Journal of Scientific & Industrial Research
Vol. 82, February 2023, pp. 192-201
DOI: 10.56042/jsir.v82i2.70219

Learning How to Detect Salient Objects in Nighttime Scenes

Nan Mu[1], Jinjia Guo[2]* & Jinshan Tang[3]*

[1]School of Computer Science, Sichuan Normal University, Chengdu, China

[2]Department of Chongqing University, University of Cincinnati Joint Co-op Institution, Chongqing University, Chongqing, China

[3]Department of Health Administration and Policy, George Mason University, Fairfax, Virginia, USA

Received 30 May 2022; revised 21 September 2022; accepted 10 October 2022

The detection of salient objects in nighttime scene settings is an essential research issue in computer vision. None of the known approaches can accurately anticipate salient objects in the nighttime scenes. Due to the lack of visible light, spatial visual information cannot be accurately perceived by traditional and deep network models. This paper proposed a Mountain Basin Network (MBNet) to identify salient objects for distinguishing the pixel-level saliency of low-light images. To improve the objects localizations and pixel classification performances, the proposed model incorporated a High-Low Feature Aggregation Module (HLFA) to synchronize the information from a high-level branch (named Bal-Net) and a low-level branch (called Mol-Net) to fuse the global and local context, and a Hierarchical Supervision Module (HSM) was embedded to aid in obtaining accurate salient objects, particularly the small ones. In addition, a multi-supervised integration technique was explored to optimize the structure and borders of salient objects. In the meantime, to facilitate more investigation into nighttime scenes and assessment of visual saliency models, we created a new nighttime dataset consisting of thirteen categories and a total of one thousand low-light images. Our experimental results demonstrated that the suggested MBNet model outperforms seven current state-of-the-art methods for salient object detection in nighttime scenes.

Keywords: High-low feature aggregation, Hierarchical supervision, Multi-supervised integration, Nighttime images, Salient object detection

Introduction

Salient object detection is crucial for computer vision, the goal of which is to enable a computer to simulate the Human Visual System (HVS) and quickly locate salient objects among an image and segment them from the complex background. Due to its ability to rapid positioning Regions of Interest (ROI), the use of salient object detection is widespread in the pre-processing steps of various vision tasks, e.g., semantic segmentation[1], image retrieval[2], object enhancement[3], object tracking[4], medical segmentation[5–8], etc.

Recently, deep learning-based saliency detection algorithms possess the capability to obtain more acceptable results than traditional methods. However, detection in nighttime situations remains a challenging research problem. The main reasons lie in the following three points: 1) nighttime scenes, especially in nighttime conditions, the feature recognition capabilities of which are relatively weak. Thus the rich semantic information and structural details of salient objects cannot be effectively obtained. 2) traditional deep networks typically employ only single-layer feature maps for prediction, which cannot accurately detect salient objects at different scales, especially for small objects. 3) some objects in the image have a small proportion and uneven distribution, which will cause incorrect detection. Meanwhile, the low Signal-to-Noise Ratio (SNR) results in blurred and incomplete boundaries of salient objects.

Because of this, different algorithms have been proposed to address these challenges, which mainly include two methods: 1) enhancing the image contrast and 2) providing an additional light source to the target. However, the contrast enhancement approach easily leads to the amplification of image noise, which is detrimental to detecting small targets. Also, providing an additional light source adds an extra step before object detection, which imposes an additional drain on the training and inference of the model.

Therefore, we propose a Mountain Basin Network (MBNet) for multi-scale salient object detection in nighttime images. The images are input to networks, high-level and low-level, respectively, by using the

*Authors for Correspondence
E-mails: Jinjia.Guo@cqu.edu.cn, jtang25@gmu.edu

residual block[9] to make the network learn the more prosperous location and category features. In addition, the concepts of the High-Low Feature Aggregation (HLFA) module and Hierarchical Supervision Module (HSM) are proposed in this paper. HLFA can aggregate the features extracted from high and low networks, enhancing network learning ability. HSM avoids overfitting caused by large datasets and deep network structures and can handle multi-scale variations of protruding objects with a minimal computational effort to obtain satisfactory accuracy. Furthermore, it can improve the accuracy of segmentation and promote fast convergence compared to the single-supervised approach.

Related Works

Some of the earliest models for visual saliency detection relied heavily on intuitive human perception and heuristic priors. However, the development of conventional image saliency detection algorithms was hampered by the need for extensive feature engineering and manual features with limited expression capability. These ineffective algorithms have trouble capturing the whole internal structure of a conspicuous object and tended to damage the underlying feature structure. Consequently, it is tricky for conventional approaches to generate adequate saliency detection findings on problematic nighttime images. With the improvement and maturation of deep learning techniques[10], deep neural network saliency detection algorithms of images have attained remarkable success. Liu *et al.*[11] proposed a contextual attention network at the pixel level that selects local or global contextual information to recognize salient objects. Although this method can improve the efficacy of salient object identification, it needs building a complicated network for feature extraction, which will increase computing costs. Chen *et al.*[12] employed a residual learning-based method to train the residual features to refine the detection of salient objects and presented a reverse attention mechanism to direct this top-downside-output residual learning. Li *et al.*[13] presented a novel Hierarchical Feature Fusion Network (HFFNet) directed by edge information for extracting image features from various levels of VGG. Then, high-level and low-level information which represent segmentation and edge respectively were produced by hierarchically combining these properties. Zhang *et al.*[14] established multi-path recurrent operation to improve the Progressive Attention-Guided Recurrent (PAGR) network for detecting salient objects, which is achieved by transmitting multi-path cyclic connections transfer global semantic information from the top convolutional layer to the deeper layers.

