 | 0.308 |

Table 5. Four-fold cross-validation results on YouTube-w-ALI and YouTube-w/o-ALI, unit: %.

| Model | 1st fold | 2nd fold | 3rd fold | 4th fold | Avg | Imp | Std |
|------------------|--------------|--------------|--------------|--------------|--------------|------|--------------|
| ConvNeXt | 73.70 | 73.70 | 73.50 | 74.50 | 73.85 | 2.01 | 0.384 |
| SECGS | 70.10 | 71.38 | 72.10 | 72.30 | 71.42 | 4.44 | 0.862 |
| HMF ² | 71.30 | 71.90 | 72.50 | 73.10 | 72.20 | 3.66 | 0.671 |
| MBFN | 74.80 | 74.70 | 74.60 | 76.20 | 75.08 | 0.78 | 0.653 |
| IQEFD | 75.18 | 75.39 | 75.72 | 77.16 | 75.86 | / | 0.773 |

Table 6. Four-fold cross-validation on the MattrSet dataset, unit: %. The best value of each metric is bold and italic.

serve as a teacher online, providing good guidance for the fused attention features, which helps further improve the discrimination of our model.

In a word, the IQEFD model is superior to current mainstream baseline models. Notably, larger performance improvements are observed on the YouTube-w/o-ALI dataset. The IQEFD model is effective and robust for the task of pavement condition identification in unfavorable weather or time conditions. More importantly, our idea of transferring enhanced material knowledge is effective, which provides a novel insight into the task of pavement condition identification.

Cross-validation results

Cross-validation is a useful strategy for objectively evaluating the performance and robustness of a classification model. Researchers usually use the *k*-fold approach to complete their cross-validation experiments. In this study, we implemented four-fold cross-validation, which is a significant complementarity to the above performance comparisons. We randomly divided the dataset into 4 (*k*=4) parts. One part was used as the validation set while the remaining *k* – 1 parts were remained as the training set. We repeated our testing four times and computed the average accuracy (Avg) as well as standard deviation (Std) of each dataset. The four-fold cross-validation results of the IQEFD model are shown in Table 5.

According to Table 5, we find that the average accuracy of our model on each dataset is very close to the accuracy value presented in Table 3 (or Table 4), which means that our model is robust and stable for pavement condition identification to a certain degree. Moreover, the standard deviation results of the IQEFD model on each dataset are also stable, with values of 0.220% and 0.308%, respectively. This further indicates that our model is very stable, promoting its practicality.

To further demonstrate the generalization ability of the IQEFD model, we completed additional cross-validation experiments on the classical material image dataset called MattrSet. The dataset contains relatively more coarse-grained material semantics information and fewer training samples, which brings a certain challenge to the proposed recognition model. The corresponding experimental results are shown in Table 6. We computed the average accuracy (Avg) and standard deviation (Std) of each baseline model. “Imp” is a relative performance improvement of the IQEFD model compared to the corresponding baseline. For example, the performance improvement of IQEFD compared to ConvNeXt is 75.86%-73.85%=2.01%.

According to the Std values in Table 6, our model is also stable for coarse-grained material image recognition. This finding firmly supports the above conclusion of Table 5. In addition, the IQEFD model significantly outperforms a group of baselines, including the ConvNeXt, SECGS, HMF², and MBFN models, with an improvement of 2.01%, 4.44%, 3.66%, and 0.78% in terms of the average accuracy, respectively, which better demonstrates the generalization ability of the IQEFD model. Hence, IQEFD is also a general method for material-related image recognition tasks, highlighting its wide application prospect.

Robustness validation results

To assess the robustness of the proposed IQEFD model, we conducted both black-box testing with noise effects and gradient-based white-box attack testing in this section. We want to simulate real-world noise effect or face adversarial attack and evaluate the robustness of our model more comprehensively. The black-box test assumes that the adversary has no access to the model’s internal information, while the white-box test allows access to the model’s parameters, gradients, and other details. Given the numerous random variables involved, we reported the average results from three experiments as the final outcome. In the black-box testing, we applied random noise to 20% of the input data and observed the corresponding changes in the model’s output. This approach simulates real-world noise effects, such as sensor noise, environmental interference, data corruption, and other factors that could introduce noise into the input image.

| Dataset | Noise type | 1st testing | 2nd testing | 3rd testing | Avg | Var |
|-----------------|-------------------|-------------|-------------|-------------|-------|------|
| YouTube-w-ALI | Gaussian noise | 97.54 | 96.62 | 96.94 | 97.03 | 1.01 |
| | Salt Pepper noise | 96.13 | 97.59 | 97.54 | 97.08 | 0.96 |
| YouTube-w/o-ALI | Gaussian noise | 95.91 | 97.33 | 95.83 | 96.36 | 2.32 |
| | Salt Pepper noise | 95.44 | 95.64 | 94.76 | 95.28 | 3.40 |

Table 7. Identification results of IQEFD after facing different noise interference, unit: %.

| Dataset | 1st testing | 2nd testing | 3rd testing | Avg | Var |
|-----------------|-------------|-------------|-------------|-------|------|
| YouTube-w-ALI | 96.04 | 96.10 | 96.07 | 96.07 | 1.97 |
| YouTube-w/o-ALI | 96.37 | 96.40 | 96.38 | 96.38 | 2.30 |

Table 8. Identification results of IQEFD after facing FGSM confrontation attack, unit: %.

For the random noise selection, we used Gaussian noise and Salt Pepper noise. Gaussian noise simulates the random, continuous noise typically encountered during sensor acquisition. It reflects subtle disturbances from the natural environment and is useful for testing the model’s stability under continuous noise. In contrast, Salt Pepper noise is more suited to simulating sudden, discrete noise caused by transmission errors, dust, localized occlusions, or sensor failures. This type of noise results in distinct pixel extremes and presents different challenges for model robustness. In the Gaussian noise experiment, the standard deviation of the noise added to each sample was randomized within the range of 0 to 5, simulating real-world uncertainty as accurately as possible. For the Salt Pepper noise, we set the ratio of Salt noise to Pepper noise at 1:1, with the maximum intensity of the noise capped at 0.1. The experimental results are shown in Table 7 and the evaluation metric is accuracy. We repeated each testing three times and got the averaged accuracy (Avg). And we computed the variance (Var) that is the comparison with before the addition of noise.

