ITC 3/52
Information Technology and Control
Vol. 52 / No. 3 / 2023
pp. 581-593
DOI 10.5755/j01.itc.52.3.32258

Saliency Detection Algorithm for Foggy Images Based on Deep Learning

Received 2022/09/11 | Accepted after revision 2022/11/14

HOW TO CITE: Zhang, L., Ji, Z., Xu, R., Zhang, D. (2023). Saliency Detection Algorithm for Foggy Images Based on Deep Learning. *Information Technology and Control*, 52(3), 581-593. https://doi.org/10.5755/j01.itc.52.3.32258

Saliency Detection Algorithm for Foggy Images Based on Deep Learning

Leihong Zhang
College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai, China; e-mail: lhzhang@usst.edu.cn

Zhaoyuan Ji
College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai, China; phone: +86 17665239280; e-mail: jzycr316@163.com

Runchu Xu
School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China; e-mail: xrc1231@163.com

Dawei Zhang
School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China; e-mail: usstoe@vip.163.com

Corresponding author: jzycr316@163.com, ORCID 0000-0002-2778-7451

The detection of salient objects in foggy scenes is an important research component in many practical applications such as action recognition, target tracking and pedestrian re-identification. To facilitate saliency detection in foggy scenes, this paper explores two issues. The construction of dataset for foggy weather conditions and implementation scheme for foggy weather saliency detection. Firstly, a foggy sky image synthesis method is designed based on the atmospheric scattering model, and a saliency detection dataset applicable to foggy sky is constructed. Secondly, we compare the current classification networks and adopt resnet50, which has the highest classification accuracy, as the backbone network of the classification module, and classify the foggy sky images into three levels, namely fogless, light fog and dense fog, according to different concentrations. Then, Residual Refinement Network (R^2Net) was selected to train and test the classified images. Horizontal and vertical flipping and image cropping were used to enhance the training set to relieve over-fitting. The accuracy of the network model was improved by using Adam as the optimizer. Experimental results show that for the detection of fogless images, our method is almost on par with state-of-the-art, and performs well for both light and dense fog images. Our method has good adaptability, accuracy and robustness.

KEYWORDS: Foggy images, Saliency detection, Image classification, Deep learning.

1. Introduction

Saliency detection, as one of the popular research directions in the field of computer vision, has wide range of applications in the fields of video surveillance [24], image thumbnail [23], and semantic segmentation [27]. The saliency detection under foggy conditions, as one of its branches, has also attracted the attention of related researchers. However, early saliency models such as Frequency Tuned (FT) [1], Histogram-based Contrast (HC) [3], Itti (IT) [8], Luminance Contrast (LC) [28] mainly rely on features such as color, contrast and contour of the image. With the development of deep learning theory, more and more network models have been proposed. The deep contrast network proposed by Li et al. [11] solves the problem of blurred saliency map in saliency detection. The Amulet network proposed by Zhang et al. [29] utilizes convolutional features from multiple layers as saliency cues for salient object detection.

Most of these studies are aimed at targets in the natural environment, which are characterized by contrasting colors and clear outlines. Although many saliency detection models can achieve good results on existing datasets, they often fail to achieve ideal results when they are actually applied to environments under foggy conditions. Foggy scene is an environment with uncertain factors such as smoke suspended solids and automobile exhaust. These factors make the captured images subject to blur, occlusion, abnormal lighting, etc., resulting in loss of image detail, low contrast and color distortion. Accurate saliency detection plays important role in related applications.

The existing saliency detection networks are not aimed at foggy images, and are not suitable for saliency detection under foggy conditions. There is also lack of foggy detection datasets. Therefore, this paper first simulates the distorted images affected by fog through the atmospheric scattering model, and generates synthetic fog images based on the DUTS [21] dataset. In order to not lose the detection quality of fogless images and have good detection ability of foggy images with different concentrations, the detection network based on R^2Net [6] is adopted to train separately according to the classification results. and optimize the parameters corresponding to the concentration.

2. Related Works

As of now, human-annotated datasets in real foggy conditions are very rare, and images for saliency detection are even rarer. Therefore, the method for synthesizing hazy images by atmospheric scattering model in this paper uses the generated dataset for training and testing the model.

The atmospheric scattering model of Koschmieder can simulate the effect of fog better. We use the atmospheric scattering model to generate a simulated fog image. The fog image $I(\mathrm{x})$ can be expressed as

$$I(\mathrm{x}) = J(\mathrm{x})t(\mathrm{x}) + A(1 - t(\mathrm{x})), \quad (1)$$

where $J(\mathrm{x})$ is the haze-free image and A is the atmospheric light value. The model usually assumes that the atmospheric light value is globally constant. $t(x)$ is the transmittance, which represents the amount of transmittance of the scene to the camera. In the case of homogeneous medium, the transmittance depends on the distance $l(x)$ from the scene to the camera:

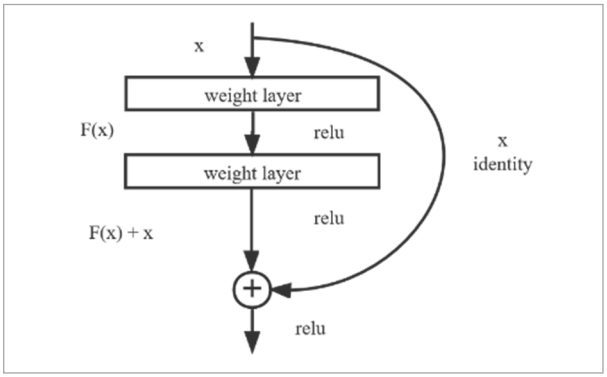
$$t(\mathrm{x}) = \exp(-\beta l(\mathrm{x})) \quad (2)$$

among them, β is called the attenuation coefficient, which can adjust the density of fog in the generated image. A larger β means that the fog density of the image is larger; conversely, the density is smaller. For haze-free image, the transmittance is first calculated using Equation (2), and the atmospheric light value A is calculated by substituting it into Equation (1); then mist and dense images are generated according to Equation (1).

In recent years, convolutional neural networks have become one of the research hotspots in many disciplines. The use of convolutional neural networks to process and analyze data has become popular trend. Classical network models such as AlexNet [10], VGG-Net [19] , GoogleNet [20] and ResNet [15] have been proposed one after another. In order to solve the foggy image classification task with small samples, this paper analyzes the current popular classification network models, and determines the basis of the foggy image classification module from the perspectives of network structure, floating-point operations and parameters. Finally, this paper adopts ResNet50 [7] as the basic model to achieve three-classification.