For salient object detection in nighttime scenes, two different approaches have been used to accomplish the task, i.e., one method uses traditional machine vision methods to find objects through mathematical transformations, and the second one trains the model using a deep neural network to find the object being searched for. In the former method, Mu *et al.*[15] utilized the discrete stationary wavelet transform to capture the saliency information from both frequency and spatial domains of low-contrast images. For the second method, Mu *et al.*[16] exploited a covariance-based convolutional neural network model to detect salient objects in low-contrast images. Xu *et al.*[17] presented an enhancement method to facilitate the deep neural network to recognize salient objects in low-light images. However, the traditional methods are inefficient and limited, and the extracted features are difficult to be applied to nighttime images with different luminance. Still, the depth feature-based methods cannot accurately detect small objects in the nighttime scenes, and the edges of the objects cannot be accurately captured.

In recent years, based on the further study of nighttime scenarios, Lore *et al.*[18] proposed an LLNet for using a variation of the stacked-sparse denoising autoencoder with whitening and denoising of dark images, low contrast image saliency detection based on deep learning networks was performed, thereby inspiring the use of end-to-end networks in low-contrast image enhancement. Zhu *et al.*[19] suggested an Edge-Enhanced Multi-Exposure Fusion Network (EEMEFN), in which the edge improvement module was used to improve the initial image and the multi-exposure fusion module was used to address the color deviation issues. Xu *et al.*[20] found that different objects had variable contrast at various frequency levels, and the low-frequency layer could make noise identification much easier than the high-frequency layer, and a frequency-based breakdown and augmentation network had been proposed. The network, under the condition of low-frequency noise layer of the image content layer, can simultaneously infer high-frequency detail. Li *et al.*[21] recently introduced a weak light image enhancement network that employed a residual block and a recursive layer to extract local features, then gradually injected the global feature map of dual attention into each stage. Although the deep learning-based saliency model has been able to obtain excellent results in various complicated scenarios, it still has room for improvement. However, because of circumstances such as low contrast, SNR, etc., the salient objects referenced by these models always lose some details of the structure and boundary portions, and they may also identify background content. This study primarily adopted the high-low feature aggregation module and hierarchical supervision module to capture richer semantic information and structural details of multi-scale salient objects and leveraged a multi-supervised integration strategy to enhance the shapes and boundaries of salient objects in nighttime conditions. Consequently, the salient objects of various dimensions, particularly the little objects, are highlighted more precisely.

Contributions of this Work

1) A novel network MBNet is proposed for nighttime image saliency detection, exploiting a high-low feature aggregation module to learn richer semantic information and structural details. The detection of small salient objects is facilitated by fusing the prediction results from different perceptual domains.

2) A multi-supervised model strategy is leveraged to balance the discrepancies caused by category inequality. It can solve the uneven distribution of background and foreground, focus on salient objects for accurate segmentation, and optimize the boundaries of salient objects under nighttime conditions. In addition, the multi-supervised strategy improved the model convergence speed and segmentation accuracy compared with the single-supervised approach.

3) We built a publicly available salient object detection dataset under low illumination: the *nighttime image segmentation* (NISeg) dataset, for studying and evaluating object segmentation and saliency detection models under low illumination. We were able to successfully prove the higher performance of the suggested model on this new dataset. Specifically, our network achieved an AUC of 86% on the NISeg dataset, which is significantly greater than the performance of other state-of-the-art networks, proving the network's superiority.

Proposed Method

Here, the MBNet, a newly salient object detection network will be a detailed description, starting from an overview of our deep neural network, which includes Bal-Net, the low-level network containing richer semantic information and feature recognition capabilities. Mol-Net, the higher-level network for determining the location of salient targets in the image and optimizing image edges, as well as high-low feature aggregation module for feature fusion and hierarchical supervision module.

Overview

Inspired by the encoder-decoder structures of UNet[22] and SegNet[23], we designed MBNet with high and low network branches for feature extraction to fully learn the feature information in the nighttime images. The nighttime inputs of size 128×128 are used as the input and then it enters the low-level network (Mol-Net) and the high-level network (Bal-Net) simultaneously for feature extraction as shown in Fig. 1. In the Mol-Net network, the input image is up-sampled first and then it will be recovered by maximum pooling. Long skip connections are used between different feature maps of the encoder and decoder layers. In the Bal-Net network, the input image is first convolved, pooled by residual blocks, and up-sampled to the original size (128×128) then. The HLFA module is applied after each image extraction of features. The feature maps are aggregated together after convolutions and then sent back to the high-level and low-level networks respectively as inputs for further learning. The HLFA module's internal structure is shown in Fig. 2. Before supervising the image output, we also used four hierarchical supervised modules to evaluate the training results of the whole model and update the parameters in advance. The structure of the HSM is shown in Fig. 3. After combining the feature maps from the high-level and low-level networks, the salient object detection map is finally produced.