The experimental results in Table 7 demonstrate that IQEFD exhibits strong noise robustness (NR), as it can still classify correctly despite noise disturbances. This indicates that the model learns more robust image features rather than merely memorizing the training data. Additionally, IQEFD shows good resistance to both continuous perturbations and discrete outliers. All these demonstrate the effectiveness of our design idea. Certainly, there is a little challenge for the YouTube-w/o-ALI dataset. Thus, from the noise effect perspective, our model is very robust for pavement condition identification.

Unlike the above attack testing, the Fast Gradient Sign Method (FGSM)⁹¹ is a gradient-based white-box adversarial attack technique usually designed to generate adversarial examples for testing the robustness of deep learning models. Hence, it is a significant complementarity to the above robustness testing. In detail, FGSM creates adversarial samples by calculating the gradient of the loss function with respect to the input data and then adding a small perturbation in the direction of the gradient, producing samples that can deceive the model, as shown in Eq. 10.

$$x' = x + \epsilon \cdot \text{sign}(\nabla_x J(\theta, x, y)) \tag{10}$$

Let x denotes the original input image, x' denotes the adversarial samples, ϵ denotes the attack strength (i.e., the size of the perturbation), and $J(\theta, x, y)$ denotes the loss function. In this case, we used the cross-entropy loss function, following previous designs. The term $\nabla_x J(\theta, x, y)$ represents the gradient of the loss function with respect to the input x . In our experiment, we set ϵ to 0.05 and selected 20% of the samples for the attack testing. This setting helps reduce computational cost while still effectively evaluating the adversarial robustness of the proposed model. The experimental results are shown in Table 8 and the evaluation metric is accuracy. We repeated each testing three times and got the averaged accuracy (Avg). And we computed the variance (Var) that is the comparison with before facing the FGSM confrontation attack.

The experimental results in Table 8 show that IQEFD still maintains high accuracy on the adversarial samples generated by FGSM, indicating that the proposed model is insensitive to small-amplitude perturbations and has already possessed some adversarial robustness. All these demonstrate the effectiveness of our design idea. Certainly, there is a little challenge for the YouTube-w/o-ALI dataset. In the future, we will test it under stronger adversarial attacks and further enhance model’s robustness on the YouTube-w/o-ALI dataset. Thus, from the adversarial attack perspective, our model is also very robust for the task of pavement condition identification.

Summarily, the above extensive results validate that the IQEFD model is much robust for pavement condition identification. Our model design idea also works well under relatively harsh environment.

Real-time performance of the model

In practical applications, on the one hand, model accuracy is very important. On the other hand, real-time performance must also be considered in the task of pavement condition identification. Therefore, we measured the model’s frame rate (i.e., the number of the images processed per second) and latency (i.e., the total time from input data entry to result generation). For each metric, we evaluated two scenarios: (1) the end-to-end time, which includes dataset loading and preprocessing, (2) the inference time, which covers only the model’s forward

| Dataset | Overall FPS | Inference FPS |
|-----------------|-------------|---------------|
| YouTube-w-ALI | 59.69 | 68.70 |
| YouTube-w/o-ALI | 59.84 | 68.76 |

Table 9. Number of images processed per second by the IQEFD model.

| Dataset | Overall latency | Inference latency |
|-----------------|-----------------|-------------------|
| YouTube-w-ALI | 0.016754 | 0.014555 |
| YouTube-w/o-ALI | 0.016712 | 0.014543 |

Table 10. Averaged latency of the IQEFD model, unit: second.

inference process. The results are presented in Tables 9 and 10. To ensure reliable measurements, we calculated the average inference time over 1000 images.

As shown in Tables 9 and 10, IQEFD demonstrates satisfactory processing speeds, making it well-suited for real-time applications. Additionally, the short processing time per image allows the system to make decisions quickly, helping reduce potential safety risks. All these real-time performance metrics accompanied with satisfactory identification accuracy support the possible deployment of the proposed IQEFD model in the field of autonomous driving.

Confusion matrix

The confusion matrix can help us fully understand the recognition ability of the IQEFD model in each category. Here, we compare our model with the state-of-the-art MBFN model according to confusion matrices Fig. 7a and b show the confusion matrix of the MBFN model on the YouTube-w-ALI and YouTube-w/o-ALI datasets, respectively, whereas Fig. 7c and d show the confusion matrices of the IQEFD model on the YouTube-w-ALI and YouTube-w/o-ALI datasets, respectively.

As shown in Fig. 7, on the YouTube-w-ALI dataset, 8 images of the dry surface condition are misclassified with sufficient light, which are 6 wet samples and 2 snow samples. This classification effect is satisfactory to some degree. As we know, with sufficient light, the dry ground doesn't have more reflective conditions, which will not affect the recognition model negatively. However, for the category of wet, the corresponding wrong results are 53 dry samples and 9 snow samples, respectively, which may be due to the reflection and scattering of the wet images under the light condition, generating a certain noise and affecting the final performance. The recognition of the snow category is similar to the wet category. However the recognition task in the snow category is slightly more challenging than in the wet category. Moreover, our IQEFD model beats the MBFN model in each category evidently, further validating the effectiveness of the proposed method.

Further, on the YouTube-w/o-ALI dataset, the IQEFD model obtains more satisfactory classification performance compared to YouTube-w-ALI. For example, only one snow sample is misclassified as wet. This is due to the fact that the reflection of light is greatly reduced under low-light conditions. Meanwhile, our model enhances the original input image and employs mid-layer feature maps to boost the corresponding performance. However, there is still a certain challenge for the dry samples. High visual similarity occurs between the dry and wet pavement conditions. We need to absorb a more effective feature representation approach to address this problem in our future work. Furthermore, our model is superior to the previous MBFN model evidently, further promoting the practicality of IQEFD.

In summary, based on the confusion matrix, our model exhibits satisfactory classification performance in the task of pavement condition identification. It outperforms the state-of-the-art baseline in most categories. In the future, we need to pay more attention to the snow samples on YouTube-w-ALI as well as the dry samples on YouTube-w/o-ALI.

