Expert Systems With Applications 234 (2023) 121151

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multi visual feature fusion based fog visibility estimation for expressway surveillance using deep learning network

Wenchen Yang [a,b], Youting Zhao [c,*], Qiang Li [d], Feng Zhu [e], Yu Su [a,b]

[a] National Engineering Laboratory for Surface Transportation Weather Impacts Prevention, Broadvision Engineering Consultants Co.Ltd., Kunming 650200, China
[b] Yunnan Key Laboratory of Digital Communications, 650103 Kunming, China
[c] School of Automobile and Transportation Engineering, Guangdong Polytechnic Normal University, 293 Zhongshan West Road, Guang Zhou 510665, China
[d] School of Computer Science and Engineering, Sun Yat-sen University, 135 Xingang West Road, Guang Zhou 510275, China
[e] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Keywords:
Intelligent transportation system
Image dataset
Multi visual feature fusion
Fog visibility estimation
Deep learning network

ABSTRACT

Visibility in foggy weather is of great value for traffic management and pollution monitoring. However, vision-based fog visibility estimation methods are usually based on a single image to approximate the visibility in foggy weather, and most existing data-driven machine learning models struggle to capture effective features and achieve high estimation accuracy due to the severe image degradation caused by reduced visibility and lack of real scene images. Therefore, this paper proposes a novel deep learning framework based on multi visual feature fusion for fog visibility estimation, named VENet, which comprises of two subtask networks (for fog level classification and fog visibility estimation) constructed in a cascade structure. A special feature extractor and an anchor-based regression method (ARM) are proposed to help improve the accuracy. Further, a standard Fog Visibility Estimation Image (FVEI) dataset containing 15,000 images of real fog scenes is established. This dataset greatly bridges the lack of suitable data in the field of vision-based visibility estimation. Extensive experiments have been conducted to demonstrate the performance of the proposed VENet, where the error of fog visibility estimation is less than 5% at 500 m and the fog level classification accuracy is at least 92.3%. In addition, the proposed VENet has been applied on Yunnan Xiangli and Mazhao Expressway surveillance with promising performance in practice.

1. Introduction

Fog is a common weather phenomenon and a potential threat to public safety, especially in transportation. Traffic control in foggy weather has become one of the most challenging tasks in traffic management due to reduced visibility. According to the statistics of domestic traffic accidents in China (Traffic Administration Bureau of the Ministry of Public, 2021), highway accidents account for about 11.3% of the total number of fatalities, of which about 1/4 are caused by severe weather conditions such as dense fog, with a fatality rate of over 40%. Therefore, accurate visibility estimation in foggy weather is key to alleviating transportation-related mishaps, and is also a part of supporting intelligent transportation systems (Xue, Xu, & Du, 2022).

In general, there are two main approaches for fog visibility estimation, namely sensor-based methods and vision-based methods. Sensor-based methods (Wang, Jia, Li, Lu, & Hua, 2020; Xian, Han, Huang, Sun, & Li, 2018) are supported by optical theories and can achieve high accuracy. However, these methods require special equipment which are often expensive and require specific installation conditions; therefore, they can only be applied in a limited number of observatories. In addition, such observatories are geographically scattered, making it difficult to detect the occurrence of fog. With the development of computer vision, vision-based methods can leverage preinstalled road surveillance cameras for the task, without relying on special equipment. Furthermore, with the high installation density of existing surveillance cameras, vision-based methods (You, Jia, Pei, & Yao, 2022) can help establish a real-time fog visibility observation system. Therefore, compared with traditional sensor-based methods, vision-based fog visibility estimation methods are superior in terms of both cost and efficiency, and are of great value in traffic management and pollution monitoring. It is also a

* Corresponding author.
E-mail addresses: tongjiywc@163.com (W. Yang), zhaoyouting@gpnu.edu.cn (Y. Zhao), liqiang27@mail2.sysu.edu.cn (Q. Li), zhufeng@ntu.edu.sg (F. Zhu), ynsuyu2022@163.com (Y. Su).

https://doi.org/10.1016/j.eswa.2023.121151
Received 22 April 2023; Received in revised form 25 July 2023; Accepted 7 August 2023
Available online 9 August 2023

popular development technology in the field of intelligent transportation systems in recent years, which can support cutting-edge technologies such as autonomous driving.

Vison-based methods for fog visibility estimation can be divided into two categories: fog level classification and fog visibility estimation. Fog level classification involves separating images into different classes based on visibility. In a previous study, Li, Lu, Tong, and Zeng (2014) classified fog levels via statistical methods, which can effectively classify images but cannot achieve high accuracy. To overcome this disadvantage, Wang, Jia, Li, Lu, and Hua (2020) proposed the construction of a fog level classifier based on machine learning models. Compared with previous methods, the classification accuracies were improved; however, their performance drops rapidly when the fog level increases. The main reason is that image degradation becomes more serious as the fog level increases, thereby causing difficulties in capturing effective features.

Fog visibility estimation can be considered as a regression problem of estimation, which is usually applied in the autonomous driving assistance system. This method requires special landmarks in the scene, such as lanes, sky, etc., which limits its application. To address this issue, learning-based methods with higher robustness are proposed (You, Lu, Wang, & Tang, 2018). However, due to the lack of sufficient data and suitable estimation models, it is difficult for existing methods to achieve high accuracy in visibility estimation.

To address the above research gaps, a novel fog visibility estimation network named VENet is proposed, which consists of two subtask networks, including a fog level classification network and a fog visibility estimation network. Inspired by Dai, He, and Sun (2016) a cascade structure using these two networks is constructed. The results of fog level classification are first input to the fog visibility estimation network to realize a coarse-to-fine process; then, an anchor-based regression method (ARM) is proposed to improve the estimation accuracy. Moreover, to overcome the challenges posed by image degradation, a special feature extractor that captures effective features from fog-scene images is built.