The residual network proposed by ResNet improves the structure of the convolutional neural network, so that it can maintain its feature expression ability while increasing the depth of the network, and effectively solve the problem of gradient disappearance or gradient explosion caused by deepening the number of layers. The introduction of residual module is a crucial part in the development process of convolutional neural network. The structure of this module is shown in Figure 1.

Figure 1
Residual structure of ResNet



The residual structure constitutes two mapping paths, identity mapping and residual mapping, in the form of cross-layer links. By adding the x-identity map in the process of common module connection, the network can effectively control the network layer parameters and the computational complexity while alleviating the problem of gradient disappearance. The residual structural unit can be expressed as:

$$x_{j+1} = x_j + F(x_j, W_j). \tag{3}$$

In Equation (3), x_j and x_{j+1} respectively represent the input and output information of the network in this layer; W_j represents the parameters to be learned in this layer. Performing recursive operation on Equation (3), the feature representation of any deep unit J can be obtained:

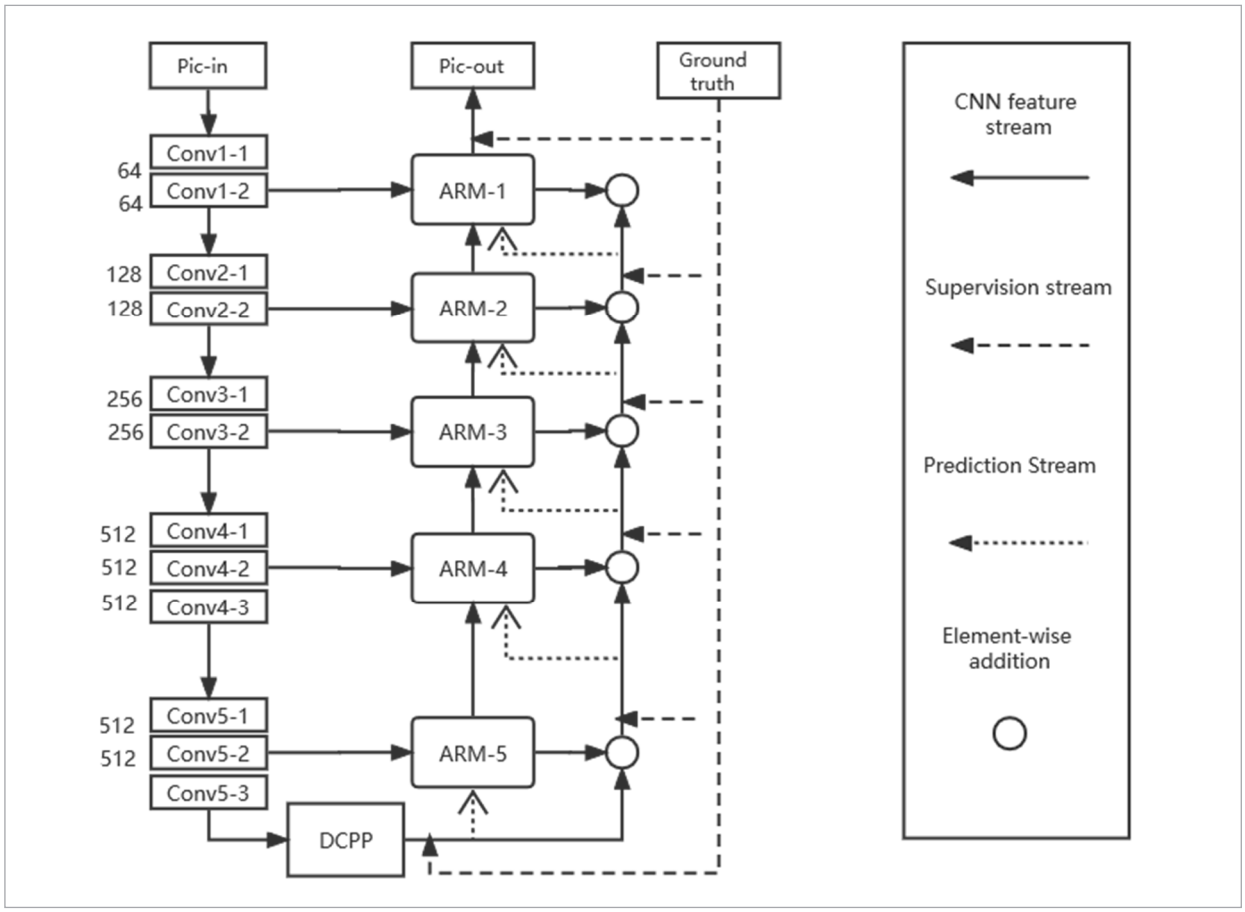
$$x_J = x_j + \sum_{i=j}^{J-1} F(x_i, W_i). \tag{4}$$

R^2Net [6] proposes a new residual learning strategy, which is different from previous models based on multi-scale, but gradually generates prediction maps of each scale. This strategy arranges the prediction maps at each scale from small to large until it matches the best ground-truth map. R^2Net employs Dilated Convolutional Pyramid Pooling (DCPP) module to generate coarse predictions based on global contextual information, which can locate the general location of target objects. The DCPP module consists of dilated convolutions at different rates to capture local and global information. This module has relatively few parameters compared to using fully connected layers. Then, by introducing new Attention Residual Modules (ARMs), the matching process of coarse predictions and ground-truth (GT) maps is guided. ARMs focus on edge details while guiding the refinement process, making the saliency map more discriminative. The R^2Net network structure is shown in Figure 2.

An important topic in deep learning is the generalisation ability of the model. The problem of over-fitting is often encountered in applications, and the correct use of regularization techniques can improve or reduce the problem of over-fitting. Zheng et al. [32] proposed a two-stage training method to improve the generalization ability of the network. In the pre-training process, the network model is trained to extract the image representation for anomaly detection. In the implicit regularization training stage, the network is retrained to regularize the feature boundary to converge based on the anomaly detection results. This approach effectively maintains a low over-fitting. Jin et al. [9] proposed computer-aided facial diagnosis for various diseases using deep transfer learning for face recognition. The overall top-1 accuracy can reach more than 90%, outperforming traditional machine learning methods and clinicians in experiments. Zheng et al. [33] proposed a spectrum interference-based two-level data augmentation method for automatic modulation classification. This is the first time that frequency domain information is considered to augment radio signals to help modulation classification.