Ablation analysis

In this section, we make extensive ablation analysis experiments, including backbone model selection, and contribution evaluation of each component. We need to obtain a deeper understanding of the IQEFD model.

First, the backbone plays a significant role in our model. In order to ensure the effectiveness and model size, we chose ConvNeXt-S as the backbone network of the proposed approach. Thus, we can obtain a good trade-off between recognition performance and model size in this way. To support this choice, we made a detailed comparison between different ConvNeXt backbones in Table 11. As shown in Table 11, the ConvNeXt-B model is nearly 0.11% better than ConvNeXt-S, whereas the number of parameters in ConvNeXt-B is nearly 1.78 times that in ConvNeXt-S. Compared with ConvNeXt-B, ConvNeXt-S is a “cheaper” but effective choice. Similarly, compared with ConvNeXt-T, ConvNeXt-S is more effective and the corresponding number of parameters is relatively acceptable. Hence, we consider using ConvNeXt-S as the backbone of the proposed IQEFD model in terms of recognition performance and model size.

Second, the IQEFD model consists of the Zero-DCE-based image enhancement module, the bidirectional fusion module embedded with the HAM module, and the feature distillation module. The actual contribution of each module needs to be verified through ablation experiments. We removed the Zero-DCE-based image

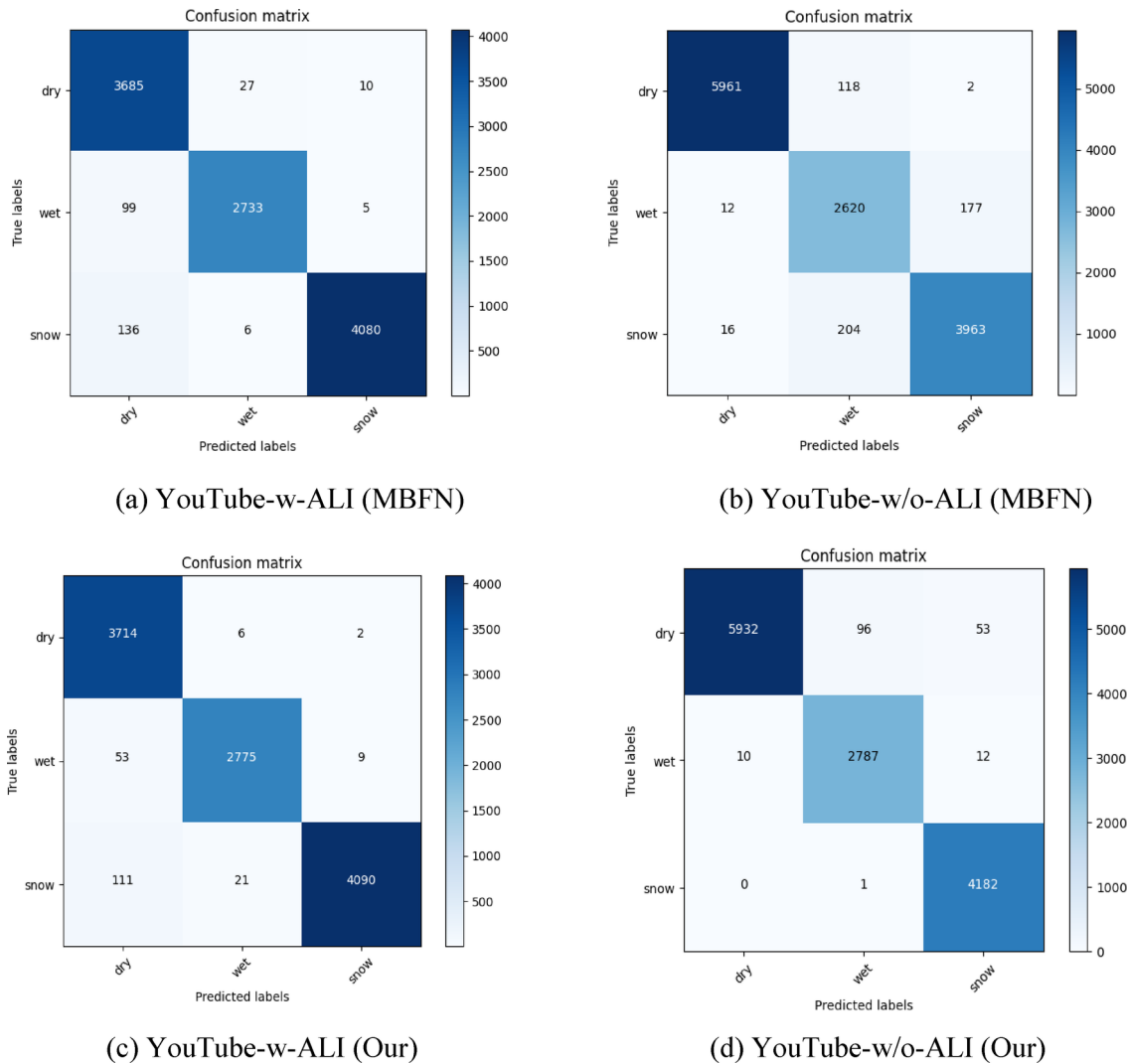


Fig. 7. Confusion matrix comparisons for pavement condition identification.

| Backbone | Accuracy on different datasets (%) | | Param(MB) |
|------------|------------------------------------|---------------|-----------|
| | YouTube-w/o-ALI | YouTube-w-ALI | |
| ConvNeXt-T | 96.23 | 96.08 | 29 |
| ConvNeXt-S | 97.71 | 97.28 | 50 |
| ConvNeXt-B | 97.82 | 97.35 | 89 |

Table 11. Efficiency and parameters of different ConvNeXt backbones. “Param (MB)” indicates the number of parameters in M (million).