It is worth noting that a standard Fog Visibility Estimation Image (FVEI) dataset containing 15,000 images of real fog scenes is established. The construction of this dataset took nearly a year of hard work, including site selection, data acquisition, data preprocessing, data annotation, and acceptance. There are 10 experts in the field of transportation and meteorology invited to participate in the data annotation task. The dataset of FVEI greatly makes up for the lack of data in the field of vision-based visibility estimation.

Extensive experiments were subsequently conducted to demonstrate the performance of the proposed VENet. The error of fog visibility estimation was observed to be less than 5% at 500 m, and the fog level classification accuracy was approximately 92.3% or higher. The main contributions of this study are summarized as follows:

- A standard fog visibility estimation image (FVEI) dataset has been constructed containing 15,000 annotated images, where each image is annotated with two labels including fog level and visibility.
- Multi feature fusion methods are used to simultaneously extract multiple features for classifier training to achieve feature complementarity and reduce the impact of inherent defects in a single feature. By analyzing the characteristics of fog images, the RGB three channel features, dark channel features, and edge features are fused together to achieve fog visibility estimation.
- A deep learning based fog visibility estimation network (VENet) on multi visual feature fusion is proposed, which adopts a multitask network cascade structure to realize a coarse-to-fine recognition process. Extensive experiments are performed and the results show that the proposed model can achieve accurate visibility estimation with an error of less than 5% in the range of 500 m.
- The VENet has been applied on Yunnan Xiangli and Mazhao Expressway surveillance, and has achieved promising results in practice.

2. Literature review

2.1. Visibility estimation

Vision-based visibility estimation has garnered increasing attention in recent years. The estimation methods can be mainly divided into two categories: rule-based methods and learning-based methods. Rule-based methods usually apply graphical operators to exact image features and then build visibility estimations according to statistical analysis. For example, Hautiere, Tarel, Lavenant, and Aubert, (2006) developed an automatic fog detection system for visibility estimation using Koschmieder's laws; Bronte, Bergasa, and Alcantarilla, (2009) proposed visibility estimations based on the sky-road limit using a monocular camera. These methods are advantageous in terms of efficiency, but their main drawback is lack of accuracy. Learning-based methods are data-driven methods that use machine learning to build visibility estimations. In the early exploration stage of these methods, models were constructed with handcrafted features. Pavlić, Belzner, Rigoll, and Ilić, (2012) proposed a classifier based on image descriptors that were extracted using fast Fourier transform (FFT) and support vector machine (SVM). Asery, Sunkaria, Sharma, and Kumar, (2016) proposed the gray-level co-occurrence matrix features to train an SVM-based classification model. These approaches achieve good accuracy in the classification of foggy and non-foggy images. With the development of deep-learning approaches, You, Lu, Wang, and Tang, (2018) proposed a relative convolutional neural network and recurrent neural network (CNN-RNN) model for feature detection and built visibility estimation models based on SVM or support vector regression (SVR). Palvanov, and Cho, (2019) proposed Visnet model using three streams of deep integrated CNNs which can further solve the problem of fog level classification.

2.2. Multitask learning

Multitask learning (MTL) is a learning paradigm in machine learning that aims to exploit useful information contained in multiple related tasks to improve the generalization performances across all tasks. This approach has been successfully applied in many machine learning applications ranging from natural language processing to computer vision (Crawshaw, 2020).

In general, MTL is performed by parameter sharing of the hidden layers, and multiple tasks are processed separately. For example, MultiNet was proposed for the detection, classification, and semantic segmentation (Teichmann123, Weber, Zöllner, & Cipolla, 2018). A novel end-to-end CNN-based multitask weather recognition network with multi-scale weather cues was proposed (Xie, Huang, Zhang, Qin, & Lyu, 2022). MTL is tailored for multiple tasks that are internally related with no apparent progressive relationship. For multiple tasks with a progressive relationship, constructing a multitask cascade would be a better solution.

2.3. Image dataset

Datasets are the cornerstone of any data-driven machine learning model and are of great significance in driving the development of various applications. For example, early efforts on the handwriting dataset MNIST (LeCun, Bottou, Bengio, & Haffner, 1998), face recognition dataset (Georghiades, Belhumeur, & Kriegman, 2001), face expression dataset (Lyons, Akamatsu, Kamachi, & Gyoba, 1998), and object recognition dataset (Griffin, Holub, & Perona, 2007) have greatly stimulated popular interest in these fields, which drives the development of many computer vision models. Datasets in recent years are often built on a large scale, such as the natural scene image dataset ImageNet

(Deng, Dong, Socher, Li, Li, & Li, 2009) and Microsoft-COCO (Lin et al., 2014), and more area-specific ones such as the medical image dataset MURA (Rajpurkar et al., 2018) and cartoon image dataset Danbooru (Branwen, & Gokaslan, 2019). These datasets have advanced the development of machine learning technologies while rendering the vision tasks more vivid and interesting.

However, for vision-based visibility estimation, there are only the synthetic dataset FROSI (Belaroussi, & Gruyer, 2014), FRIDA (Tarel, Hautière, Cord, Gruyer, & Halmaoui, 2010) and FRIDA2 (Tarel, Hautière, Caraffa, Cord, Halmaoui, & Gruyer, 2012) which contain multiple synthetic scenes and traffic signs, and the annotation accuracy of fog concentration levels can only meet the classification requirements to a certain extent. Moreover, since the dataset is not composed of real scene images, it cannot accurately describe uneven fog concentrations in real scenes. It only contains the fog concentration levels without detailed information such as accurate visibility values. The lack of data limits the application of data-driven model in visibility estimation. To bridge the gap, this study presents a new dataset (i.e., FVEI) of real fog-scene images. The comparison of FVEI with FROSI, FRIDA, and FRIDA2 is shown in Table 1.