When deep neural networks process larger scale data, the excessive computation affects the learning and inference speed of the model and cannot meet the demand in practical applications. Therefore, improving the computational speed has important application value. A new faster Mean-shift algorithm is proposed by Zhao et al. [30] By introducing a novel online seed

Figure 2
Residual refinement network mode



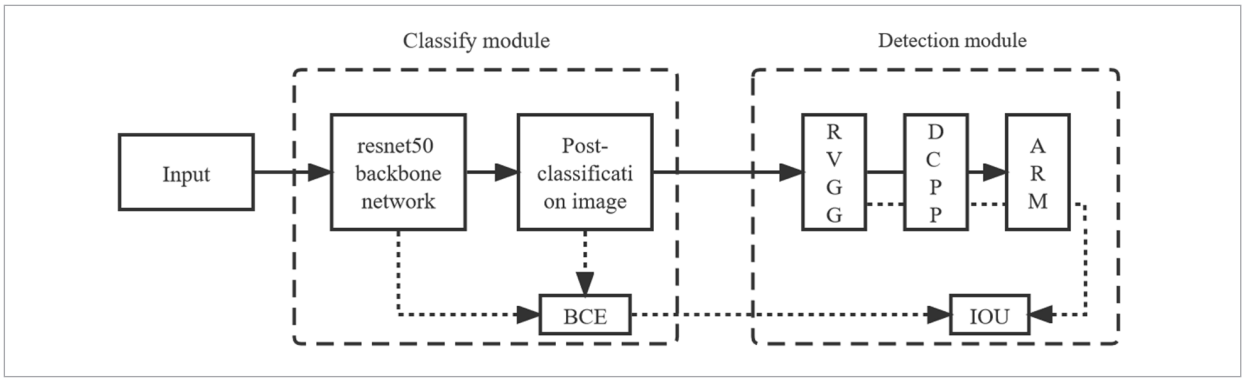
optimization policy (OSOP), the minimum number of seeds is determined adaptively to speed up the computation and optimize GPU memory. You et al. [26] extended and improved the Mean-shift algorithm with a novel Seed Selection & Early stopping method, which greatly improves the computing speed and reduces GPU memory consumption.

In summary, making saliency algorithms applied to foggy scenes, the generalization ability of the algorithm, computing speed and detection accuracy are undoubtedly among the main goals. To focus on the shortcomings in the current saliency network models, a method combining the resnet50-based classification module is proposed. The fog images with different concentrations are trained and tested using R^2Net. Data augmentation [31] and Adam optimizer are used to alleviate the overfitting problem and improve the generalization ability of the model. Use to speed up training and inference through enhanced CPU usage. Finally, the detection method of this paper is tested under foggy images with different concentrations by comparing it with traditional algorithms and currently models.

3. Proposed Method

The fog density is not stable under real foggy conditions, and the existing image dehazing algorithms are mostly aimed at the application of single image, and have no adaptive ability for fog image detection under different concentrations. Therefore, step of fog classification is added before the saliency detection, and the detection network is trained and optimized according to the classification results, thereby improving the

Figure 3
Saliency detection network in foggy weather



detection accuracy. Based on the R²Net [6] network model, we add ResNet50-based [7] foggy image classification module. The module divides the images into three types: fogless, light fog and dense fog, and then implements different detection strategies according to different types, so as to propose saliency detection method under foggy conditions.

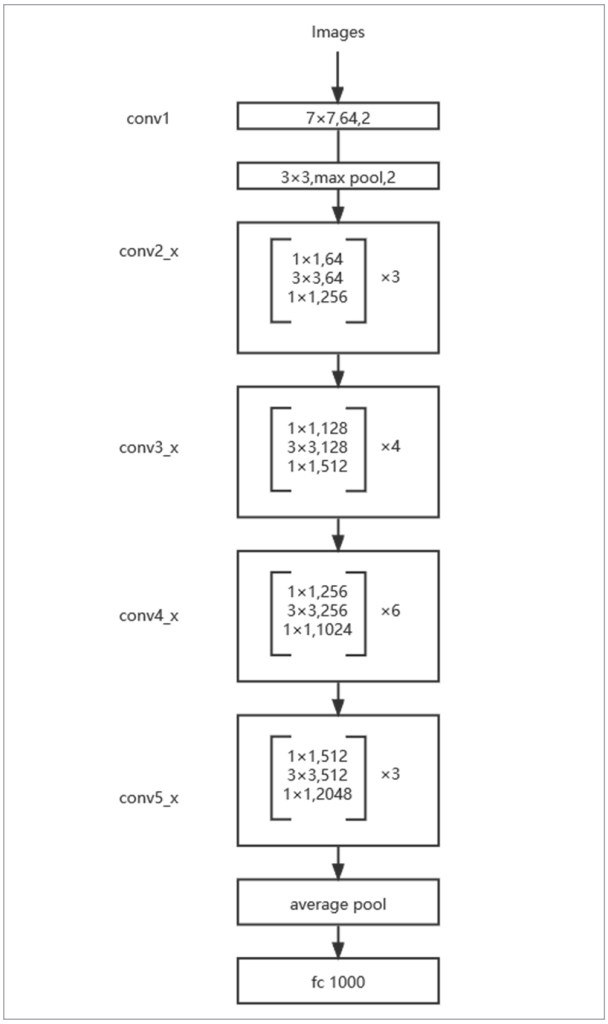
Directly predicting the best result under the influence of high foggy particles makes the saliency detection task very challenging. R²Net's residual learning strategy can gradually refine the coarse predictions. The residuals are predicted to compensate for the errors between the coarse saliency map and ground truth masks. It can generate coarse predictions through the DCPP module and guide the residual learning process through ARM. Even if the target profile is not successfully detected, the finest saliency map can be greatly approximated. The method structure is shown in Figure 3.

3.1. Classify Module

We use ResNet [7] as the basic network of classification modules. ResNet consists of several residual blocks. The principle of the residual block is to directly skip the data output of the previous layers and introduce it into the input part of the subsequent data layer. He uses $F(x)$ to represent two-layer network without skip connections, then the residual block can be expressed as $H(x) = F(x) + x$, introducing more abundant reference data for x, so that the network can learn more plentiful content.