Bal-Net

Inspired by UNet[22], we designed a model with an encoder-decoder for feature extraction of image data in nighttime. In the encoder, the input image is extracted using the residual network module, where the convolution kernel size is 1. The same padding approach is employed to ensure that the image size after convolution is unaltered. The size has been reduced by half, but the number of channels has not

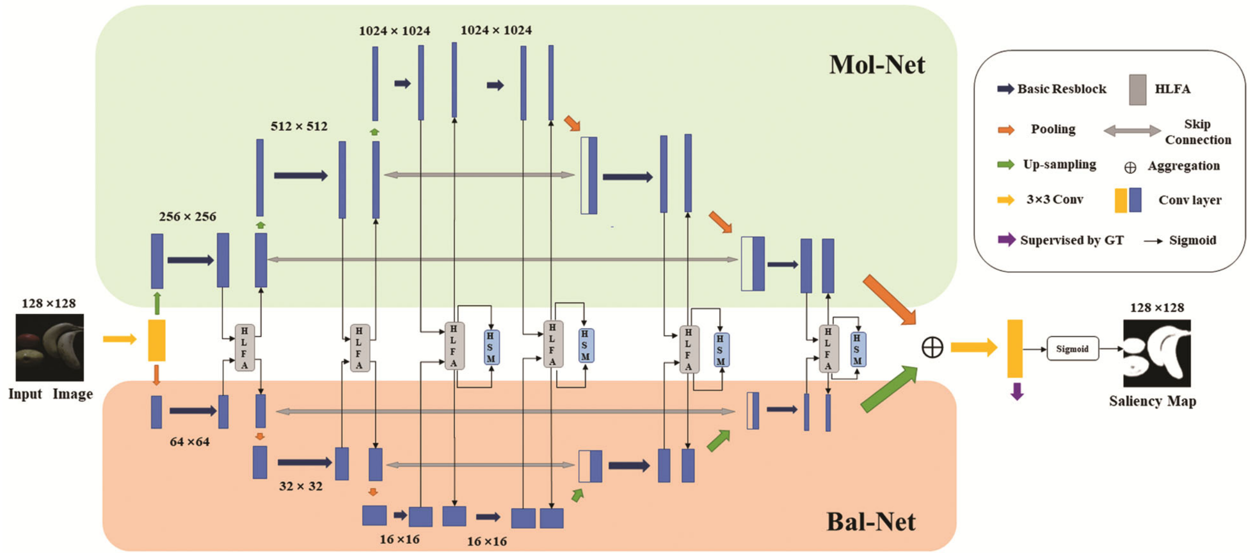

Fig. 1 — The architecture of MBNet network containing the Mol-Net and Bal-Net components that supervised by the composite loss

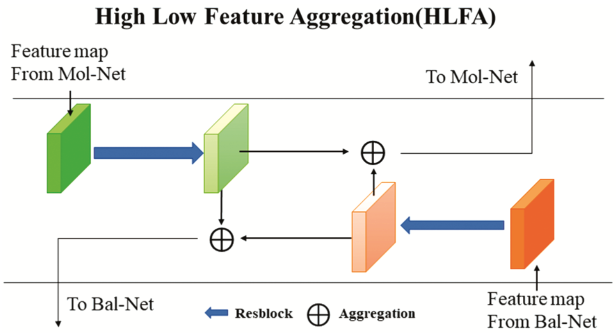

Fig. 2 — The HLFA module synergizes the information of Mol-Net and Bal-Net to fuse the global and local contexts

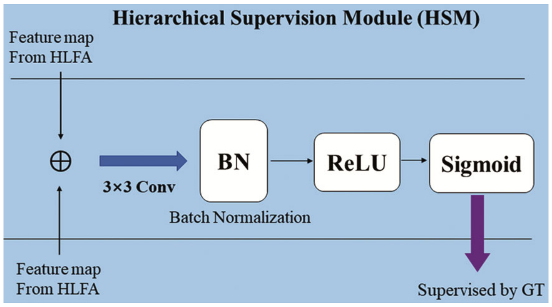

Fig. 3 — HSM is embedded on each layer of the decoder to assist in obtaining a clear and accurate salient object

changed. After down-sampling three times, the decoder was used to recover the image to its origin, selecting bilinear interpolation for up-sampling, and up-sampling three times to get a feature map of similar size as the original image. To deepen the network and generate a more comprehensive feature map of the image in a nighttime environment, a basic residual neural network module is added at the encoder-decoder connection. Through the convolution procedure, the receptive field steadily expands, and the semantic information it contains grows richer, which is essential for learning the shape characteristics and speckles of salient objects in nighttime conditions. To further fuse the global and local features among the image in nighttime conditions and reduce signal-to-noise ratio interference, we employed a skip-connection technique like UNet to aggregate the individual feature maps. However, the model's concentration capacity deteriorates, making it less capable of learning features like blurred edges and noisy textures. Consequently, we introduced the low-level network Mol-Net to resolve this issue.

Mol-Net

To handle the object localization problem in nighttime images, a new branch is added to the original network: the low-level network Mol-Net, whose structure is quite similar to Bal-Net. In Mol-Net, after the image has been exposed to feature extraction by the residual block, to reduce the feature map's receptive field, we use bilinear interpolation. so that it can pay greater attention to the object's edge information. The image size is increased, and the number of channels is decreased after up sampling. Likewise, a residual block is added between the decoder layers and the encoder layers. After decoding, the generated feature map is recovered to its original image size, and the decoder uses a skip connection to maintain both shallow and deep features. By including Mol-Net, the network does not lose the edge information properties of the object when the image's luminance and contrast are low in a low-light background, even as the network is gradually deepened. Before outputting the final probability map, we combine the feature maps acquired by Bal-Net and Mol-Net and tweak the number of channels using convolution to produce the final salient target prediction binary map. At this stage, the image possesses both rich and deep semantic knowledge about the salient objects and the capacity to distinguish fuzzy edges.