3. Pipeline of the proposed method

The pipeline of the proposed method is illustrated in Fig. 1, which consists of five main steps: (1) fog image construction, (2) multi features extraction, (3) the proposed multi-network cascade structure to realize fog level classification and fog visibility estimation, (4) experiment verifications, and (5) engineering application.

Convolutional neural network models are data-driven, so first it need to build a dataset of images with fog level and visibility value labels to train the network model (see Section 4). Secondly, the quality of image features is critical to the final recognition, greatly affecting the accuracy, efficiency, and generalization performance of classification of deep learning network. Therefore, it is necessary to extract and select features that can effectively capture the fog image through comparative analysis, such as dark channels, edge features, et al. (see Section 5). Thirdly, a novel deep learning framework based on multi visual feature fusion for fog visibility estimation, which comprises of two subtask networks constructed in a cascade structure (see Section 6). Finally, the validity of the deep learning network model is verified and an application system is built for practical application at expressways (see Section 7).

4. Dataset

As convolutional neural network models are data-driven, in this section a dataset of images with fog level and visibility value labels is built to train the neural network model. The boom in datasets has fueled the development of computer vision methods in all walks of life and industry. In previous studies, only a few datasets were available for vision-based fog visibility estimation. Furthermore, these datasets usually use synthetic image data that lack real visibility measurement; thus, they are inadequate for actual scenes. The amount of actual scenes visibility data is very limited, which might lead to over fitting or underfitting. In this work, a real fog-scene image dataset is built (named the FVEI dataset), where the data comes from three real highway

Table 1
Comparison between the FROSI, FRIDA, FRIDA2, and FVEI datasets.

Statistic	FROSI	FRIDA	FRIDA2	FVEI
Number	3024	90	330	15,000
Image Size	640*480	640*480	640*480	1920*1080
Scenes	16	18	66	3
Source	Synthetic	Synthetic	Synthetic	CCTV Camera
Label	Fog level only	Fog level only	Fog level only	Fog level and visibility

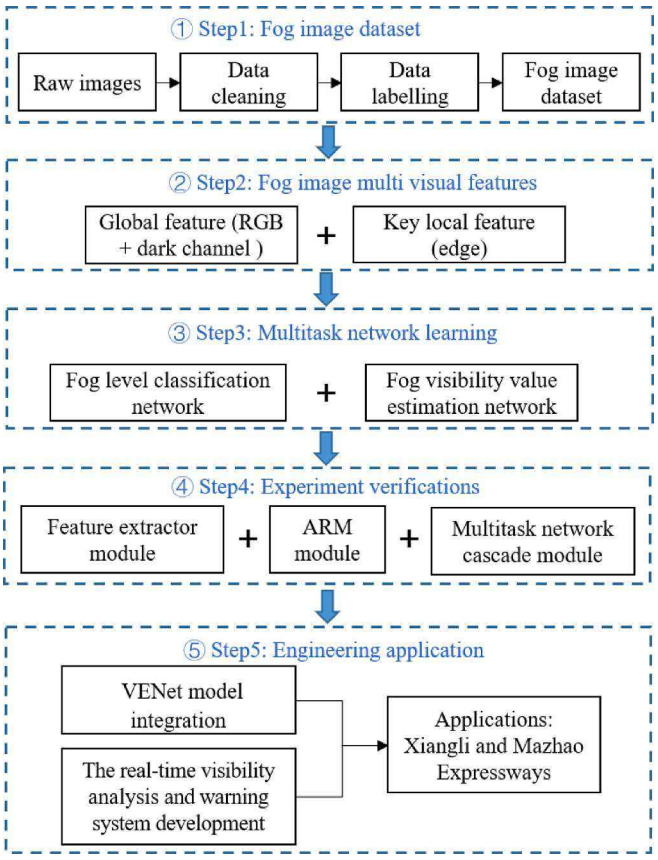

Fig. 1. Pipeline of the proposed multi visual feature fusion based fog visibility estimation method.

monitoring scenes, as shown in Fig. 2. This dataset greatly compensates for the lack of data in the field of vision-based visibility estimation, and provides great support for future research. The dataset with its ground truth is partly publicly available (https://pan.baidu.com/s/1JP8WBue5C2WTG5I_plc9ew?pwd=s427).

4.1. Collection of images

When setting up the data collection points, scenes with rich backgrounds are prioritized. The main reason for this is that reduced visibility can lead to severe image degradations, which are mainly manifested by reduced contrast and color saturation. Therefore, the fog visibility of an image is estimated from the presence of background information. In addition, locations with point light sources in the background are selected so as to enable visibility estimation of fog at night.

After determining the locations at which equipment should be installed, the devices are set up accordingly. The equipment chosen for capturing the data is the dome camera (camera model: DH-SD6A82C-HN), and the visibility measurement equipment is PWD22. They are mounted at the same location and separated by a distance of less than 50 cm to avoid inconsistencies between images and visibility measurements due to uneven distribution of fog concentration.

4.2. Image labelling

The dome camera and visibility measurement equipment are located in the same location, less than 50 cm apart. The visibility data of the visibility measurement equipment is consistent with the visibility data of the image captured by the dome camera. First, the images have to be aligned with the time stamps of the measurement data from the visibility meter to obtain the fog visibility data for each image, while the invalid

Fig. 2. Three real expressway surveillance scenes in the FVEI dataset.

image data would be removed. If the camera lens is covered by dew or if data is lost due to encoding and decoding distortions during image transmission, the image is considered invalid data. Further, there may be errors in the visibility meter measurements due to inconsistencies with the scene; therefore, in order to avoid the interference of visibility-related errors, such abnormal data will also be discarded.