The ResNet50 [7] structure is shown in Figure 4. The residual network consists of several residual blocks,

Figure 4
ResNet50 network structure



and the structure with multiple residual blocks is called a layer. The initial layer is an ordinary convolution structure, layer1 contains 3 residual blocks, layer2 contains 4 residual blocks, layer3 contains 6 residual blocks, layer4 contains 3 residual blocks, and finally there is a full connection layer. From layer1 to layer5, after the image data with a size of 224 pixel × 224 pixel is transmitted, the residual network extracts features for learning and training, and finally reduces the size to 7 × 7 pixel. After the residual network training, the images are input to the average pooling layer and averaged, and finally the image category is divided by the Softmax function of the fully connected layer.

3.2. Detection Module

We use R²Net [6] as the basic network of detection modules.R²Net is a novel residual structure-based saliency detection network. Unlike existing methods, the network progressively modifies the error of the prediction map and the saliency mask until it best matches the ground truth. R²Net mainly includes the R-VGG module, DCPP module and ARM module. The R-VGG module is modified from the VGG16 [19] network.

The DCPP module structure employs four dilated convolutional layers, which are used to generate the coarse saliency map. The resulting rough saliency map is fed into the bottom residual learning branch of the residual module. Except for the different rate parameters, the four dilated convolution layers are all implemented using atrous convolution. The purpose of using atrous convolution is to enlarge the receptive field without losing spatial resolution. Atrous convolution has another advantage, by setting different dilation rates to get different receptive fields. Information at different scales can be obtained from different receptive fields, which plays an important role in vision tasks.

When the image is converted into a two-dimensional matrix $x[i, j]$ and convolved with a filter $w\left[k_i, k_j\right]$ with a kernel size of K, $y[i, j]$ can be expressed as:

$$y[i, j]=\sum_{k_i=1}^{K}\sum_{k_j=1}^{K} x\left[i+r\cdot k_i, j+r\cdot k_j\right] w\left[k_i, k_j\right]\cdot \tag{5}$$

The parameter r is the similarities and differences between the atrous convolution and the classical convolution. As shown in

$$F=k+(k-1)(r-1)\,, \tag{6}$$

the expansion rate r controls the distance between adjacent elements in the convolution kernel, and its change controls the size of the receptive field F of the convolution kernel, and will not boost the number and computation of parameters.

R²Net adopts four dilated convolutional layers to form a dilated convolutional pyramid pooling module for predicting coarse global saliency maps. The network uses four dilated convolutional layers with $k = 3$ filters but different rate parameters. In order to ensure the extraction of global view and multi-scale features, the rate of the four dilated convolutional layers is set to $r=1,5,9,13,$ respectively, and the number of output channels is 16. In this paper, in order to alleviate the saliency detection of small-scale objects subject to foggy conditions, the rate is set to $r = 1, 2, 5, 9$ according to [22], and the number of output channels is unchanged. In the end, the network can still accurately extract local and global features.

3.3. Loss Function

In this paper, the binary cross entropy loss (L_{BCE}), which is often used in classification problems, is used as the loss function of the classification module. R²Net adopts the standard cross-entropy loss to calculate the per-pixel loss, ignoring the global structure of the image. To remedy this deficiency, this paper uses the IOU loss [17] (L_{IOU}) to focus on the global structure, thereby forming global constraints on the network. For our optimized network, the loss function (L_{BCE}) obtained in the classification module is used, and then the classified foggy image is sent to the detection network for training, and finally the detection result of the target is obtained. To sum up, the loss function of the model in this paper is defined as: $L=\lambda L_{BCE}+L_{IOU}$. Among them, L is the overall loss function, L_{BCE} is the loss function of the classification network module, L_{IOU} is the loss function of the detection network module, and λ is the balance factor.

4. Experiment

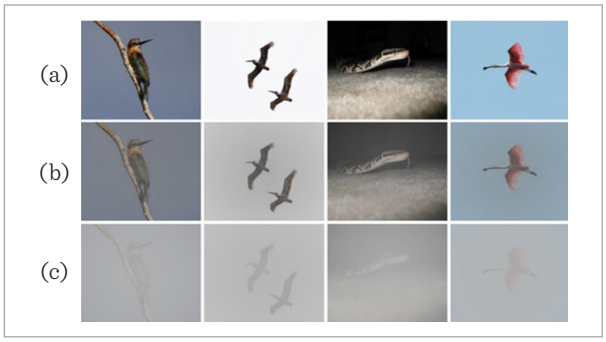
4.1. Datasets

Sakaridis [18] uses the Foggy-Cityscapes to process synthetic foggy images. There are three different concentrations of fog in this dataset. Different concentrations of fog have different β values in the atmospheric

scattering model. In this paper, $0.04 \le \beta \le 0.08$ corresponds to light fog, and $0.12 \le \beta \le 0.16$ corresponds to dense fog, the fog-free images use the DUTS [21]. Figure 5 is example of the dataset used in the text.

Figure 5

The data set (a) no fog dataset; (b) light fog dataset; (c) dense fog dataset



This paper uses DUTS-TRAIN [21] and the generated light fog and dense fog datasets as training sets, and uses DUTS-TEST [21], ECSSD [25], HKU-IS [12] and PASCAL-S [14] as test sets.

4.2. Evaluation Metrics

We evaluates the proposed method by adopting Maximun F-measure [1], S-measure [4], E-measure [3] and Mean Absolute Error (MAE). Among them, Maximun F-measure and MAE are calculated in pixel-by-pixel manner, which cannot fully capture the structural information of the prediction graph. Therefore, S-measure is supplemented to compute structural similarity and E-measure to evaluate image-level properties.

This paper adopts Mean Absolute Error (MAE) to predict the pixel-wise mean absolute error between the saliency map and the ground truth map, as follows:

$$\mathrm{MAE} = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left| S_{i,j} - G_{i,j} \right|, \tag{7}$$

where W and H represent the width and height of the saliency map, respectively, and the MAE value is normalized to {0,1} interval value. The MAE represents the similarity between the significance map and the ground truth map.

