An Improving Blue-Green Algae Detection in Water Bodies Using ResNet50

Pranita V. Shitole
M.E. Student,
(Artificial Intelligence and Data science)
AISSMS COE, Pune,
Maharashtra, India

Shashikant .V. Athawale
Associate Professor,
Department of Computer Engineering,
AISSMS COE, Pune,
Maharashtra, India

---***---

Abstract

Blue-green algae blooms are a growing problem that have a significant impact on human communities due to the extreme transforms in the state of the ecosystem. Deep learning models are essential for real-time image data extraction and classification to detect blue-green algae in water bodies and enable subsequent alerts and interventions. However, there are several obstacles that present studies in this area must overcome, starting with the fact that standard convolutional methods are not very good at extracting features like patterns and outlines from images. Second, classification errors frequently result from the subject and surrounding areas of collected pictures sharing comparable settings. In order to deal with these problems, this work first integrates a detail feature extraction module into the ResNet50 architecture, improving its feature representation capabilities. A binary mask image fusion technique is then suggested to centre the model's attention on acquiring water region features. When viewed alongside previous classification models, our built model performs better in real-time camera data classification.

Keywords: ResNet;ResNet50; UNet ; Enviroment Evaluation; Image classification

Introduction:

The issue of eutrophication in water bodies is growing more significant as modernization and rising populations pick up speed. The leading and dangerous condition is blue-green algae blooms. In the past, the extent and location of blue-green algae outbreaks were assessed by human monitors. Due to manpower , labor constraints, this approach finds it difficult to offer thorough examination of every possible epidemic site. In order to enhance the effectiveness and coverage of blue-green algae handling and offer a more scientific and effective solution to the water bloom crisis, this paper investigates the application of image classification technology, employing deep learning techniques that effortlessly identify and classify image data.

Convolutional neural networks' (CNNs') extensive past and ongoing improvements in technology contribute to their sophisticated use in picture classification. CNNs were first inspired by the 19th-century "receptive fields" idea, which has been shown to be successful in a variety of models [1]. LeNet-5, for example, showed up CNNs' promise in handwritten digit recognition [2], and AlexNet introduced ReLU and Dropout approaches to address overfitting and vanishing gradients [3]. Through the use of deeper networks and novel technologies like skip connections, later models like VGG and ResNet improved performance [4][5]. Accuracy and training speed were enhanced by the Inception series' optimized structures and technologies, such as batch normalization [6][7]. Lightweight depthwise separable convolutions are used by MobileNet, which is optimized for mobile devices [8]. Models' ability to deal with features has been further improved by the use of focus methods like Residual Attention Networks and SENet [9], [10], and [11].

A well-known deep learning model for image recognition, ResNet50 belongs to the ResNet family of models. The bottleneck modules that make up the model enable it to function effectively. Bottleneck modules are unique network building components that enable the model to learn complex characteristics while lowering the amount of parameters (the parts of the model that must be learned).Three different kinds of bottleneck modules—three, four, six, and three bottleneck blocks—are employed in the second and fifth layers of the ResNet50 model.

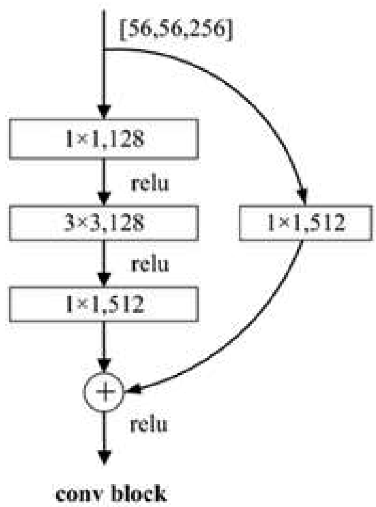


Fig. 1. Two different types of Block structures

We further investigate the use of the ResNet50 architecture in the identification and classification of blue-green algae, suggesting two improvements, even if residual network models excel at picture classification tasks:

(a) a multi-channel fusion technique for model input; and

(b) a detail extraction module built into the original residual blocks to capture complex image elements like textures and contours.