After performing the above processing steps, the data are ready for annotation. The whole annotation process is divided into three stages. In the first stage, according to national standards in China (GB/T 31445-2015 Standard for expressway traffic safety control under fog weather conditions), the visibility levels are divided according to Table 2, where images that are obviously inconsistent with the current levels are deleted. In the second stage, the images are checked at each level, and are deleted or adjusted if they are inconsistent with visibility sorting via comparisons between images. In the third stage, the data within the boundaries of each category is focused to delete those that are obviously ambiguous and cannot be classified accurately. At the end, the results of annotations by all the experts and output data with the same annotation are combined. In all, about 15,000 images (some instances are shown in Fig. 3) are compiled in the dataset.

4.3. Characteristics of the dataset

The size of the image in the final dataset is 1920*1080, and the total number of images in the dataset is approximately 15000. Due to that there are fewer cases of dense fog and heavy fog in the actual environment, the dataset is distributed as follows: clear: 28%, low fog: 20%, medium fog: 20%, high fog: 18% and dense fog: 14%, among them, the distribution of visibility value for each category is uniform. The images comprise data of different seasons (spring, summer, autumn and winter), lighting conditions (from 6 am to 7 pm), and different weathers (sunny, snowy, rainy, foggy and cloudy), along with the overall data distribution. Some examples of the dataset are shown in Fig. 4.

Compared with the existing visibility dataset, our new dataset FVEI has a larger amount of data, and the images are real scenes, which can better represent the differences between the real conditions and object concentration changes in the images. Furthermore, the FVEI dataset not only includes classifications based on visibility levels, but also has more accurate ground truth visibility measurements.

5. Fog image multi visual features

The quality of image features is critical to the final recognition, greatly affecting the accuracy, efficiency, and generalization performance of classification of deep learning network. Therefore, in this section it is necessary to extract and select features that can effectively capture the fog image through comparative analysis. In image recognition problems, a feature is usually only sensitive to changes in certain characteristics of the image, while not sensitive to changes in other characteristics. Therefore, when two types of images have different feature sensitive features, classifiers based on single feature training cannot output correct classification. In addition, the complex background noise in the image will also lead to the degradation of feature data. One way to solve this problem is to use multi feature fusion methods to simultaneously extract multiple features for classifier training to achieve feature complementarity and reduce the impact of inherent defects in a single feature. By analyzing the characteristics of fog images, dark channel and edge features are extracted based on the three channel features of RGB.

Table 2
Correspondence between visibility and fog levels.

Traffic control level	Fog level	Visibility (m)
Null	Clear	$500 \leq V$
Null	Low fog	$200 \leq V < 500$
Fourth level	Medium fog	$100 \leq V < 200$
Third level	High fog	$50 \leq V < 100$
First/Second level	Dense fog	$V \leq 50$

5.1. Dark channel features

The concept of dark channel is defined after statistical analysis of a large number of exterior fog-free images, which means that in most local areas not covered by the sky, some pixels typically have very low intensity values in at least one color (R, G, B) channel. The minimum value of light intensity in this area is very small, approaching zero, and is therefore called a dark pixel (He, SUN, &Tang, 2011). Some image examples as show in Fig. 5 (a-c). The dark-channel I_d is computed by:

$$I_d = \min_{y\in\Omega(x)}(\min_{c\in\{r,g,b\}} I^c(y)) \tag{1}$$

where $I^c(y)$ represents a color channel, and$\Omega(x)$ represents the window at pixel x.

5.2. Edge features

Fog can cause the boundary lines of objects in the image to become more blurred, as shown in some image examples in Fig. 5 (d-f). Therefore, edge features are also one of the important factors to distinguish foggy images and clear images. The sobel operators are used, as shown in Equation (2), to detect the edges of the collected images.

$$\Delta_x f = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \Delta_y f = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2}$$

Each point in the image is convolved using these two operators. $\Delta_x f$ corresponds to the maximum horizontal edge response, and $\Delta_y f$ corresponds to the maximum vertical edge response. The maximum value of the two convolutions is taken as the output point value, as shown in Equation (3).

$$\begin{aligned} s(i,j) = \max(|\Delta_x f|, |\Delta_y f|) = \max(|(f(i-1,j-1) + 2f(i-1,j) + f(i-1,j \\ +1)) - (f(i+1,j-1) + 2f(i+1,j) + f(i+1,j+1))|, |(f(i-1,j-1) \\ + 2f(i,j-1) + f(i+1,j-1)) - (f(i-1,j+1) + 2f(i,j+1) + f(i+1,j \\ +1))|) \end{aligned} \tag{3}$$

where $s(i,j)$ is the output point value, and $f(i,j)$ is the input point value.