enhancement module, all HAM modules, feature distillation module, as well as using both and neither of them, respectively. Thus, we get five variants of our model. The corresponding results are shown in Table 12. From Table 12, it can be found that removing any of the modules on both datasets could affect the final accuracy negatively. Specifically, the most substantial performance decay occurs after removing the Zero-DCE-based image enhancement module, which is 0.51% on the YouTube-w-ALI dataset and 0.65% on the YouTube-w/o-ALI dataset, respectively. This means that the Zero-DCE-based image enhancement module is the most significant component of our model. It improves the image quality to some extent without considering any light conditions. So, it curbs the noise in low quality images and reduces the negative impact on classification at the source, ultimately improving the pavement condition identification accuracy. Moreover, the feature distillation module is slightly better than the HAM module. This module employs the enhanced features to guide the learning procedure of the fused attention features online, which facilitates enhanced material knowledge transfer and bridges the semantics gap between the image enhancement and feature extraction modules, improving

| Variant | Zero-DCE | HAM | Feature distillation | YouTube-w-ALI | YouTube-w/o-ALI | Param (MB) |
|---------|----------|-----|----------------------|---------------|-----------------|---------------|
| 1 | × | × | × | 97.28 | 97.71 | 75.81 |
| 2 | × | √ | √ | 97.53 | 98.03 | 76.05 |
| 3 | √ | × | √ | 97.85 | 98.43 | 163.36 |
| 4 | √ | √ | × | 97.76 | 98.38 | 163.61 |
| 5 | √ | √ | √ | 98.04 | 98.68 | 163.61 |

Table 12. Ablation analysis results, unit: %. The best value of each metric is bold and italic. “Param (MB)” indicates the number of parameters in M (million).

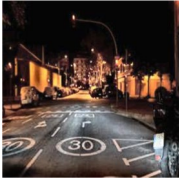

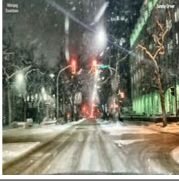
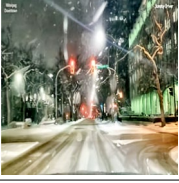
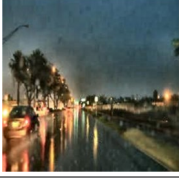
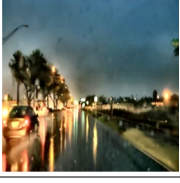
| Pavement conditions | Original image | Enhanced image |
|---------------------|--|--|
| Dry |  |  |
| Snow |  |  |
| Wet |  |  |

Table 13. Image quality enhancement results.

the discriminative ability of the fused attention features and thus boosting the final classification performance. Therefore, the contribution descending order of the three modules is “Zero-DCE > Feature Distillation > HAM”. The contribution of all the modules is the same on both datasets, which also indicates the stability of the proposed model.

In conclusion, each component of IQEFD plays an important role in pavement condition identification, and all the components form a synergy to jointly drive the accuracy improvement of the model.

Visualization analysis
Image quality enhancement results

In the IQEFD model, we utilize the Zero-DCE network to enhance each input image and extract the corresponding enhanced features. Subsequently, the enhanced features are used to guide the learning procedure of the fused attention features online through feature distillation. It bridges the gap between the image enhancement network and feature extraction network, thus contributing to improving the final recognition performance. So, Zero-DCE plays a significant role in our model. The image quality enhancement results of Zero-DCE are shown in Table 13. We sampled three images from the YouTube-w/o-ALI dataset to complete this visualization analysis.

As shown in Table 13, the original image contains more noise or blurred local regions, which hurts the final classification. However, there is a certain denoising effect after introducing Zero-DCE, which is beneficial for the subsequent feature distillation and can improve the training quality of the IQEFD model (Please refer to Table 12. According to ablation analysis, the Zero-DCE-based image enhancement module plays the most significant role in our model). Thus, the Zero-DCE network is effective and robust for our task, which creates a solid foundation for the subsequent feature fusion and feature distillation.

Feature visualization results

In this section, we validate the effectiveness of IQEFD from the perspective of feature visualization, which also helps enhance the interpretability of IQEFD. The features after the last pooling layer of the ConvNeXt, ViT, MBFN, and IQEFD models are visualized using t-SNE, respectively, where the visualization results are shown in Fig. 8.

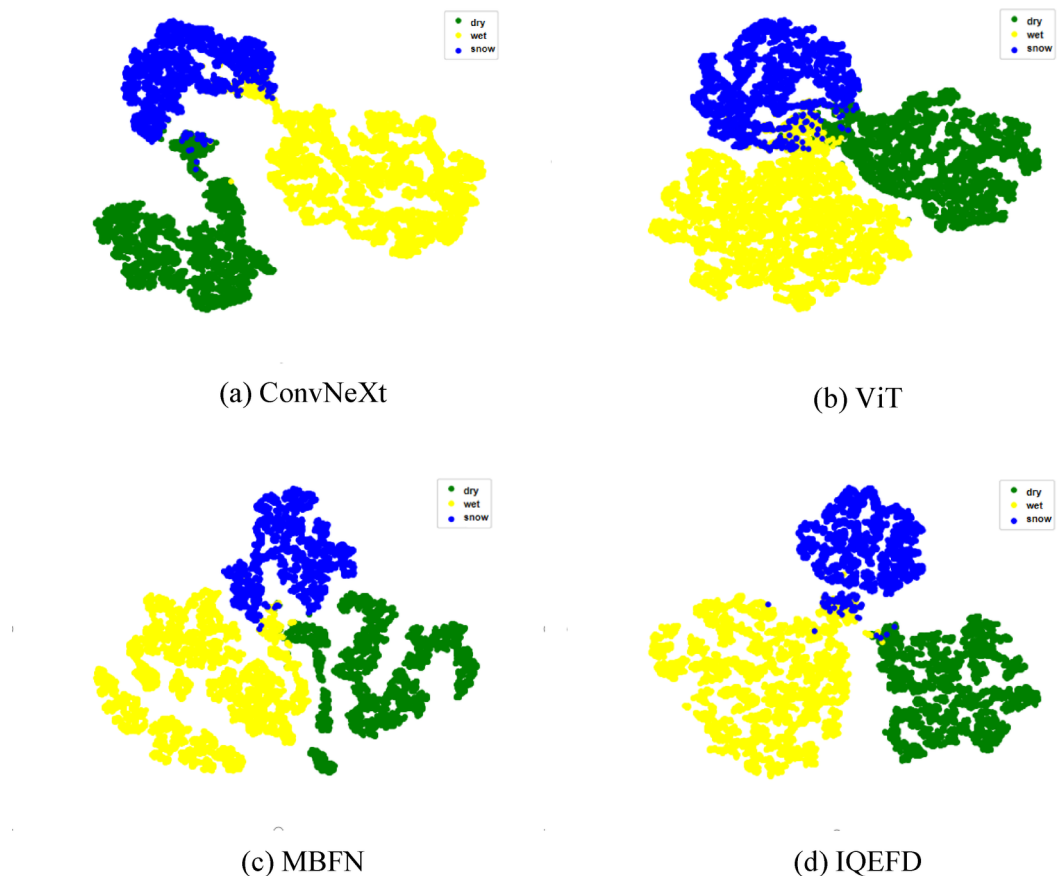


Fig. 8. Feature visualization results on the YouTube-w/o-ALI dataset.