The following succinctly describes our primary contributions:
To enhance and extract delicate features, our detail feature extraction module uses differential convolution to collect the gradient information of images. The model structure is stabilized through the incorporation of an attention mechanism. In order to direct the model to concentrate more on the target area, we added a binary mask to the original input dimensions.

Related Work：

Concatenation, summation, and multiplication are the main ways that channel fusion approaches improve the accuracy of models in deep learning. As is frequently observed in encoder-decoder structures where low-level and high-level features are mixed, concatenation enables the combining of feature maps at multiple points in a given dimension, improving the model's visualization of features capabilities. Summation fusion helps balance feature contributions and lessen noise interference by adding numerous feature maps element-wise. Through element-wise multiplication, multiplication fusion highlights the significance of features and improves the model's capacity to collect meaningful information and details. Each of these approaches has practical uses and benefits, yet they all process information differently. In contrast to earlier research, we re-examined the model input's characteristics. We used combination in a feature fusion strategy to better focus the model on the target area without appreciably increasing model complexity. This technique keeps the original background data while enabling later steps to focus the model's attention on the desired region.

Local Binary Patterns (LBP)[12], a texture classification technique that translates local pixel variations into decimal values, is where the idea of differential convolution originated. Following the broad success of Convolutional Neural Networks (CNNs) in image tasks, Xu et al. [13] expanded on this idea by introducing Local Binary Convolution (LBC), which combines linear convolution layers with nonlinear activation functions to encode pixel differences. Additionally, Centered Differential Convolution (CDC), which enables the direct encoding of pixel differences with fully learnable differential weights, was introduced by Yu et al. [14]. Other forms of differential convolution, including Cross-Centered Differential Convolution[15] and Pixel Differential Convolution[16], were later proposed by researchers. These methods improve the efficiency of feature extraction by capturing gradient information in photos. In order to enhance efficiency, we accordingly thought about adding differential convolution to our image classification model.

Methodology :

3.1. Fusion of Binary Layers :

Increasing the correctness of model training is frequently hampered by background effects in real-world detection scenarios. Our goal was to highlight the target area by altering the model's input section. In order to distinguish the target water body from other backgrounds, we first used a segmentation method. As seen in Figure 2, tests showed that the Unet model worked well for segmenting our dataset and provided quick prediction results (the Unet model is solely utilized for dataset construction in this case).

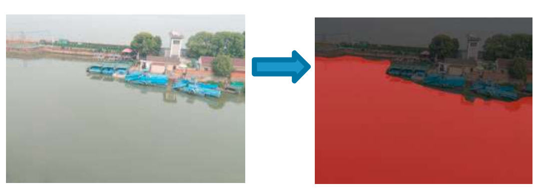
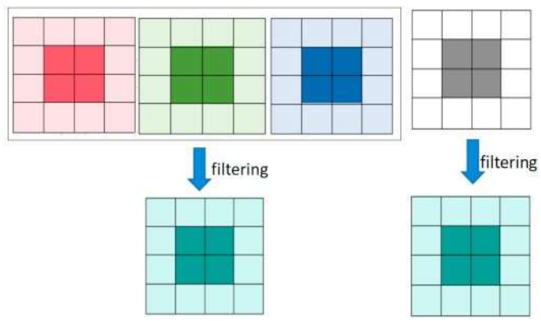


Fig. 2. Segmentation results

Unet was then used to segment the collected data, and the generated masks were used for training or prediction with the initial pictures. We kept the raw picture data in its original RGB three-channel format. For example, the target mask area was made brighter and the surrounding mask region was made less bright in order to apply the segmentation-generated mask to the initial picture. Figure 3(a) illustrates the additional RGB channel components that were created by weighting the original pictures based on their mask results.

Assuming that these inputs go via a filter in the model, where the convolution results at corresponding places accumulate and the middle position solutions stay greater compared to the edges, the various hues of color indicate the backdrop and target. The extra data produced by the mask effect is precisely equivalent to a binary mask in this processing. The background and target areas, separated by the binary mask, are filtered to produce effects akin to those described, as seen in Figure 3(b). This lowers the complexity of computation while successfully emphasizing the target area.