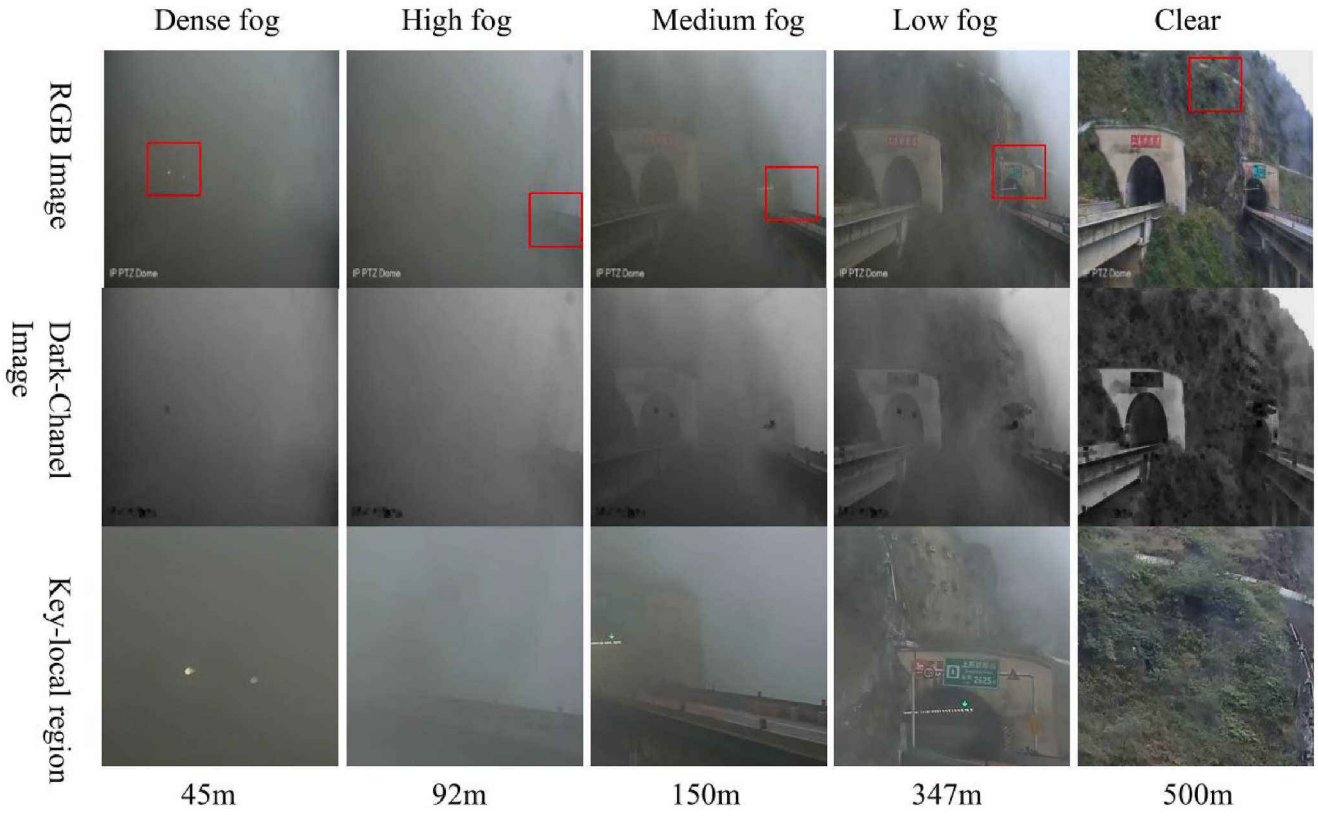

Fig. 3. Differences between dense fog and high fog images. The First line is the raw image in FEVI dataset. The red boxes show the main differences between the two types of images. The Second line is dark channel image. The third line is its corresponding key local region. The last line is the fog visibility value.

Fig. 4. Instances in FVEI dataset.

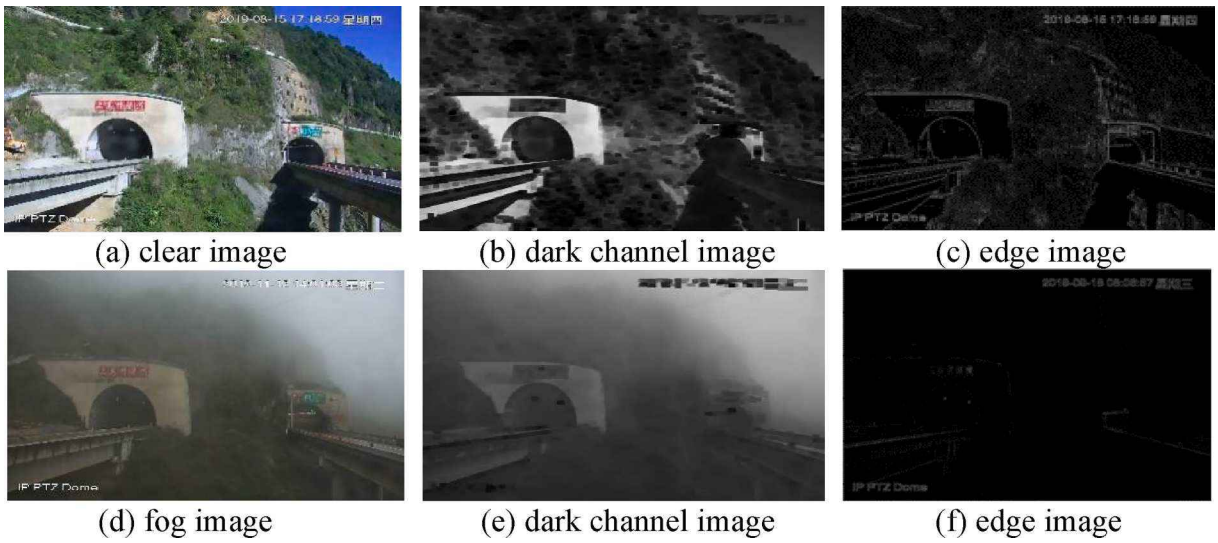

Fig. 5. Examples of fog image features, the first column illustrates the original image, the second column illustrates the corresponding dark channel image, the third column illustrates the corresponding edge image.

6. Methods

In this section, the visibility estimation deep learning network is proposed, namely VENet, which is a cascade multi-network designed to estimate the visibility of a scene accurately from a single image. Given an input image $I_{raw} \in R^{H\times W\times 3}$, VENet aims to predict the fog level and visibility. As shown in Fig. 6, the framework of VENet consists of two modules, namely the feature extractor module and multitask deep learning network cascade module.

6.1. Feature extractor module

A special feature extractor consisting of a global feature extractor and a key local feature extractor are built to capture the discriminative features from a fog-scene image. The details for the construction of the global and local feature extractors are as follows.