High-Low Feature Aggregation

The construction of the high-low feature aggregation module is illustrated in Fig. 2. In the task of detecting salient objects, we must not only differentiate the categories of targets but also segment them at the pixel level. However, the pooling layer in the Bal-Net network architecture causes spatial domain information to be compressed, resulting in less accurate segmentation results. Consequently, using HLFA, we fuse the feature map with higher feature resolution in Mol-Net and the semantic information obtained in Bal-Net so that the feature map can obtain more target features in the subsequent learning process without losing the edge position information of the image, thereby obtaining a feature map with enhanced discriminative ability. In addition, the HLFA module may effectively eliminate gradient disappearance and network degradation issues, as well as facilitate training convergence. During backpropagation, the gradient flow simultaneously traverses both branches and the parameters are updated.[24]

Hierarchical Supervision Module

In the Fig. 3, the structure of the hierarchical supervision module is explained. Inspired by the Feature Pyramid Network (FPN)[25], we added the HSM module in the network for improving the robustness of detection at various image scales. For hierarchical supervised learning, the prediction results are based not only on the output of one layer but also on the output of numerous layers. Simultaneously, HSM can change the parameters of the underlying feature maps in a timely manner, allowing the network to attain greater accuracy with tiny datasets or a limited number of iterations.

Multi-Supervised Integration

The concept of aggregated loss function was introduced by Jadon *et al.*[26], inspired by the fact that we also defined our loss function as a composite loss function, including a fusion of distribution-based loss (focal loss), region-based loss (dice soft loss), and boundary-based loss (boundary loss).[27] To achieve high-quality regional segmentation and clearly defined borders, the composite loss can be calculated as:

$$L = \sum_{i=1}^{I} \alpha_i \left(l_{focal}^{(i)} + l_{dice}^{(i)} + l_{boundary}^{(i)} \right), \qquad \ldots(1)$$

where, $l_{focal}^{(i)}$ is the focal loss, $l_{dice}^{(i)}$ is the dice soft loss, $l_{boundary}^{(i)}$ is the boundary loss, I represents the quantities of outputs, α_i represents the weight. After performing the HLFA module and SHM, four initial saliency maps are produced and the final outputs are generated by 1×1 convolution of our training process, i.e., $I = 5$

Distribution-based Loss

In general, traditional losses transform the similarity of pixels into a possibility and aim to reduce the difference between prediction and reality. Therefore, these losses learn every pixel in the image equally. But in most cases, the classes in the dataset are unevenly distributed (e.g., under nighttime conditions, some salient objects occupy only a small part of pixels, and most of the pixels are in the background region), thus the trained model tends to predict the pixels as the class with a large number of pixels. For balancing the gap caused by uneven categories, different weights are given to positive samples and negative samples. In addition, contemplating the different costs of learning samples caused by different image features (e.g., brightness, contrast, and texture details) in the nighttime dataset, the pixels are divided according to the difficulty of learning that can more specifically improve the detection accuracy and promote faster convergence. Thus, the proposed distribution-based loss l_{focal} is defined as blow, in which the loss of the positive sample is l_1, the loss of the negative samples is l_2 and the final loss is the summation.

$$\begin{aligned} l_{focal} &= -l_1 - l_2, \\ l_1 &= \alpha(1-y_{pred})^{\gamma} \times y_{true} \log(y_{pred}), \\ l_2 &= (1-\alpha) y_{pred}{}^{\gamma} \times (1-y_{true}) \log(1-y_{pred}), \end{aligned} \qquad \ldots(2)$$

where, γ is usually set to 2, and $\alpha = N_{neg} / N_{pos}$, y_{pred} is our predicted saliency map, y_{true} is the ground truth.

Region-based Loss

Region-based loss aims to minimize the mismatch region between ground truth y_{true} and the predicted saliency map y_{pred}, which mainly calculates the similarity between two samples from the perspective of the whole via:

$$l_{dice} = 1 - \frac{2\sum_{P} y_{true} y_{pred}}{\sum_{P} ({y_{true}}^2 + {y_{pred}}^2)}, \quad \ldots (3)$$

where, P represents all pixels in the binary ground truth or predicted probability map. Overall, l_{dice} is suitable for dealing with nighttime images, where the foregrounds and backgrounds are unevenly distributed. Furthermore, l_{dice} focuses more on capturing salient regions in the training process, which is undoubtedly appropriate for our research.

Boundary-based Loss

Due to the intrinsic properties of low brightness and SNR of nighttime images, the contours of salient objects are not clear. To handle this challenge, inspired by Zhu *et al.*[28], accurate boundary segmentation can be achieved by minimizing the boundary distance of S_θ and G, i.e., $l_{boundary}$ With the progress of training, the loss continues to decline, and a clearer and more accurate boundary can be obtained in the later stage. $l_{boundary}$ is calculated as follows:

$$\phi_G = \begin{cases} -D_G(q) & q \in G \\ D_G(q) & otherwise \end{cases}, \qquad \ldots (4)$$
$$l_{boundary} = \int_{\Omega} \phi_G(q) s_\theta(q) dq,$$

where, Ω is image set, S_θ is softmax predictions, G is ground truth, $-D_G(q)$ is distance map corresponding to ∂G.