Maximum F-measure is a commonly used evaluation method, which considers both precision and recall, and uses the beta parameter to trade off precision and recall:

$$F_\beta = \frac{\left(1 + \beta^2\right) \cdot \mathrm{precision} \cdot \mathrm{recall}}{\beta^2 \cdot \mathrm{precision} + \mathrm{recall}}. \tag{8}$$

This paper follows [1] to set β^2 to 0.3 to enhance the accuracy, and we use the maximum value from all precision and recall pairs.

Structure-measure [4] considers both region-oriented and object-oriented structural similarity measures. To capture the importance of structural information in the image, S is used to evaluate the structural similarity between region-awareness (S_r) and object-awareness (S_o). Therefore, S can be defined as:

$$S = \alpha * S_o + (1 - \alpha) * S_r, \tag{9}$$

where $\alpha \in [0, 1]$ is the balance parameter.

Enhanced-alignment measure [5] is recently proposed method that considers both pixel and image level properties of expression composition, which is an effective and efficient way to evaluate saliency maps. E is proposed based on cognitive vision research to obtain image-level statistical information and its local pixel matching information. Therefore, E can be defined as:

$$E = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{i=1}^{H} \phi_{FM}(i, j), \tag{10}$$

where W and H are the height and width of the map, respectively, and ϕ_{FM} represents the enhanced diagonal matrix.

4.3. Classification Experiment

In order to better express the performance of deep learning on foggy image classification, this paper designs three classic convolutional neural network models (AlexNet [10], VGG16 [19] and ResNet50 [7]) for classification experiments and comparisons. In order to verify the effectiveness of the proposed scheme in this paper, training and testing are carried out through the DUTS [21] dataset. In the experiment, the three networks use the same parameters, batch size is set to 32, the loss function use Cross Entropy Loss, and the optimizer uses the Adam.

In the early days of CNN, researchers focused on improving the classification accuracy of the network. While CNN has developed so far, in order to reduce the time cost and hardware limitations of training and testing, it has high image classification accuracy and a small amount of parameters. Therefore, this paper adopts the Resnet50 [7] network as the basic model of the foggy image classification module.

Table 1
Performance comparison of classified networks

Method	Accuracy(%)	FLOPs	Parameters	Size
AlexNet	89.54	7×10^{9}	62369155	233MB
VGG16	93.52	1.5×10^{11}	138357544	528MB
ResNet50	94.35	3.8×10^{10}	46159168	98MB

Under the same parameter settings, it can be seen from Table 1: (1) AlexNet [10], as the most classic convolutional neural network, still has an accuracy rate of 89.54% in the three-class experiment, but its 0.7 GFLOPs cannot meet the network requirements; (2) VGG16 [19], due to its deep network layers, although the accuracy rate is as high as 93.52%, the complexity and size of the model are not conducive to the optimization of the overall network; (3) Resnet50 [7] can achieve 94.35% with the least parameters and the smallest memory footprint It can meet the classification task performance while shortening the training time and reducing the complexity of the training model.

4.4. Experimental Results and Analysis

In this paper, the experimental environment used are Windows 10, the CPU is Inter(R) Core i9-9900K @3.6GHz and the GTX2080TI GPU is used for training. The Python version uses Version 3.7, Torch version uses version 1.2.0. During training, the batch size is set to 8. We set the momentum parameter to 0.9, the weight decay to 0.001, and the learning rate to 5e-5. The Adam are selected to train our networks.

4.4.1. Comparative Analysis

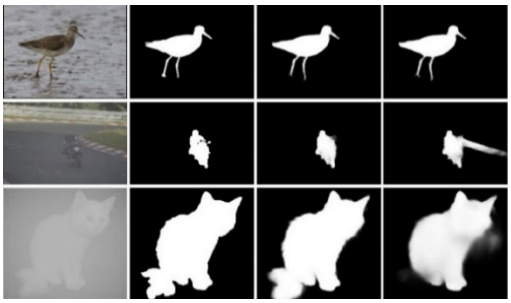
The experimental training data set adopts the DUTS-TRAIN [21], which contains 5019 fogless images, 5019 simulated light fog images, and 5019 simulated dense fog images. By comparing whether the detection performance is improved before and after the introduction of the classification module, the method is verified. The effectiveness of the module, the experimental results are shown in Table 2.

Table 2
Comparison results of R^2Net network before and after improvement; M-F(Maximum F-Measure, Larger is Better); E-m(E-Measure, Larger is Better); S-m(S-Measure, Larger is Better); MAE(Small is Better); Fogless, Light and Dense(The degree of fog); The best results of fogless are shown in bold

		M-F↑	E-m↑	S-m↑	MAE↓	M-F↑	E-m↑	S-m↑	MAE↓
Methods		DUTS-TEST				ECSSD			
OURS	Fogless	0.861	0.926	0.886	0.040	**0.937**	**0.958**	**0.928**	0.038
	Light	0.860	0.926	0.885	0.040	0.936	0.956	0.926	0.038
	Dense	0.836	0.914	0.8675	0.048	0.920	0.945	0.911	0.049
		HKU-IS				PASCAL-S			
	Fogless	**0.924**	0.958	0.919	0.032	**0.849**	**0.895**	**0.862**	0.069
	Light	0.921	0.957	0.917	0.033	0.844	0.894	0.859	0.070
	Dense	0.912	0.953	0.910	0.038	0.819	0.878	0.838	0.082
R^2Net		DUTS-TEST				ECSSD			
	Fogless	**0.863**	**0.927**	**0.886**	**0.040**	0.935	0.956	0.926	**0.039**
	Light	0.828	0.906	0.862	0.047	0.925	0.948	0.917	0.043
	Dense	0.477	0.662	0.602	0.110	0.629	0.744	0.661	0.136
		HKU-IS				PASCAL-S			
	Fogless	0.923	**0.959**	**0.920**	**0.033**	0.842	0.890	0.858	**0.071**
	Light	0.905	0.945	0.905	0.037	0.815	0.872	0.838	0.080
	Dense	0.642	0.765	0.691	0.106	0.499	0.648	0.577	0.177

It can be concluded from Table 2: (1) Under the fogless datasets, the four evaluation indicators of the DUTS-TEST [21] have maximum difference of about 0.14% with that of R²Net; under the HKU-IS [12], the E-measure is similar to the S-measure. Compared with R²Net, it is 0.0007 behind; under the ECSSD [25] and PASCAL-S [14], the indicators are similar to R²Net. (2) In the light fog datasets, the evaluation indicators of the improved network in each data set are slightly higher than R²Net. (3) Under the dense fog datasets, the improved network has significant advantages over R²Net in various evaluation indicators under each datasets. It can be seen from Figure 6 that: (1) In the case of no fog, the network before and after the improvement can accurately identify the outline of the bird and even the beak. (2) In light fog, our method successfully detects small-scale objects. (3) In the case of dense fog, our method can still clearly detect the outline of the cat, and the network performance before and after the improvement is clearly distinguished. It can be seen that the robustness of R²Net to images with different degrees of foggy degradation is not strong, and the network is more suitable for the situation where both training images and test images are clear images.