(a) RGB channels (b) binary mask.

Fig. 3.The convolution results of RGB channels and binary mask.

This study created a four-channel input by combining the original RGB three channels with a converted single-channel grayscale mask binary mask to guarantee the model's input maintains the original visual information while adding more data. The convolution results over its target region add to the original data, changing the feature matrix of that area and continually impacting subsequent convolutional processes, while the background area of the mask usually stays the same after convolution. The original image in the R, G, and B channels as well as the mask in a grayscale image are shown in Figure 4; the various extracted grayscale values match the corresponding original pixel values of red, green, and blue. We anticipate that this approach will direct later modules to concentrate on feature extraction from the target area without appreciably raising the computing load.

(a) R channel (b) G channel (c) B channel (d) mask

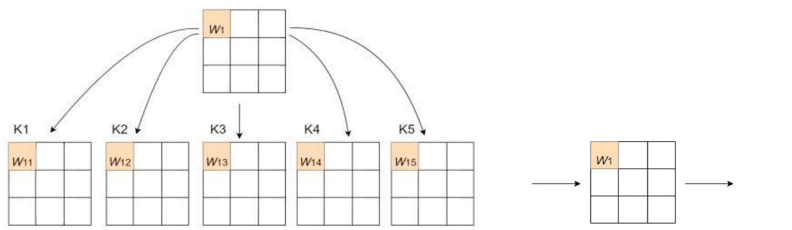
Fig. 4. The gray level of the four channel

Detail Feature Extraction Module

High-frequency details like edges and contours in images are frequently difficult for standard convolution layers to capture. We created an Edge Enhanced Convolution (EEConv) layer in response to research that demonstrated the use of edge priors in dehazing models [17], [18], and [19]. In order to accurately define blurred borders and properly extract texture features from blue-green algae patches, this layer incorporates previously collected data.

Four differential convolution layers [20] and one conventional convolution layer were coupled in our built EEConv to extract information concurrently. Heterogeneous convolution improves explanatory capacity and generalization by first calculating pixel differences before convolving, as per earlier research [21], [22]. The Center Differential Convolution (CDC), Angular Differential Convolution (ADC), Horizontal Differential Convolution (HDC), and Vertical Differential Convolution (VDC) are the differential convolution layers in EEConv [23]. In order to increase the model's capacity for detail recognition, these layers concentrate on improving gradient information. We also investigated the idea of combining several parallel convolution kernels into a single standard kernel at matching points in order to streamline computation and increase efficiency.

The weights of the convolutional kernels in the EEConv can be individually modified applying gradient descent throughout backpropagation, as shown in Figure 5(a). The altered convolutional kernel weights in the forward propagation process are determined by adding the weights at each location, as seen in Figure 5(b).



(a)Back propagation (b)forward propagation

Fig. 5. The forward and back propagation processes of EEConv.

We foresee combining the Convolutional Block Attention Module (CBAM)[25], EEConv layer, and the 3x3 convolution of the Bottleneck module to produce the proposed EEA-Block, as shown in Figure 6, in light of the effective integration of focus methods in earlier ResNet architectures[24]. The Channel Attention Module (CAM) and the Spatial Attention Module (SAM), two separate sub-modules of the CBAM, carry out focus activities on the channel and spatial dimensions, namely.

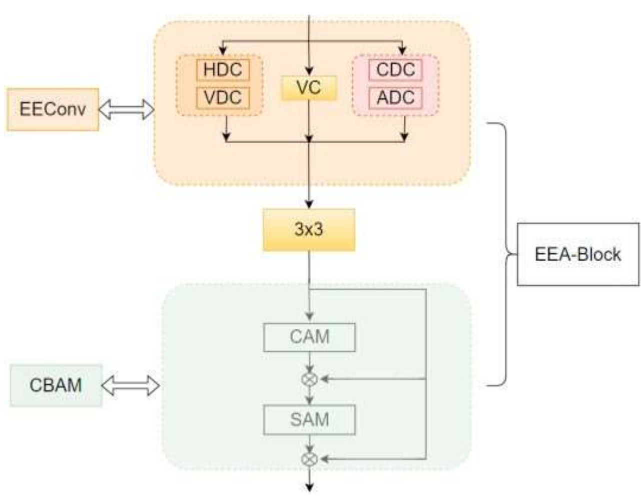


Fig. 6.The overall framework of EEA-Block

Experiments Setup and Results:

Training Configuration:

The PyTorch framework and an NVIDIA RTX 2070 GPU were used in the studies, which were carried out on a 64-bit Windows 10 computer. A batch size of 32 was chosen, and the starting learning rate was set at 1e-4. The original image was preprocessed by resizing its shorter side to 256 pixels and then center cropping it to create a 224x224 pixel rectangle for training.

Dataset Construction and Metrics:

The 900 real-time photographs taken by cameras surrounding Lake Taihu are used to generate the experimental data; these images contain intricate background details and satisfy the specifications for actual environmental monitoring. There are 600 training photos, 200 validation images, and 100 test images in the dataset. Figure 9 displays the monitoring areas. We uniformly gathered photos from four places, keeping an equal amount of photographs in and out of blue-green algae to guarantee consistent model training and the accuracy of experimental outcomes.

(a) Region 1 (b) Region 2 (c) Region 3 (d)Region 4

Fig. 7. Represents each of the four regions

Metrics including Accuracy, Precision, F1-Score, and Recall were used to assess the effectiveness of the ResNet50 model as well as comparable models like ShuffleNet and DenseNet. FN stands for a false negative, FP for a false positive, TN for a true negative, and TP for a true positive. These are known as False Negatives (FN), False Positives (FP), True Negatives (TN), and True Positives (TP).

Experiments Results:

To put the suggested EEA-Block structure into action, we made related changes to the Bottleneck core of the ResNet50 network design. Figure 8 shows one of these altered Blocks.

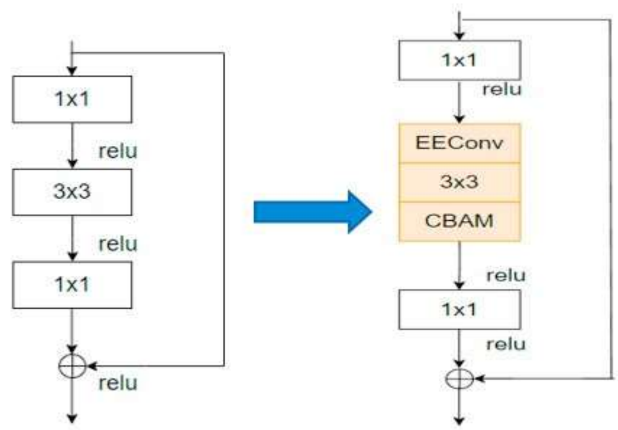
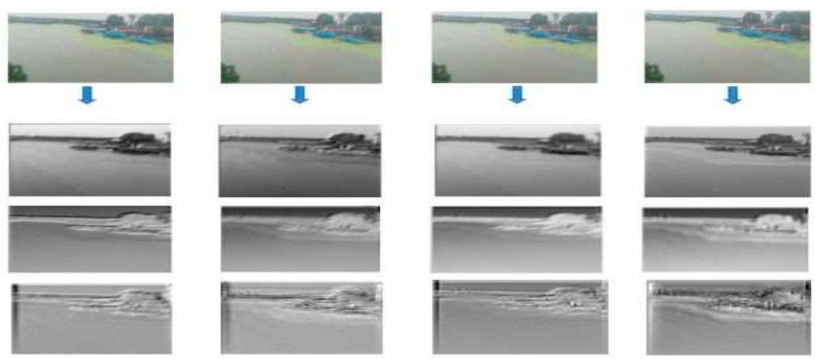


Fig. 8. Improvements to feature extraction in the Block

We show the outcomes of numerous module blends in order to illustrate the impact of feature maps following the addition of various modules to ResNet50. In order to more clearly highlight crucial aspects, we chose outputs from the early to mid-stages of the structure and displayed feature maps in grayscale. Both the original model and the model with the CBAM module added exhibit less reactivity to the blue-green algae area, as seen in Figure 9. The EEA-Block is more sensitive to the textural characteristics of the blue-green algae waters, which helps to highlight the borders and textures of the algae sections. The addition of the EEConv module enhances the results compared to the first two.