Above all, the combination of l_{dice} and $l_{boundary}$ will be very effective. Since the former uses regional information from a global perspective, the latter pays attention to boundary knowledge from local information. Similarly, l_{focal} is essentially an improved BCELoss, it focuses on the distribution of nighttime image datasets, through dynamically scales BCELoss to prevent submerging simple negative samples. In general, the composite loss combines the above three kinds of loss and has an excellent performance in nighttime images.

Results and Discussion

Experimental Datasets

MSRA[29], DUT-OMRON[30], PASCAL-S[31], HKU-IS[32], and DUTS[33] are currently the most popular datasets for salient object detection. These datasets have a different number of images, image resolutions, image sceneries, number of salient items in images, and detection difficulty. In addition, pixel-level manual calibration benchmarks of these datasets were provided as performance comparison references, which were ideal for evaluating the performance of saliency models. The majority of traditional image datasets consisted of situations with favorable environmental conditions, such as visible light, and had very few nighttime photographs. However, in everyday life, backlight, non-uniform illumination, and low light bother individuals owing to lighting or technical restrictions. Existing saliency detection methods had difficulty in distinguishing salient items in nighttime scene settings with precision. In addition, the lack of a comprehensive dataset for salient object detection in nighttime scene settings hindered the resolution of this issue. The proposed publicly accessible nighttime scene datasets are NI-A[34], LOL[35], and Exdark[36]. Specifically, the NI-A datasets featured nighttime images annotated at the pixel level, although the amount of data was minimal, and the scene was reasonably straightforward. The LOL dataset included 500 image pairs that corresponded to low-light and normal-light images, but these images do not accurately depict a low-light situation. Exdark dataset images were captured in actual low-light environments, although neither LOL nor Exdark datasets included pixel-level annotated images.

In light of this, we created the NISeg dataset, which contains 1,000 low-light images captured in high-quality nighttime conditions. The majority of the nighttime images were downloaded from Baidu search and Google search, and some images were obtained by equipment shooting, and the collection time was during the period before the sky darkened into night, during which the visible light gradually became weaker, the color information of objects gradually degraded, and the SNR of the scene (containing low light, underexposure, and noise) became lower and lower. From the acquired images, a total of one thousand nighttime images for thirteen item categories were picked and the carefully annotated baseline ground facts were provided. The selection of these images was based on the following criteria: 1) All images have low light; 2) They contain multi-scale items, such as small and huge objects, whose area ratio to the image area is less than 0.1 or larger than 0.7; and 3) several salient objects are unconnected.

Experimental Details

In this work, 1000 nighttime images from the NISeg dataset were used for trials. To conduct an exhaustive evaluation, we employ six performance metrics: 1) the curve of precision-recall (PR) measured by precision and recall; 2) the curve of the receiver operating characteristic (ROC) generated from false positive rate and true positive rate; 3) the curve of F-measure; 4) the area under the curve (AUC) of ROC; 5) the Mean Absolute Error (MAE) obtained by comparing GTs and saliency maps; 6) the Overlapping Ratio (OR) between GTs and saliency maps.

We developed the recommended method utilizing the PyTorch with a graphics card RTX3090 and CPU of i9-10900k including 128G RAM. The MBnet model was trained by NISeg dataset. The training set has 800 images, test set contains 200 images, a total of 100 epochs that took approximately 10 hours to complete. All models were tested in the same environment and with the same test set.

Comparison with the State-of-the-art Models

During this section, the proposed network is compared with four advanced salient object recognition models (BAS[37], F^{3}Net[38], MPI[39], and MEUN[40]) and three image segmentation models (UNet[22], AUTO[28], and KIU[41]).

The results of the visual comparison are demonstrated in Fig. 4. It is clear that the majority of techniques produce the desired effects, but are inferior to the MBNet we presented, and that some competitive models cannot identify all prominent items in nighttime environments (e.g., UNet, AUTO, and KIU). The proposed MBNet provides good detection results for salient objects of various sizes, especially for small objects, and the obtained salient objects have defined and complete borders. This illustrates that our model is highly capable and dependable for saliency detection in complex nighttime environments.

Comparisons between the proposed MBNet and seven models on a quantitative level at the forefront of the field can be found in Table 1 and Fig. 5(a–c). On the ROC curve, PR curve, AUC score, and OR score, the proposed model achieves excellent results; on the F-measure curve, it ranks third; and on

Table 1 — Comparing the quantitative performance of different deep learning models using AUC, MAE, and OR metrics

	AUC↑	MAE↓	OR↑
AUTO	0.6659	0.1733	0.3709
BAS	0.8451	0.0484	0.8031
F^{3}Net	0.8412	0.0564	0.7914
KIU	0.7068	0.1717	0.4205
MEUN	0.8408	0.0508	0.7964
MPI	0.8494	0.0527	0.7964
UNet	0.7063	0.1445	0.4603
MBNet	0.8607	0.0494	0.8083