6.1.1. Global feature extractor

In fog-scene images, severe image degradation leads to loss of effective information, which makes accurate visibility estimation difficult. To address this problem, a multi-channel integration strategy is adopted by introducing an additional dark channel I_d to the data input, which is obtained by Equation (1).

In addition, it is difficult to extract effective features from the general network in a fog-scene image; therefore, it is necessary to design a specialized network for feature detection. By comparing images with different fog concentration levels, the number of image edges and color saturation significantly decrease as the fog level increases.

Inspired by Qin, Yu, Liu, and Chen (2018) a feature extractor with only four convolutional layers and one fully connected layer are designed. It is worth noting that the information extracted by a shallow CNN mainly includes image edges and colors. Atrous convolution (Chen, Papandreou, Schroff, & Adam, 2017) is thus adopted in the first two layers to increase the receptive field without introducing too many parameters. The whole process is expressed as follows:

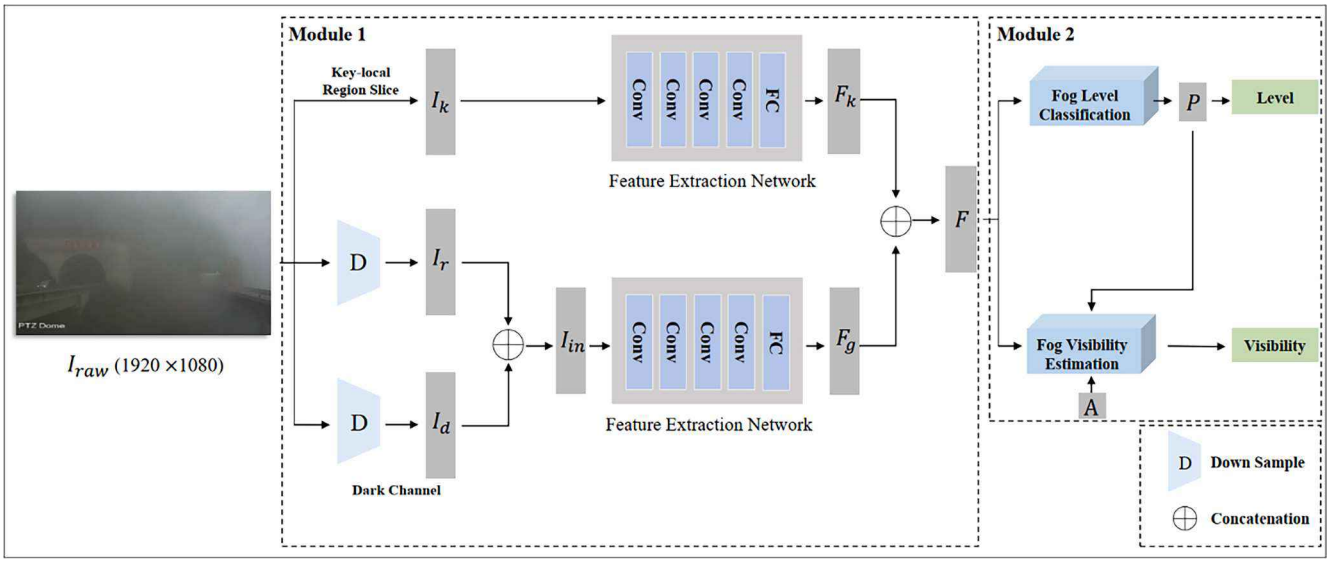

Fig. 6. Framework of VENet. The input is a single image, and the entire framework consists of two modules, namely the feature extractor module (module 1) and multitask cascade module (module 2). In module 1, the global features F_g and key local features F_k are extracted. In module 2, the fog level classification and fog visibility estimation networks are cascaded.

$$I_{in} = I_d \oplus I_r \tag{4}$$

$$F_g = \text{softmax}(\omega_g, I_{in}) \tag{5}$$

where I_r is the RGB three-channel image, I_{in} is the multichannel image after channel fusion, $\oplus$ is the concatenation operation, and ω_g denotes the weights of the feature extraction network.

6.1.2. Key local feature extractor

As the fog level increases, there are only subtle differences among images, as shown in Fig. 7, where the red boxes are the key local regions. The network cannot accurately estimate the visibility by only relying on global features and more information is needed. In this case, the key local features extracted from the key local regions of an image provide the essential details. This idea mainly comes from the data annotation process. When labelling the FVEI dataset, since the human visual system is not sensitive to slight differences in the scene, the annotators often combined regions with obvious texture features for comprehensive judgment. To define key local regions in an image, the local regions with the highest densities of edge points are considered, since regions with high densities of edge points often tend to contain more texture information.

Based on this definition, the key local feature extractor is built. First, the key local region I_k is obtained using Algorithm 1. Then, it is inserted into the feature extraction network, which has the same structure as the global feature extractor. The key local feature extractor is calculated by:

$$F_k = \text{softmax}(\omega_k, I_k) \tag{6}$$

where I_k denotes the key local region, ω_k denote the network parameters, and F_k is a feature vector. The fused feature F is obtained by concatenating F_g and F_k as:

$$F = F_k \oplus F_g \tag{7}$$

Algorithm 1: *The capture of key local region in a fog-scene image.*

Input:
The original fog-scene image, I_{raw};
The target size of the key local region, $S \times S$;
Output:
The key local region image, I_k;
1: Convert RGB image I_{raw} to Gray image I_{gray};
2: Convert I_{gray} to edge image I_{edge} using sobel operator;
3: Let $p_x \leftarrow 0$, $p_y \leftarrow 0$, $W \leftarrow 1920$, $H \leftarrow 1080$;
4: $R = [R_0, R_1, \cdots R_{W-1}], R_i = \sum_{j=0}^{H-1} I_{edge}(i,j)$;
5: $C = [C_0, C_1, \cdots C_{H-1}], C_j = \sum_{i=0}^{W-1} I_{edge}(i,j)$;
6: $P_x = \underset{x}{\text{argmax}}(\sum_{i=x}^{x+S} R_i)$ and $x \in [0, W-S] \wedge x \in Z$;
7: $P_y = \underset{y}{\text{argmax}}(\sum_{j=y}^{y+S} C_j)$ and $y \in [0, H-S] \wedge y \in Z$;
8: Cropping I_{raw} based on the left high position (p_x, p_y) and size S, obtain I_k;
9: return I_k;