As shown in Fig. 8, for the ConvNeXt and ViT models, it can be seen that different kinds of samples are mixed together, especially for ViT, which means that the features extracted by the pure deep learning backbone are insufficient for pavement condition identification. However, for the MBFN and IQEFD models, the 3 clusters are more concentrated, and the inter-class distance is larger while the intra-class distance is smaller, which helps create a clearer decision boundary and improves the final classification performance. We also find that it is difficult to classify the snow and wet samples as well as the dry and snow samples for the MBFN model. On the contrary, for the proposed IQEFD model, the classification confusion results are significantly reduced and more satisfactory intra-class and inter-class distances are observed, which helps promote the final recognition performance. In addition, it can be seen that the wet and snow categories are the most difficult categories to be classified by all the models. Hence, more attention needs to be paid to these two categories in future work. In summary, the feature visualization results validate the effectiveness of the IQEFD model from another significant perspective, which also provides a more intuitive visualization of our model and improves its interpretability. More importantly, the visualization results light up our future research directions.

Grad-CAM visualization results

To further support the superiority of our approach, we present the IQEFD visualization results from a more comprehensive perspective. First, we used the well-known CAM technique to visualize different variant operations to reveal their significance more intuitively, as shown in Fig. 9. To learn more about each variant, please refer to Table 12. Variant 5 represents the proposed model. We want to know whether our model could accurately locate the key local regions of the pavement images. As Fig. 9 shows, Variant 1 without any proposed module gets the worst pavement heat-map and some false detection results. Variant 2 combining HAM and feature distillation is better than Variant 1, but it also gets some false detections. Variant 3 considering Zero-DCE and feature distillation obtains relatively satisfactory visualization results due to material knowledge transfer and image enhancement. Unlike Variant 3, Variant 4 combining HAM and Zero-DCE can focus more on the key region areas of the pavement, which is significant for pavement condition identification. Evidently, our model (Variant 5) gets more sufficient context information and is more robust for nighttime condition, which is the best variant for our task.

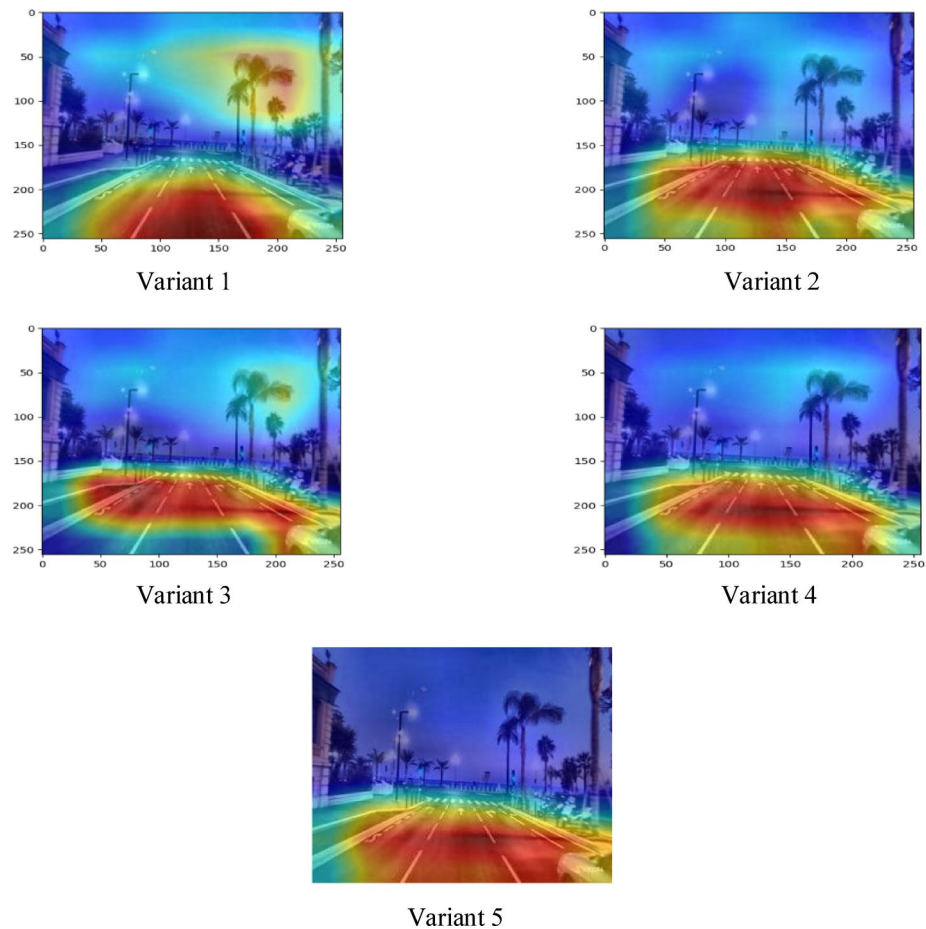


Fig. 9. Grad-CAM visualization results for each variant on the YouTube-w-ALI dataset.