(a)original model (b) EEConv (c) CBAM (d)EEA-Block

Fig. 9. The effects of different modules

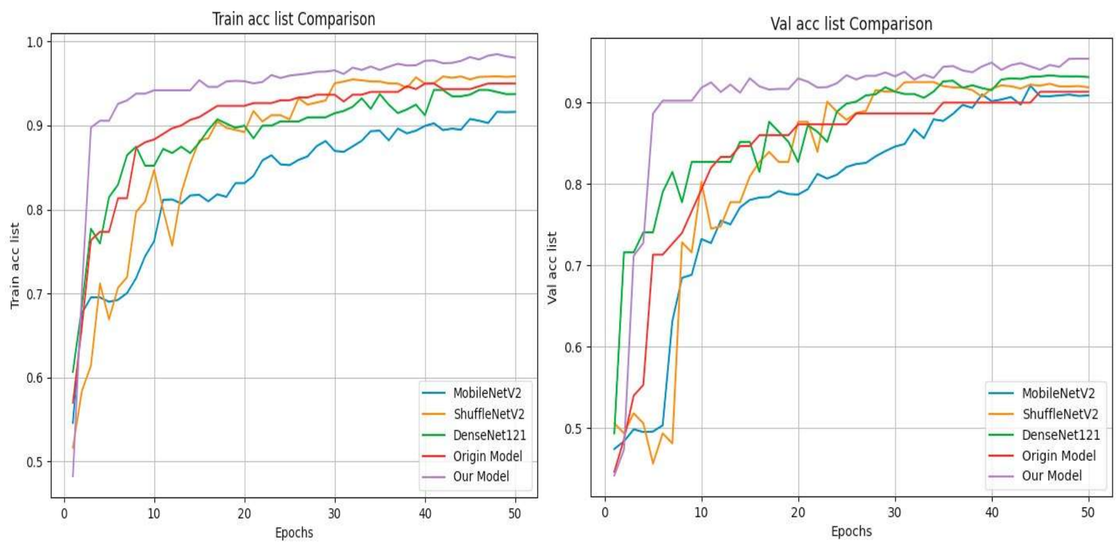
We carried out comparison studies with various module combinations, as indicated in Table 1, in order to confirm the influence of the various components suggested in the article on experimental results:

Table 1. Module Performance Comparison.

Model	Binary Mask	CBAM	EEConv	Precision	F1
				0.9568	0.8977
	√			0.9582	0.9080
ResNet50	√	√		0.9631	0.9101
	√		√	0.9718	0.9380
	√	√	√	0.9826	0.9570

It is clear via the findings of Table 1 that the pairing of Mask and CBAM partially supports our hypothesis, with the Mask influencing the network's attention to target areas; the utmost efficient pairing in terms of model balance and classification results is the use of Mask and EEA-Block; the combination of EEConv and Mask performs better, demonstrating enhancements in evaluation metrics.

The mask and EEA-Block pair suggested in this research was tested against MobileNetv2, ShuffleNetV2, DenseNet121, and the original ResNet50 model under identical testing conditions in order to verify its performance. Test accuracy, precision, and F1 score were all significantly enhanced by the upgraded ResNet50 network, especially when compared to MobileNetv2, where the gains were most pronounced. The analysis of the accuracy curves during the training and validation procedures is shown in Figure 10. The improved ResNet50 network outperforms the original model in identifying and categorizing datasets of blue-green algal water bodies, according to a thorough evaluation.



(a)Training accuracy (b) Verification accuracy

Fig. 10. Experiment results

Table 2. Results of comparison between models.