6.2. Multitask network cascade

Fog level classifications and visibility estimation are highly related tasks. Essentially, fog level classification is a form of visibility estimation with lower accuracy. Therefore, a multitask deep learning network cascade module based on the idea of coarse-to-fine refinement is constructed. Fog visibility estimation relies on the results of fog level classifications, whose framework is illustrated in Fig. 8, and the relevant parameters in the figure are explained in the latter section.

6.2.1. Fog level classification network

First, a simple network is built consisting of two fully connected layers, and the exponential linear unit (ELU) (Clevert, Unterthiner, & Hochreiter, 2015) is adopted as the activation function. The network outputs a vector P, and the network can be denoted as:

$$P = \text{softmax}(\omega_{cls}, F) = [P_0, P_1, \cdots, P_{N-1}]^T \tag{8}$$

where P is a vector with N elements, and each element P_i represents the probability of the corresponding fog level i. And N is the number of fog levels, ω_{cls} represents all the classification network parameters to be optimized, and the fog level V_f is the maximum index of P given by:

$$V_f = \underset{i}{\text{argmax}} P_i \tag{9}$$

6.2.2. Visibility value estimation network

In the second stage, the fused convolutional features F and stage-1 probability P are used as input, and the fog visibility estimation value is output. When a neural network is applied to solve the regression problem, it requires a large amount of evenly distributed data as support. However, the amount of visibility data is limited in our study, which might lead to over fitting or under-fitting. Further, as shown in Fig. 9, the distribution of the visibility data is uneven, which will affect the convergence of the training process.

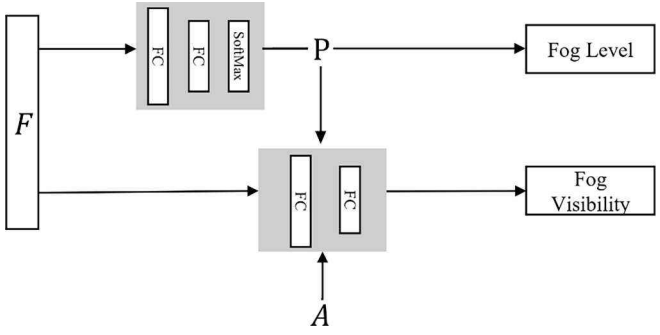

Fig. 8. Structure of the multitask network cascade module.

Fig. 7. Differences between dense fog and high fog images. The red boxes show the main difference between the two types of images.

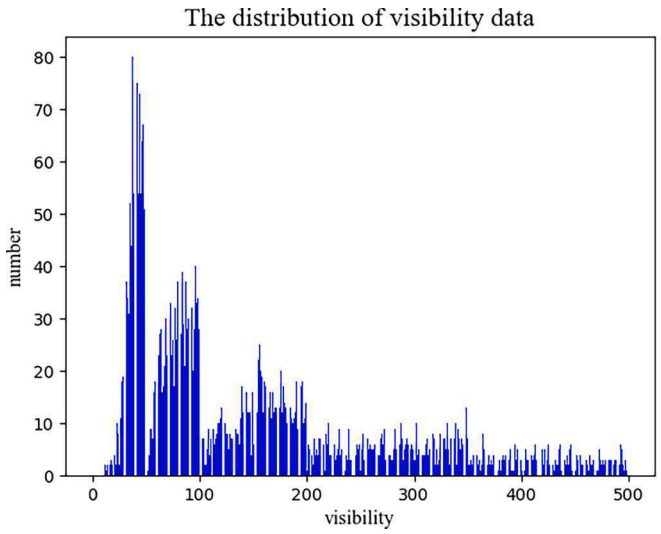

Fig. 9. Histogram of the visibility data distribution.

Inspired by the research on target detection (Dai, He, & Sun, 2016), an anchor-based regression method is proposed (ARM), by introducing additional anchor data whereby the regression problem of a single value is transformed into the estimation problem of multiple coefficients. At the same time, the introduction of anchors increases the constraints for model training, which allows the model to converge more easily. The procedure of this method is as follows.

First, the mean values of the visibility data for each fog level class are obtained and consider these mean values as the anchors A:

$$A = [A_0, A_1, \cdots, A_{N\text{-}1}] \tag{10}$$

$$A_i = \frac{1}{n\text{-}1}\sum_{j=0}^{n\text{-}1} v_i^j \tag{11}$$

where A_i is the *ith* anchor data, n is the number of images in the fog level i, and v_i^j is the raw visibility data at the fog level i. In particular, all these data are obtained from the training set.

Then, the same network used for classification is adopted to obtain the coefficients $E = \{E_0, E_1, E_2, E_3, E_4\}$ for each anchor data.

Finally, the classification results P and coefficients E are multiplied in a pointwise manner with the anchors A. The final estimation result is computed by

$$V_d = P{\cdot}A \times E = \sum_{i=0}^{N} P_i{\cdot}A_i{\cdot}E_i \tag{12}$$

where V_d is the result from visibility estimation.