Figure 6
Subjective visual contrast



4.4.2. Comparison with Other Methods

This paper compares the improved network with other methods, including four deep learning methods (PoolNet [15], U²Net [16], PurNet [13], CSNet [2]) and four traditional algorithms (FT [1], HC [3], IT [8], LC [28]). For fair comparison, other deep learning networks are trained and tested in the same environment, and the same dataset is used for both training and testing.

As shown in Table 3, this paper presents the objective evaluation results of saliency map and saliency segmentation. In this paper, MAE and S-measure are used to evaluate non-binary saliency maps, and Max F-measure and E-measure are used to evaluate binary saliency segmentation. It can be seen from Figure 7 and Table 3: (1) From the subjective vision, for

Table 3
Detection results under the fogless dataset. The best results of fogless are shown in bold. D(DUTS); E(ECSSD); H(HKU-IS); P(PASCAL-S)

Methods		OURS	U²Net	PoolNet	PurNet	CSNet	FT	HC	IT	LC
D	M-F↑	0.861	0.823	0.852	0.859	0.779	0.291	0.224	0.185	0.284
	E-m↑	0.926	0.895	0.919	0.922	0.875	0.606	0.584	0.603	0.631
	S-m↑	0.886	0.858	0.879	0.882	0.823	0.472	0.420	0.410	0.478
	MAE↓	0.041	0.054	0.039	0.038	0.074	0.233	0.327	0.356	0.254
E	M-F↑	0.937	0.929	0.931	0.936	0.895	0.370	0.316	0.288	0.364
	E-m↑	0.958	0.949	0.950	0.955	0.929	0.595	0.569	0.588	0.619
	S-m↑	0.928	0.918	0.919	0.923	0.891	0.447	0.413	0.413	0.459
	MAE↓	0.038	0.041	0.039	0.036	0.066	0.290	0.362	0.386	0.304
H	M-F↑	0.924	0.915	0.913	0.927	0.881	0.373	0.287	0.255	0.369
	E-m↑	0.958	0.948	0.950	0.958	0.933	0.628	0.585	0.612	0.657
	S-m↑	0.919	0.908	0.910	0.917	0.882	0.477	0.427	0.420	0.494
	MAE↓	0.032	0.037	0.033	0.030	0.059	0.252	0.342	0.371	0.268
P	M-F↑	0.849	0.815	0.842	0.841	0.795	0.357	0.301	0.292	0.360
	E-m↑	0.895	0.867	0.891	0.886	0.860	0.552	0.514	0.541	0.581
	S-m↑	0.862	0.832	0.852	0.849	0.817	0.427	0.384	0.397	0.451
	MAE↓	0.069	0.086	0.071	0.069	0.102	0.313	0.390	0.384	0.316

Figure 7
Visually detected results at different concentrations

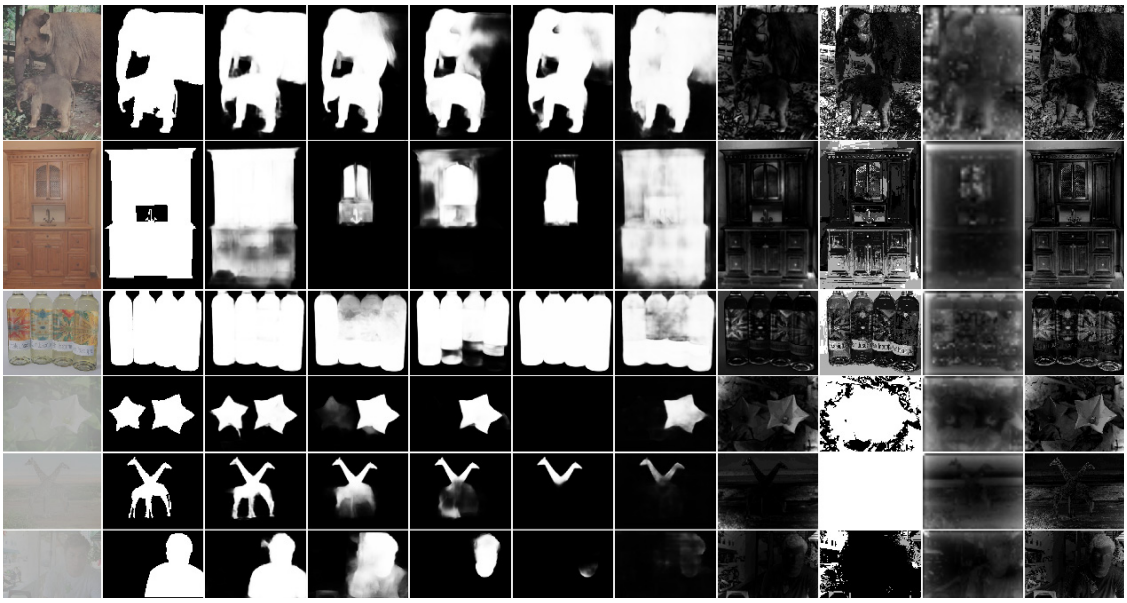


Table 4
Detection results under the light fog dataset. The best results of light fog are shown in bold. D(DUTS); E(ECSSD); H(HKU-IS); P(PASCAL-S)