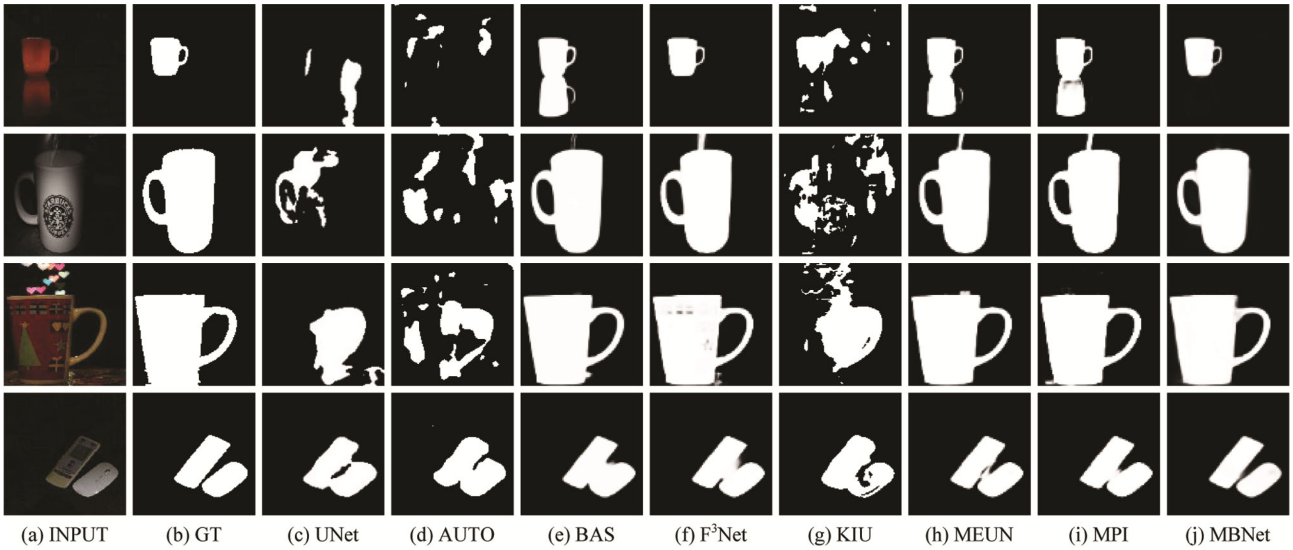

Fig. 4 — Visual comparisons of saliency maps generated by different models on the proposed NISeg dataset

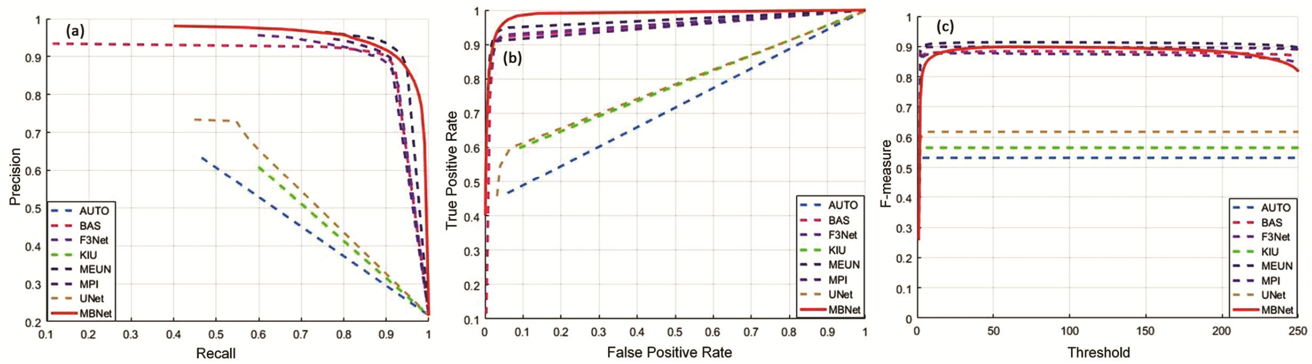

Fig. 5 — On the NISeg dataset, performance comparisons between the existing state-of-the-art saliency models and the proposed model (which is signed by the red solid line): (a) precision-recall curve, (b) ROC curve, (c) F-measure curve

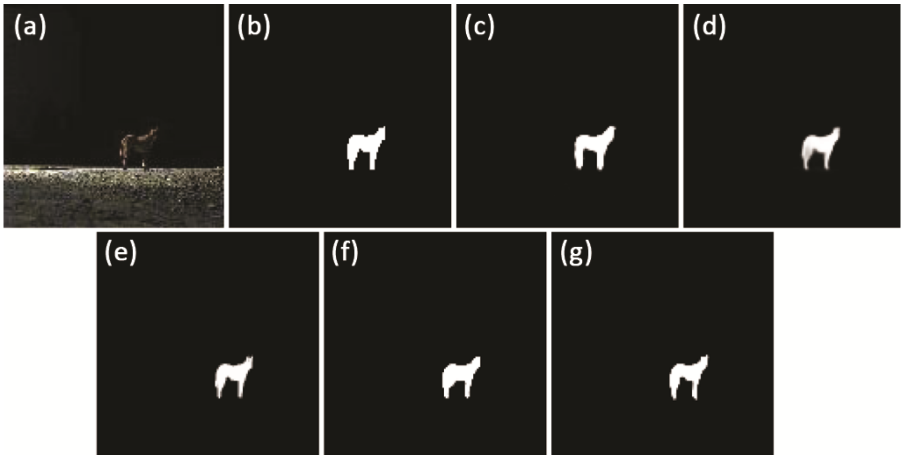

Fig. 6 — Visualization of five ablations of the proposed MBNet model: (a) Input, (b) GT, (c) Ablation 1, (d) Ablation 2, (e) Ablation 3, (f) Ablation 4, (g) Ablation 5