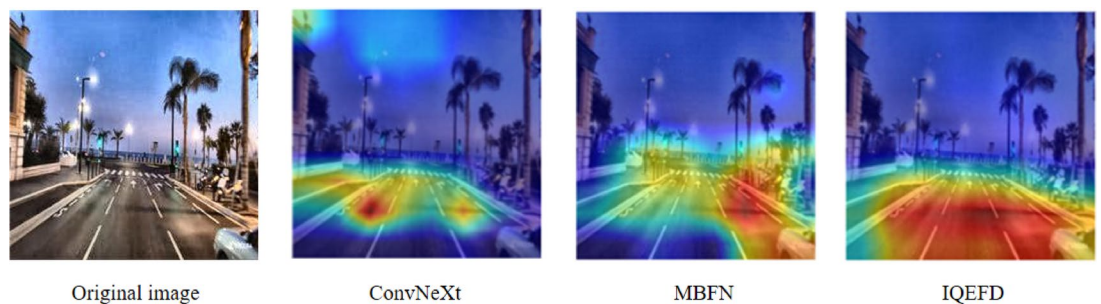


Fig. 10. Grad-CAM visualization results on the YouTube-w-ALI dataset.

Second, we exhibited the visualization results of ConvNeXt, MBFN, and IQEFD on the YouTube-w-ALI dataset, where the features in the layer before the classification head are selected for visualization, so as to more intuitively compare the ability of different models to capture key focal regions. The results are shown in Fig. 10.

As shown in Fig. 10, the ConvNeXt only focuses on some fragmented and localized pavement regions because the single ConvNeXt network is unable to extract global information. As another representative baseline, the MBFN model can better detect the key pavement regions compared to ConvNeXt, which lays an important foundation for improving the corresponding recognition performance. However, some critical pavement regions like the left two lanes are still ignored by MBFN. As expected, the IQEFD model accurately detects most pavement regions which is important for the subsequent pavement condition identification. As described above, IQEFD captures both the channel and spatial information from pavement images using the HAM module. Moreover, the proposed bidirectional fusion module retains multi-scale semantics information from the successive up-sampling and down-sampling operations. All these enrich the proposed model with a kind of capability of extracting the global contextual information, thus detecting the key regions of the pavement

images and improving the final recognition performance. Summarily, the Grad-CAM visualization results help better support the effectiveness of our model. More significantly, these results enhance the interpretability of the IQEFD model.

Conclusions and future work

In this study, we proposed the novel idea of transferring enhanced material knowledge by constructing the IQEFD model for pavement condition identification. The model achieves the accuracies of 98.04% and 98.68% on the YouTube-w-ALI and YouTube-w/o-ALI datasets, respectively, outperforming the state-of-the-art baselines. We aimed to reduce the negative effect of image quality and the loss of context information. First, the image enhancement network Zero-DCE was employed to enhance the quality of pavement conditions images. Then we designed the bidirectional fusion module embedded with a HAM module to extract the refined feature containing multi-scale semantics information. Last, we bridged the gap between the image enhancement network and the feature extraction network. We used the enhanced features to guide the optimization procedure of the fused attention features through online feature distillation. Extensive experimental results showed that the IQEFD model not only outperforms the mainstream baselines but also showed powerful robustness and generalization ability. It can better process the pavement condition images under very complex light environments. Moreover, we further validate the effectiveness of our model using diverse visualization methods, which also enhanced the model's interpretability. Summarily, the idea of transferring enhanced material knowledge works well, which provides a novel insight into pavement condition identification.

Although the IQEFD model obtains satisfactory performance in pavement condition identification, there is still a room for improvement in the model. Our method may overlook the extraction of specific texture information which is significant to depict pavement conditions. Meanwhile, the proposed model ignores other modalities like textual content. The proposed model could combine the image modality with the textual descriptions to provide more comprehensive but complementary contextual information to improve the final accuracy.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author Zejiu Wu on reasonable request via e-mail wzjhdjd@qq.com.

Received: 19 November 2024; Accepted: 11 April 2025

Published online: 21 April 2025

References

- Bellone, M. et al. *Autonomous Driving in the real-world: the Weather Challenge in the Sohjoa Baltic Project Towards Connected and Autonomous Vehicle Highways: Technical, Security and Social Challenges* 229–255 (Springer, 2021).
- Banerjee, D., Upadhyay, D. & Rawat, R. S. *Enhanced Road Surface Classification using CNN and Random Forest Models*. Paper presented at the 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS). (2024).
- Observatory, E. R. S. *Annual Accident Report 2018* (European Road Safety Observatory Brussels, 2018).
- Seo, J. W., Kim, J. S., Yang, J. H. & Chung, C. C. A Spatiotemporal Deep Learning Architecture for Road Surface Classification Using LiDAR in Autonomous Emergency Braking Systems. *IEEE Access* (2023).
- Eriksson, J. et al. *The pothole patrol: using a mobile sensor network for road surface monitoring*. Paper presented at the Proceedings of the 6th international conference on Mobile systems, applications, and services. (2008).
- Ruiz-Llata, M., Rodríguez-Cortina, M., Martín-Mateos, P., Bonilla-Manrique, O. E. & López-Fernández, J. R. *LiDAR design for road condition measurement ahead of a moving vehicle*. Paper presented at the 2017 IEEE Sensors (2017).
- Casselgren, J., Sjö Dahl, M. & LeBlanc, J. P. Model-based winter road classification. *Int. J. Veh. Syst. Model. Test.* 7(3), 268–284 (2012).
- Jokela, M., Kutila, M. & Le, L. Road condition monitoring system based on a stereo camera. In *Paper presented at the 2009 IEEE 5th International conference on intelligent computer communication and processing* (2009).
- Jonsson, P., Casselgren, J. & Thörnberg, B. Road surface status classification using spectral analysis of NIR camera images. *IEEE Sens. J.* 15(3), 1641–1656 (2014).
- Shin, J., Park, H. & Kim, T. Characteristics of laser backscattering intensity to detect frozen and wet surfaces on roads. *J. Sens.* 2019(1), 8973248 (2019).
- Zhao, J., Wu, H. & Chen, L. Road surface state recognition based on SVM optimization and image segmentation processing. *J. Adv. Transp.* 2017(1), 6458495 (2017).
- Marianingsih, S. & Utaminigrum, F. Comparison of support vector machine classifier and Naïve Bayes classifier on road surface type classification. In *Paper presented at the 2018 International Conference on Sustainable Information Engineering and Technology (SIET)* (2018).
- Fauzi, A. A., Utaminigrum, F. & Ramdani, F. Road surface classification based on LBP and GLCM features using kNN classifier. *Bull. Electr. Eng. Inf.* 9(4), 1446–1453 (2020).
- Kaya, V. & Akgül, I. Detection of defects in printed circuit boards with machine learning and deep learning algorithms. *Avrupa Bilim Ve Teknoloji Dergisi* 41, 183–186. (2022).
- Min, X. et al. Exploring rich subjective quality information for image quality assessment in the wild. *arXiv preprint arXiv:2409.05540*. (2024).
- Min, X. et al. Blind quality assessment based on pseudo-reference image. *IEEE Trans. Multimed.* 20(8), 2049–2062 (2017).
- Min, X. et al. Unified blind quality assessment of compressed natural, graphic, and screen content images. *IEEE Trans. Image Process.* 26(11), 5462–5474 (2017).
- Hashemzadeh, M., Asheghi, B. & Farajzadeh, N. Content-aware image resizing: an improved and shadow-preserving seam carving method. *Sig. Process.* 155, 233–246 (2019).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.*, 25 (2012).
- Lochan, R. N., Tomar, A. S. & Srinivasan, R. Plant detection and classification using fast region-based convolution neural networks. In *Paper presented at the Artificial Intelligence and Evolutionary Computations in Engineering Systems* (2020).
- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. (2014).