Methods		OURS	U2Net	PoolNet	PurNet	CSNet	FT	HC	IT	LC
D	M-F↑	0.861	0.779	0.798	0.823	0.732	0.251	0.189	0.179	0.251
	E-m↑	0.926	0.868	0.885	0.896	0.846	0.607	0.584	0.604	0.629
	S-m↑	0.885	0.831	0.841	0.856	0.793	0.447	0.394	0.358	0.456
	MAE↓	0.040	0.063	0.048	0.045	0.084	0.229	0.331	0.357	0.248
E	M-F↑	0.936	0.911	0.909	0.919	0.869	0.327	0.287	0.279	0.325
	E-m↑	0.956	0.936	0.934	0.939	0.916	0.584	0.564	0.563	0.612
	S-m↑	0.926	0.905	0.903	0.907	0.871	0.416	0.379	0.327	0.425
	MAE↓	0.038	0.047	0.047	0.043	0.074	0.292	0.369	0.406	0.308
H	M-F↑	0.921	0.890	0.887	0.912	0.847	0.331	0.255	0.231	0.333
	E-m↑	0.957	0.931	0.932	0.947	0.912	0.624	0.591	0.597	0.652
	S-m↑	0.917	0.890	0.889	0.904	0.856	0.450	0.401	0.351	0.464
	MAE↓	0.033	0.044	0.039	0.034	0.069	0.253	0.347	0.382	0.271
P	M-F↑	0.844	0.791	0.809	0.805	0.755	0.318	0.285	0.283	0.327
	E-m↑	0.894	0.847	0.867	0.853	0.832	0.547	0.530	0.524	0.571
	S-m↑	0.859	0.813	0.826	0.821	0.788	0.393	0.355	0.326	0.411
	MAE↓	0.070	0.096	0.083	0.082	0.116	0.308	0.385	0.399	0.317

multiple targets (the first row), we accurately detect two targets with different scales. For large objects (second row), we pinpoint the location of the object. Our method is also robust to complex background objects (fourth row). (2) In the case of light fog and dense fog, the MAE and S-measure indicators of the improved network are better than other methods on the four datasets, which shows that our saliency map is similar to the ground truth map and has good region-aware and object-aware structural similarity. (3) Max F-measure and E-measure show that our saliency map has consistently high confidence in the target region, which can efficiently detect the location of the most prominent target and segment it.

Table 5

Detection results under the dense fog dataset. The best results of dense fog are shown in bold. D(DUTS); E(ECSSD); H(HKU-IS); P(PASCAL-S)

Methods		OURS	U2Net	PoolNet	PurNet	CSNet	FT	HC	IT	LC
D	M-F↑	**0.836**	0.553	0.496	0.348	0.536	0.180	0.179	0.179	0.185
	E-m↑	**0.914**	0.733	0.664	0.537	0.728	0.60	0.415	0.595	0.618
	S-m↑	**0.867**	0.675	0.591	0.568	0.634	0.409	0.314	0.332	0.403
	MAE↓	**0.048**	0.118	0.110	0.107	0.140	0.204	0.399	0.330	0.236
E	M-F↑	**0.920**	0.730	0.634	0.399	0.691	0.279	0.279	0.279	0.279
	E-m↑	**0.945**	0.816	0.727	0.543	0.792	0.577	0.395	0.543	0.579
	S-m↑	**0.911**	0.770	0.650	0.568	0.674	0.374	0.295	0.282	0.358
	MAE↓	**0.049**	0.109	0.136	0.161	0.156	0.281	0.450	0.393	0.311
H	M-F↑	**0.912**	0.702	0.641	0.491	0.672	0.245	0.231	0.231	0.246
	E-m↑	**0.953**	0.804	0.759	0.628	0.797	0.609	0.399	0.588	0.625
	S-m↑	**0.910**	0.752	0.662	0.622	0.686	0.401	0.312	0.314	0.397
	MAE↓	**0.038**	0.104	0.108	0.117	0.134	0.239	0.417	0.363	0.269
P	M-F↑	**0.819**	0.602	0.530	0.283	0.568	0.283	0.283	0.283	0.283
	E-m↑	**0.878**	0.722	0.665	0.410	0.690	0.534	0.396	0.521	0.543
	S-m↑	**0.838**	0.666	0.561	0.464	0.588	0.364	0.306	0.285	0.351
	MAE↓s	**0.082**	0.159	0.177	0.208	0.204	0.285	0.401	0.390	0.311

5. Conclusion

We propose a foggy saliency detection network based on R^2Net. In terms of network structure, R^2Net is selected as the basic network, and the fog concentration classification module is added, so that the network can judge the fog concentration in the image and select the subsequent work module accordingly. Through the atmospheric scattering model, the foggy degradation process was simulated, and two types of simulated foggy images of «light fog» and «dense fog» were generated to expand the datasets. Compared with the original R^2Net, the algorithm effectively improves the accuracy of saliency detection in foggy weather, improves the robustness and generalization ability of the network, and provides a new idea for saliency detection in foggy images. In addition, the universality of the fog density classification module to other network models still needs to be improved, and the light-weight architecture of the fog density classification module is also worthy of further research.

Acknowledgement

This work was supported by the Natural Science Foundation of Shanghai under Grant 18ZR1425800, the National Natural Science Foundation of China under Grant 61875125, 62275153 and 62005165, the development fund for Shanghai talents under Grant 2021005 and Shanghai Industrial Collaborative Innovation Project HCXBCY-2022-006.