the MAE score, the differences between our model and BAS model's best results was only 0.01. The UNet, AUTO, and KIU models should not be employed for nighttime detection, as is obvious from the data. These objective performance comparisons demonstrate that MBNet detects notable objects in nighttime images effectively. The proposed model especially employed a multi-supervised integration method that may successfully improve object-to-background contrast. HLFA and HSM also fine-tune the boundaries and structures of salient areas.

The ↑ indicates a higher value and better performance; The ↓ has the opposite meaning

Ablation Study

To illustrate the contributions of proposed MBNet, ablation experiments on the NISeg dataset by testing five variants of our model are performed, including: 1) Ablation 1, which only uses the basic encoder-decoder architecture of Mol-Net and Bal-Net; 2) Ablation 2, which uses the residual connection for HLFA; 3) Ablation 3, which adds the HSM at the last layer. 4) Ablation 4, which utilizes a multi-supervised integration strategy for all the decoder layers; 5) Ablation 5 combines the whole configurations. The subjective performance comparisons of saliency maps obtained from the five ablations are provided in Fig. 6.

It can be observed that Ablation 1 can roughly capture the salient object in nighttime scenes, but the detected object lacks details. By adding the HLFA module, Ablation 2 can locate more accurate salient regions. In addition, the HSM is exploited to detect the details, e.g., the ears of the dog (Fig. 6(e)). Furthermore, with the multi-supervised integration strategy, the contour of the object is clearer (Fig. 6(f)). After combining all these components, the detected salient objects are closer to GT, with accurate foregrounds and clear contours (Fig. 6(g)). In general, these proposed ablations are beneficial for detecting salient objects in nighttime scenes.

Conclusions

We proposed an MBNet for detecting salient objects in nighttime scene settings and created a new nighttime dataset for object detection that is open to the public in this research study. In our newly proposed network, we designed an HLFA module to fuse global and local contexts and an embedded HSM module into each layer of the decoder to gradually obtain rich multi-scale semantic information and structural information at different levels, thereby enhancing the precision of salient object location determination and structure refinement. A multi-supervised integration technique was also employed to maximize the salient object's form and border. This provides a boost to the detection of small objects. The suggested approach can efficiently extract additional discriminative features from nighttime images, enabling precise recognition of salient objects. On the proposed datasets, extensive trials show that this model surpassed the most advanced saliency detection techniques and it is valuable for more image processing tasks like object tracking in nighttime surveillance.

Acknowledgment

Nan Mu was supported by the National Natural Science Foundation of China (62006165).

References

1 Zhang Z, Cui Z, Xu C, Yan Y, Sebe N & Yang J, Pattern-affinitive propagation across depth, surface normal and semantic segmentation, in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* (IEEE) 2019, 4106–4115.

2 Tang J & Acton S, An image retrieval algorithm using multiple query images, in *7th Int Symp Signal Process Appl* (IEEE) 2003, 193–196.

3 Tang J, Sun Q & Agyepong K, An image enhancement algorithm based on a contrast measure in the wavelet domain for screening mammograms, in *IEEE Int Conf Imag Process* (IEEE) 2007*,* 29–32.

4 Wang Q, Zhang L & Bertinetto L, Fast online object tracking and segmentation: A unifying approach, in *IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2018, 1–13.

5 Mu N, Wang H, Zhang Y, Jiang J & Tang J, Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images, *Pattern Recognit*, **120** (2021) 1–12.

6 Liu X, Yuan Q, Wang B, Tang X, Tang J & Shen D, Weakly supervised segmentation of COVID-19 infection with scribble annotation on CT images, *Pattern Recognit,* **122** (2022) 1–15.

7 He J, Zan Q, and Zhang K, Yu P & Tang J, An evolvable adversarial network with gradient penalty for COVID-19 infection segmentation, *Appl Soft Comput*, **133** (2021) 1–10.

8 Zhao C, Xu Y, He Z, Tang J, Zhang Y, Han J, Shi Y & Zhou W, Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images, *Pattern Recognit*, **119** (2021) 1–14.

9 He K, Zhang X & Ren S, Deep residual learning for image recognition, in *IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2019, 30–42.

10 Liu X, Yu A, Wei X, Pan Z & Tang J, Multimodal MR image synthesis using gradient prior and adversarial learning, *J Sel Top Signal Process*, **14** (2020) 1176–1188.

11 Liu N, Han J & Yang M, PiCANet: Learning pixel-wise contextual attention for saliency detection, in *IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2018, 3089–3098.

12 Chen S, Tan X, Wang B & Hu X, Reverse attention for salient object detection, in *Euro Conf Comput Vis* (Computer Vision Foundation) 2018, 234–250.

13 Li X, Song D & Wang B, Hierarchical feature fusion network for salient object detection, *IEEE Trans Imag Process*, **29** (2020) 9165–9175.