22. Iandola, F. N. et al. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and 5 MB model size (2016).
23. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
24. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. *Densely connected convolutional networks*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition. (2017).
25. Liu, Z. et al. A convnet for the 2020s. In *Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022).
26. Cheng, L., Zhang, X. & Shen, J. Road surface condition classification using deep learning. *J. Vis. Commun. Image Represent.* **64**, 102638 (2019).
27. Garcea, F. et al. Self-supervised and semi-supervised learning for road condition Estimation from distributed road-side cameras. *Sci. Rep.* **12**(1), 22341 (2022).
28. Zhang, H. et al. Multi-supervised bidirectional fusion network for road-surface condition recognition. *PeerJ Comput. Sci.* **9**, e1446 (2023).
29. Wang, P. Research on comparison of lidar and camera in autonomous driving. *J. Phys. Conf. Ser.* (2021).
30. Guo, C. et al. Zero-reference deep curve estimation for low-light image enhancement. In *Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020).
31. Liu, C. & Gu, J. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(1), 86–98 (2013).
32. Kumar, R., Jones, M. & Marks, T. K. Morphable reflectance fields for enhancing face recognition. In *Paper presented at the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2010).
33. Li, W. & Fritz, M. Recognizing materials from virtual examples. In *Paper presented at the Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7–13, 2012, Proceedings, Part IV 12 (2012).
34. Liu, C., Sharan, L., Adelson, E. H. & Rosenholtz, R. Exploring features in a bayesian framework for material recognition. In *Paper presented at the 2010 IEEE computer society conference on computer vision and pattern recognition* (2010).
35. Sharan, L., Liu, C., Rosenholtz, R. & Adelson, E. H. Recognizing materials using perceptually inspired features. *Int. J. Comput. Vision* **103**, 348–371 (2013).
36. Bell, S., Upchurch, P., Snavely, N. & Bala, K. Material recognition in the wild with the materials in context database. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2015).
37. Zhang, H., Xue, J. & Dana, K. Deep ten: Texture encoding network. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
38. Zhang, H. et al. Novel framework for image attribute annotation with gene selection XGBoost algorithm and relative attribute model. *Appl. Soft Comput.* **80**, 57–79 (2019).
39. Asheghi, B., Salehpour, P., Khiavi, A. M., Hashemzadeh, M. & Monajemi, A. DASOD: Detail-aware salient object detection. *Image Vis. Comput.* **148**, 105154 (2024).
40. Ullah, A., Javaid, N., Asif, M., Javed, M. U. & Yahaya, A. S. Alexnet, adaboost and artificial bee colony based hybrid model for electricity theft detection in smart grids. *IEEE Access* **10**, 18681–18694 (2022).
41. Kim, S., Lee, J. & Yoon, T. Road surface conditions forecasting in rainy weather using artificial neural networks. *Saf. Sci.* **140**, 105302 (2021).
42. Omer, R. & Fu, L. An automatic image recognition system for winter road surface condition classification. In *Paper presented at the 13th international IEEE conference on intelligent transportation systems* (2010).
43. Smolyakov, D. & Burnaev, E. Software System for Road Condition Forecast Correction. *arXiv preprint arXiv:2003.09957* (2020).
44. Roychowdhury, S., Zhao, M., Wallin, A., Ohlsson, N. & Jonasson, M. *Machine learning models for road surface and friction estimation using front-camera images*. Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN) (2018).
45. Fink, D., Busch, A., Wielitzka, M. & Ortmaier, T. Resource efficient classification of road conditions through CNN pruning. *IFAC-PapersOnLine* **53**(2), 13958–13963 (2020).
46. Guo, K., He, C., Yang, M. & Wang, S. A pavement distresses identification method optimized for YOLOv5s. *Sci. Rep.* **12**(1), 3542 (2022).
47. Jiang, J. et al. High-accuracy road surface condition detection through multi-sensor information fusion based on WOA-BP neural network. *Sens. Actuators A: Phys.* **378**, 115829 (2024).
48. Karunasekera, H. & Sjöberg, J. In Search for Better Road Surface Condition Estimation–Using Non-Road Image Region. In *Paper presented at the 2024 IEEE Intelligent Vehicles Symposium (IV)* (2024).
49. Coltuc, D., Bolon, P. & Chassery, J. M. Exact histogram specification. *IEEE Trans. Image Process.* **15**(5), 1143–1152 (2006).
50. Ibrahim, H. & Kong, N. S. P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **53**(4), 1752–1758 (2007).
51. Lee, C., Lee, C. & Kim, C. S. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans. Image Process.* **22**(12), 5372–5384 (2013).
52. Stark, J. A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **9**(5), 889–896 (2000).
53. Guo, X., Li, Y. & Ling, H. LIME: Low-light image enhancement via illumination map Estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2016).
54. Park, S., Yu, S., Moon, B., Ko, S. & Paik, J. Low-light image enhancement using variational optimization-based retinex model. *IEEE Trans. Consum. Electron.* **63**(2), 178–184 (2017).
55. Li, M., Liu, J., Yang, W., Sun, X. & Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **27**(6), 2828–2841 (2018).
56. Gu, Z., Li, F., Fang, F. & Zhang, G. A novel retinex-based fractional-order variational model for images with severely low light. *IEEE Trans. Image Process.* **29**, 3239–3253 (2019).
57. Ren, X., Yang, W., Cheng, W. H. & Liu, J. LR3M: robust low-light enhancement via low-rank regularized retinex model. *IEEE Trans. Image Process.* **29**, 5862–5876 (2020).
58. Hao, S., Han, X., Guo, Y., Xu, X. & Wang, M. Low-light image enhancement with semi-decoupled decomposition. *IEEE Trans. Multimed.* **22**(12), 3025–3038 (2020).
59. Lore, K. G., Akintayo, A. & Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recogn.* **61**, 650–662 (2017).
60. Lv, F., Lu, F., Wu, J. & Lim, C. MBLEN: Low-light image/video enhancement using cnns. Paper presented at the Bmvc (2018).
61. Jiang, Y. et al. Enlightengan: deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **30**, 2340–2349 (2021).
62. Zhu, A. et al. Zero-shot restoration of underexposed images via robust retinex decomposition. In *Paper presented at the 2020 IEEE International Conference on Multimedia and Expo (ICME)* (2020).
63. Liu, R., Ma, L., Zhang, J., Fan, X. & Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021).
64. Zhai, G. & Min, X. Perceptual image quality assessment: A survey. *Sci. China Inform. Sci.* **63**, 1–52 (2020).
65. Hinton, G. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531* (2015).