References

1. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S. Frequency-tuned Salient Region Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, 1597-1604. https://doi.org/10.1109/CVPR.2009.5206596
2. Cheng, M. M., Gao, S. H., Borji, A., Tan, Y. Q., Lin, Z., Wang, M. A Highly Efficient Model to Study the Semantics of Salient Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44, 8006-8021. https://doi.org/10.1109/TPAMI.2021.3107956
3. Cheng, M., Mitra, N., Huang, X., Huang, X., Hu, S. Global Contrast Based Salient Region Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37, 569-582. https://doi.org/10.1109/TPAMI.2014.2345401
4. Fan, D., Cheng, M., Liu, Y., Li, T., Borji, A. Structure-measure: A New Way to Evaluate Foreground Maps. International Conference on Computer Vision, 2017, 4548-4567. https://doi.org/10.1109/ICCV.2017.487
5. Fan, D., Gong, C., Cao, Y., Ren, B., Cheng, M. M., Borji, A. Enhanced-alignment Measure for Binary Foreground Map Evaluation. International Joint Conference on Artificial Intelligence, 2018, 698-704. https://doi.org/10.24963/ijcai.2018/97
6. Feng, M. Y., Lu, H. C., Yu, Y. Z. Residual Learning for Salient Object Detection. IEEE Transactions on Image Processing, 2020, 29, 4696-4708. https://doi.org/10.1109/TIP.2020.2975919
7. He, K., Zhang, X., Ren, S., Sun, J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90
8. Itti, L., Koch, C., Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20, 1254-1259. https://doi.org/10.1109/34.730558
9. Jin, B., Cruz, L., Goncalves, N. Deep Facial Diagnosis: Deep Transfer Learning from Face Recognition to Facial Diagnosis. IEEE Access, 2020, 8, 123649-123661. https://doi.org/10.1109/ACCESS.2020.3005687
10. Krizhevsky, A., Sutskever, I., Hinton, E. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 2017, 60, 84-90. https://doi.org/10.1145/3065386
11. Li, G. B., Yu, Y. Z. Deep Contrast Learning for Salient Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 478-487. https://doi.org/10.1109/CVPR.2016.58
12. Li, G. B., Yu, Y. Z. Visual Saliency Based on Multiscale Deep Features. IEEE Conference on Computer Vision and Pattern Recognition, 2015, 5455-5463.
13. Li, J., Su, J. M., Xia, C. Q., Ma, M. C., Tian, Y. H. Salient Object Detection with Purificatory Mechanism and Structural Similarity Loss. IEEE Transactions of Image Processing, 2021, 30, 6855-6868. https://doi.org/10.1109/TIP.2021.3099405
14. Li, Y., Hou, X., Koch, C., Rehg, J., Yuille, A. The Secrets of Salient Object Segmentation. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 280-287. https://doi.org/10.1109/CVPR.2014.43
15. Liu, J. J., Hou, Q. B., Cheng, M. M., Feng, J. S., Jiang, J. A Simple Pooling-Based Design for Real-Time Salient Object Detection. IEEE Conference on Computer Vision and Pattern Recognition, 2019, 3912-3921. https://doi.org/10.1109/CVPR.2019.00404
16. Qin, X. B., Zhang, Z. C., Huang, C. Y., et al. U-2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. Pattern Recognition, 2020, 106, 107404. https://doi.org/10.1016/j.patcog.2020.107404
17. Qin, X. B., Zhang, Z. C., Huang, C. Y., Gao, C., Dehgan, M., Jagersand, M. BASNet: Boundary-Aware Salient Object Detection. IEEE Conference on Computer Vision and Pattern Recognition, 2019, 7471-7481. https://doi.org/10.1109/CVPR.2019.00766
18. Sakaridis, C., Dai, D., van Gool, L. Semantic Foggy Scene Understanding with Synthetic Data. International Journal of Computer Vision, 2018, 126, 973-992. https://doi.org/10.1007/s11263-018-1072-8
19. Simonyan, K., Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556v6. 2014.
20. Szegedy, C., Liu, W., Jia, Y. Q., Sermanet, P., et al. Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594
21. Wang, L. J., Lu, H. C., Wang, Y. F., et al. Learning to Detect Salient Objects with Image-level Supervision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 3796-3805. https://doi.org/10.1109/CVPR.2017.404
22. Wang, P. Q., Chen, P. F., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G. Understanding Convolution for Seman-

tic Segmentation. IEEE Conference on Computer Vision and Pattern Recognition, 2018, 1451-1460. https://doi.org/10.1109/WACV.2018.00163

23. Wang, W. G., Shen, J. B., Yu, Y. Z., Ma, K. L. Stereoscopic Thumbnail Creation via Efficient Stereo Saliency Detection. IEEE Transactions on Visualization and Computer Graphics, 2017, 23, 2014-2021. https://doi.org/10.1109/TVCG.2016.2600594
24. Xing, W. W., Bai, P. P., Zhang, S. L., Bao, P. Scene-specific Pedestrian Detection Based on Transfer Learning and Saliency Detection for Video Surveillance. Automatic Control and Computer Sciences, 2017, 51, 180-192. https://doi.org/10.3103/S0146411617030099
25. Yan, Q., Xu, L., Shi, J., Jia, J. Hierarchical Saliency Detection. IEEE Conference on Computer Vision and Pattern Recognition, 2013, 1155-1162. https://doi.org/10.1109/CVPR.2013.153
26. You, L., Jiang, H., Hu, J., Chang, C. H., Chen, L. GPU-accelerated Faster Mean Shift with Euclidean Distance Metrics. 2022, 2022 IEEE 46th Annual Computers, Software, and Applications Conference, 211-216. https://doi.org/10.1109/COMPSAC54236.2022.00037
27. Yu, Z., Zhuge, Y. Z., Lu, H. C., Zhang, L. H. Joint Learning of Saliency Detection and Weakly Supervised Semantic Segmentation. Proceedings of IEEE International Conference on Computer Vision, 2019, 7222-7232. https://doi.org/10.1109/ICCV.2019.00732
28. Zhai, Y., Shah, M. Visual Attention Detection in Video Sequences Using Spatiotemporal Cues. Proceedings of the ACM International Conference on Multimedia, 2006, 815-824. https://doi.org/10.1145/1180639.1180824
29. Zhang, P., Wang, D., Lu, H. C., Wang, H. Y., Ruan, X. Amulet: Aggregating Multi-level Convolutional Features for Salient Object Detection. Proceedings of the IEEE International Conference on Computer Vision, 2017, 202-211. https://doi.org/10.1109/ICCV.2017.31
30. Zhao, M. Y., Jha, A., Liu, Q., Millis, B. A., Mahadevan-Jansen, A., Lu, L., Landman, B. A., Tyska, M. J., Huo, Y. Faster Mean-shift: GPU-accelerated Clustering for Cosine Embedding-based Cell Segmentation and Tracking. Medical Image Analysis, 2021, 71. https://doi.org/10.1016/j.media.2021.102048
31. Zheng, Q. H., Yang, M. Q., Tian, X. Y., et al. A Full Stage Data Augmentation Method in Deep Convolutional Neural Network for Natural Image Classification. Discrete Dynamics in Nature and Society, 2020. https://doi.org/10.1155/2020/4706576
32. Zheng, Q. H., Yang, M. Q., Yang, J., et al. Improvement of Generalization Ability of Deep CNN via Implicit Regularization in Two-Stage Training Process. IEEE Access, 2018, 6, 15844-15869. https://doi.org/10.1109/ACCESS.2018.2810849
33. Zheng, Q. H., Zhao, M. Y., Li, Y., Wang, H. J., Yang, Y. Spectrum Interference-based Two-level Data Augmentation Method in Deep Learning for Automatic Modulation Classification, 2020, 33, 7723-7745. https://doi.org/10.1007/s00521-020-05514-1