Compare module/indicator	Accuracy	Precision	F1	Recall
MobileNetV2	0.9137	0.8722	0.8730	0.8738
ShuffleNetV2	0.9610	0.9356	0.9334	0.9312
DenseNet121	0.9641	0.9401	0.9369	0.9337
Origin Model	0.9568	0.9024	0.8977	0.8931
Our Model	0.9826	0.9581	0.9570	0.9559

Conclusion:

By adding variable convolutions and attention methods to the conventional ResNet50 residual network, this article greatly increases the detection accuracy of blue-green algae in aquatic bodies. In order to improve system interoperability and classification efficiency, the next study will concentrate on streamlining network parameters and accelerating computations to facilitate implementation on industrial devices with constrained resources. Additionally, we intend to investigate Vision Transformer (ViT) models, which might provide improved performance and parallelism. This will entail a comparative analysis to assess the advantages and disadvantages of ViT use. To further improve feature representation, we are further looking into more sophisticated attention processes outside of the Convolutional Block Attention Module (CBAM). In order to further appreciate the role of each component and to further develop our method for improved environmental monitoring and management, subsequent research will involve more thorough ablation experiments.

References:

[1] SHERRINGTON C S. Observations on the scratch-reflex in the spinal dog[J]. The Journal of Physiology, 1906, 34(1-2):1-50.

[2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] KRIZHEVSKY A, SUTSKEVER I, E. HINTON G. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada. Red Hook: Curran Associates Inc, 2012: 1097-1105.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference of Learning Representation, May 7-9, 2015, San Diego, CA. arXiv:1409.1556v6 [cs.CV] , 2015.

[5] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, Nevada. Los Alamitos: IEEE Computer Society, 2016: 770-778.

[6] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition, Juny 7-12, 2015, Boston, MA, USA. Piscataway: IEEE Press, 2015: 1-9.

[7] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning, July 6-11, 2015, Lile, France. [S.l.: s.n.], 2015: 448-456.

[8] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861v1 [cs.CV] .2017.

[9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//31st Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. [S.l.: s.n.], 2017.

[10] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. Piscataway: IEEE Press, 2017: 6450-6458.

[11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, New York, USA. Piscataway: IEEE Press, 2018: 7132-7141.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971–987, 2002.

[13] F. Juefei-Xu, V. Naresh Boddeti, and M. Savvides, "Local binary convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 19–28.

[14] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face antispoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5295–5305.

[15] Z. Yu, Y. Qin, H. Zhao, X. Li, and G. Zhao, "Dual-cross central difference network for face anti-spoofing," in Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, Z.-H. Zhou, Ed. International Joint Conferences on Artificial Intelligence Organization, 8 2021, pp. 1281–1287, main Track. [Online]. Available: https://doi.org/10.24963/ijcai.2021/177

[16] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietik¨ainen, and L. Liu, "Pixel difference networks for efficient edge detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5117–5127.

[17] H. Zhang and V. M. Patel, “Densely connected pyramid dehazing network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3194–3203.

[18] H. Bai, J. Pan, X. Xiang, and J. Tang, “Self-guided image dehazing using progressive feature fusion,” IEEE Transactions on Image Processing, vol. 31, pp. 1217–1229, 2022.

[19] C. Wang, H.-Z. Shen, F. Fan, M.-W. Shao, C.-S. Yang, J.-C. Luo, and L.-J. Deng, “Eaa-net: A novel edge assisted attention network for single image dehazing,” Knowledge-Based Systems, vol. 228, p. 107279, 2021.

[20] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, “Searching central difference convolutional networks for face antispoofing,” in CVPR, 2020, pp. 5294–5304.

[21] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietik¨ainen, and L. Liu, “Pixel difference networks for efficient edge detection,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5117–5127.

[22] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, “Nas-fas: Staticdynamic central difference network search for face anti-spoofing,” IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 9, pp. 3005–3023, 2020.

[23] Chen Z, He Z, Lu Z M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024.

[24] Hassan S M, Maji A K. Pest Identification based on fusion of Self-Attention with ResNet[J]. IEEE Access, 2024.

[25] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.