66. Zhang, Y., Xiang, T., Hospedales, T. M. & Lu, H. Deep mutual learning. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2018).
67. Mirzadeh, S. I. et al. Improved knowledge distillation via teacher assistant. In *Paper presented at the Proceedings of the AAAI conference on artificial intelligence* (2020).
68. Tung, F. & Mori, G. Similarity-preserving knowledge distillation. In *Paper presented at the Proceedings of the IEEE/CVF international conference on computer vision* (2019).
69. Kim, J., Park, S. & Kwak, N. Paraphrasing complex network: Network compression via factor transfer. *Adv. Neural. Inf. Process. Syst.* **31** (2018).
70. Chen, D. et al. Cross-layer distillation with semantic calibration. In *Paper presented at the Proceedings of the AAAI conference on artificial intelligence* (2021).
71. Tian, Y., Krishnan, D. & Isola, P. Contrastive representation distillation. *arXiv preprint arXiv:1910.10699*. (2019).
72. Yang, J., Martinez, B., Bulat, A. & Tzimiropoulos, G. Knowledge distillation via softmax regression representation learning (2021).
73. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2018).
74. Gao, Z., Xie, J., Wang, Q. & Li, P. Global second-order pooling convolutional networks. In *Paper presented at the Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition* (2019).
75. Wang, Q. et al. ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020).
76. Lee, H., Kim, H. E. & Nam, H. Srm: A style-based recalibration module for convolutional neural networks. In *Paper presented at the Proceedings of the IEEE/CVF International conference on computer vision* (2019).
77. Mnih, V., Heess, N. & Graves, A. Recurrent models of visual attention. *Adv. Neural. Inf. Process. Syst.* **27** (2014).
78. Jaderberg, M., Simonyan, K. & Zisserman, A. Spatial transformer networks. *Adv. Neural. Inf. Process. Syst.* **28** (2015).
79. Wang, F. et al. Residual attention network for image classification. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
80. Park, J. Bam: Bottleneck attention module. *arXiv preprint arXiv:1807.06514* (2018).
81. Misra, D., Nalamada, T., Arasanipalai, A. U. & Hou, Q. Rotate to attend: Convolutional triplet attention module. In *Paper presented at the Proceedings of the IEEE/CVF winter conference on applications of computer vision* (2021).
82. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Paper presented at the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18* (2015).
83. Lin, T. Y. et al. Feature pyramid networks for object detection. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
84. Zhang, H. et al. Mining sufficient knowledge via progressive feature fusion for efficient material recognition. *Sci. Program.* **2021**(1), 8971349 (2021).
85. Zhang, H., Sehab, R., Azouigui, S. & Boukhni, M. Application and comparison of deep learning methods to detect night-time road surface conditions for autonomous vehicles. *Electronics* **11**(5), 786 (2022).
86. Alexey, D. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv: 2010.11929* (2020).
87. Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep neural networks. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
88. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Paper presented at the International Conference on Machine Learning* (2019).
89. Yang, J., Li, C., Dai, X. & Gao, J. Focal modulation networks. *Adv. Neural. Inf. Process. Syst.* **35**, 4203–4217 (2022).
90. Zhang, H. et al. Gathering effective information for real-time material recognition. *IEEE Access* **8**, 159511–159529 (2020).
91. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

Acknowledgements

We should give thanks to the authors for collecting and organizing the MattrSet, YouTube-W-ALI and YouTube-w/o-ALI datasets, and we also would like to thank the editor and the reviewers for their helpful suggestions.

Author contributions

Z.W.: Conceptualization, Validation, Investigation, Writing—Review & Editing, Supervision. Y.Z.: Software, Resources, Investigation, Formal Analysis, Writing—Original Draft. B.L.: Software, Resources, Investigation, Formal Analysis, Writing—Original Draft. Z.L.: Software, Validation, Methodology, Formal Analysis, Investigation, Resources, Visualization, Data Curation, Writing—Original Draft. D.J.: Conceptualization, Formal analysis. H.Z.: Conceptualization, Project Administration, Validation, Investigation, Funding Acquisition.

Funding

This research was partly funded by the National Natural Science Foundation of China (Grant Nos. 62361027 and 62161011), the Key Research and Development Plan of Jiangxi Provincial Science and Technology Department (Key Project) (Grant No. 20223BBE51036), the Humanity and Social Science Fund of Ministry of Education of China (Grant No. 23YJA870005), the Natural Science Foundation of Jiangxi Provincial Department of Science and Technology (Grant Nos. 20224BAB202016, 20232BAB202022, and 20232BAB202004).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025