14 Zhang X, Wang T & Qi J, Progressive attention guided recurrent network for salient object detection, in *IEEE Conf Comput Pattern Recognit* (IEEE) 2018, 18–22.

15 Mu N, Xu X, Zhang X & Lin X, Discrete stationary wavelet transform based saliency information fusion from frequency and spatial domain in low contrast images, *Pattern Recognit Lett*, **115** (2018) 84–91.

16 Mu N, Xu X, Zhang X & Zhang H, Salient object detection using a covariance-based CNN model for low-contrast images, *Neural Comput Appl*, **29(8)** (2018) 181–192.

17 Xu X, Wang S, Wang Z, Zhang X & Hu R, Exploring image enhancement for salient object detection in low light images. *ACM Trans Multimed Comput Commun Appl*, **17(1s)** (2021) 1–19.

18 Lore K G, Akintayo A & Sarkar S, LLNet: A deep autoencoder approach to natural low-light image enhancement, in *IEEE Conf Comput Vis* (IEEE) 2017, 650–662.

19 Zhu M, Pan P, Chen W & Yang Y, EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network, in *Proc* AAAI *Conf Artif Intell*, **34(7)** (2020), 106–113.

20 Xu K, Yang X, Yin B & Lau R W H, Learning to restore low-light images via decomposition-and-enhancement, in *IEEE Conf Compu Vis Pattern Recognit* (IEEE) 2020, 2281–2290.

21 Li J, Feng X & Hua Z, Low-light image enhancement via progressive-recursive network, *IEEE Trans Circuits Syst Video Technol*, **31(11)** (2021) 4227–4240.

22 Ronneberger O, Fischer P & Brox T, U-net: Convolutional networks for biomedical image segmentation, in *Med Image Comput Comput Assist Interv* (Springer, Cham) 2015, 234–241.

23 Badrinarayanan V, Kendal A & Cipolla R, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans Pattern Anal Mach Intell*, **39(12)** (2017) 2481–2495.

24 Tang J, Millington S, Acton S T, Crandall J & Hurwitz S, Ankle cartilage surface segmentation using directional gradient vector flow snakes, in *Int Conf Imag Process* (IEEE) 2004, 2745–2748.

25 Lin T, Dollar P, Girshick R, He K, Hariharan B & Belongie S, Feature pyramid networks for object detection,

in *IEEE Conf Comput Vis and Pattern Recognit* (IEEE) 2017, 2117–2125.

26 Jadon S, A survey of loss functions for semantic segmentation, *IEEE Conf Comput Intell Bioinformat Computat Biol* (CIBCB) (IEEE) 2020, 1–6

27 Lin T, Goyal P, Girshick R, He K & Dollar P, Focal loss for dense object detection, in *IEEE Int Conf Comput Vis* (IEEE) 2017, 2999–3007.

28 Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Du N, Fan W & Xie X, AnatomyNet: deep learning for fast and fully Automated whole-volume segmentation of head and neck anatomy, *arXiv preprint 1808.05238*, (2018) 1–13.

29 Lin T, Yuan Z & Sun J, Learning to detect a salient object, *IEEE Trans Pattern Anal Mach Intell*, **33(2)** (2011) 353–367.

30 Yang C, Zhang L & Lu H, Saliency detection via graph-based manifold ranking, in *IEEE Conf Compu Vis Pattern Recognit* (IEEE) 2013, 3166–3173.

31 Li Y, Hou X & Koch C, The secrets of salient object segmentation, in *IEEE Conf Comput Vis Pattern Recognit* (IEEE) 2014, 280–287.

32 Li G, Lu H & Wang Y, Visual saliency based on multiscale deep features, in *IEEE Conf Comput Vis* (2015) 5455–5463.

33 Wang L, Lu H & Wang Y, Learning to detect salient objects with image-level supervision, in *IEEE Conf Comput Vis Pattern Recognit* (2017) 136–145.

34 Mu N, Xu X & Zhang X, Salient object detection in low contrast images via global convolution and boundary refinement, in *IEEE Conf Comput Vis Pattern Recognit Works* (2019) 1–9.

35 Wei C, Wang W & Yang W, Deep retinex decomposition for low-light enhancement, *arXiv preprint 1808.04560* (2019) 1–12.

36 Loh Y & Chan C, Deep residual learning for image recognition, in *IEEE Conf Comput Vis Pattern Anal Mach Intell* (2016) 770–778.

37 Qin X, Zhang Z, Huang C, Gao C, Dehghan M & Jagersand M, Basnet: Boundary-aware salient object detection, in *IEEE Conf Compu Vis Patt Recog* (2019) 7479–7489.

38 Jun W, Wang S & Huang Q, F^{3}Net: fusion, feedback, and focus for salient object detection, in *Proc AAAI Conf Artifi Intell*, **34(7)** (2020) 12321–12328.

39 Sun H, Jun C & Liu N, MPI: Multi-receptive and parallel integration for salient object detection, in *IEEE Conf on Comput Vis Pattern Recognit* (2021) 1–10.

40 Sun H, Bain Y & Liu N, Multi-scale edge-based U-shape network for salient object detection, in *IEEE Conf Comput Vis Pattern Recognit* (2021) 1–14.

41 Valanarasu J, Sindagi V, Hacihaliloglu I & Patel V, Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations, in *Med Image Comput Comput Assist Interv* (Springer, Cham) 2020, 363–373.