6.3. Loss function

In this section, the multitask deep learning network cascade module is presented. To support end-to-end training, the loss function of the entire cascade is defined as:

$$L = \gamma * L_{cls}(P) + (1-\gamma) * L_{reg}(V_d|P) \tag{13}$$

This loss function contains two parts: classification loss L_{cls} and regression loss L_{reg}, while γ is a hyper-parameter that is set to 0.25. For the first part, a normal cross-entropy loss is adopted, which can be calculated as:

$$L_{cls}(P) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k}\log(P_{ij}[lable]) \tag{14}$$

where m is the number of samples, k is the number of the fog levels, and $P_{ij}[lable]$ represents the probability of the ground truth.

For the second part, the visibility estimation results depend on the output of the classification, which might pose a challenge for back-propagation. Here the classification results to visibility estimation through pointwise multiplication is fused, and the calculation process is derivable, which ensures backpropagation. The expression of L_{reg} is:

$$L_{reg}(V_d|P) = \frac{1}{m}\sum(V_d - V_d^{'})^2 \tag{15}$$

where $V_d^{'}$ is the ground truth fog visibility value.

7. Experiments

In this section, the validity of the deep learning network model is verified and an application system is built for practical application at expressways.

7.1. Implementation details

The experimental setup consists of a single computer with two Intel Xeon-E5 CPUs, a 4 × TITAN RTX GPU, and 64 GB RAM. The models were implemented using Pytorch-1.4.1 and CUDA-10.0.

For model training, an end-to-end process is applied where the model is trained using a learning rate of 0.001 for 100 iterations. The stochastic gradient descent (SGD) method is adopted to optimize the model. The weight decay is set to 0.0001, momentum is set to 0.9, batch size is 32, and hyper parameter γ is set to 0.25. The FVEI dataset is divided into 70%, 10%, and 20% sets for training, validation, and testing, respectively. The raw images are downsampled to sizes of 448 × 448. The size of the key local region is also set as 448 × 448.

The performance of fog level classification is evaluated based on the precision and recall rates. For fog visibility estimation, the absolute error is calculated by:

$$error = \frac{|y^{'} - y|}{500} \tag{16}$$

where $y^{'}$ and y are respectively the regressed value and ground truth, while the range of visibility value is 0–500 m.

7.2. Effectiveness of the feature extractor module

To verify the effectiveness of the proposed feature extractor module, the proposed model is compared with the following models: AlexNet (Krizhevsky, Sutskever, & Hinton, 2017), VGG16, VGG19 (Simonyan, & Zisserman, 2014), ResNet-50, ResNet-101, ResNet-150 (He, Zhang, Ren, & Sun, 2016), and Vision transformer (Dosovitskiy et al., 2020). It is worth noting that all these models are used as feature extractors that only output a single feature vector; then the feature vector (1280D) is processed using a multitask network cascade module. The comparison

Table 3
Comparison of the feature extraction abilities of AlexNet, VGG16, VGG19, ResNet-50, ResNet-101, ResNet-152, Vision transformer, VENet without key local features (VENet-NK), and VENet.

Model	Precision (%)	Recall (%)	Error (%)
AlexNet	88.3	87.9	6.9
VGG16	89.1	88.4	6.4
VGG19	88.0	86.6	7.3
ResNet-50	86.1	86.0	7.5
ResNet-101	87.3	86.1	7.0
ResNet-152	86.6	86.5	7.4
Vision transformer	86.2	86.1	7.2
VENet-NK	91.7	91.2	6.0
VENet	92.3	90.8	4.6

results are shown in Table 3.

In addition, an ablation experiment is conducted where the performances of the feature extractors with and without key local features are compared, and the recognition accuracy of different fog levels in detail are also compared in the confusion matrix, as shown in Fig. 10. In the confusion matrix, the horizontal axis is the ground truth, namely, the true fog level classification, and the vertical axis is the predicted fog level classification.

From Fig. 10, the following observations can be obtained: (1) Compared with other models, the proposed VENet achieves better performance for visibility estimation with at least 2.2% and 2.4% improvement in accuracy and recall, and at least 1.8% reduction in absolute error. (2) In the ablation experiment, the feature extractor with added key local features shows better accuracy in visibility estimation, and the error is reduced by 1.4%. It is also found from the confusion matrix that by adding key local feature, the recognition accuracy of high fog and mid fog are highly improved. The results promisingly demonstrate that the proposed feature extractor has advantages in capturing effective features from the fog-scene images, and adding key local features could help capture subtle differences between images.

7.3. Effectiveness of ARM

In this subsection, the effectiveness of ARM in improving the accuracy of visibility estimation is verified. A baseline model is constructed with the same network structure as VENet but without adopting ARM (VENet-NA). And the training process and test results of these two models are compared.

In the training process (Fig. 11), VENet has a greater convergence speed and better convergence effect than VENet-NA. Then, 200 samples are randomly selected for testing, as shown in Fig. 12. From the test results, it can see that the accuracy and stability of VENet are clearly better than those of VENet-NA. The reasons for the significant improvement with ARM include the following: 1) The introduction of anchors can provide a better initial solution for the regression model, 2) Anchors contain some prior information that can help the model obtain accurate estimations.

7.4. Effectiveness of the multitask network cascade module

This subsection aims to answer the following question: does the cascade structure help improve model performance? Three models for comparison with VENet are set as follows: 1) a single task network for fog level classification (FLC), 2) a single task network for fog visibility estimation (FVE), and 3) a multitask network without a cascade structure (VENet-NC). Table 4 illustrates the results of the comparison among the four models. It can be seen that the proposed cascade structure has the best performance in fog visibility estimation. It is also noted that these models have similar performance in the classification of fog level. It is feasible to build a multitask network cascade to achieve a coarse-to-fine process from the fog level to visibility.

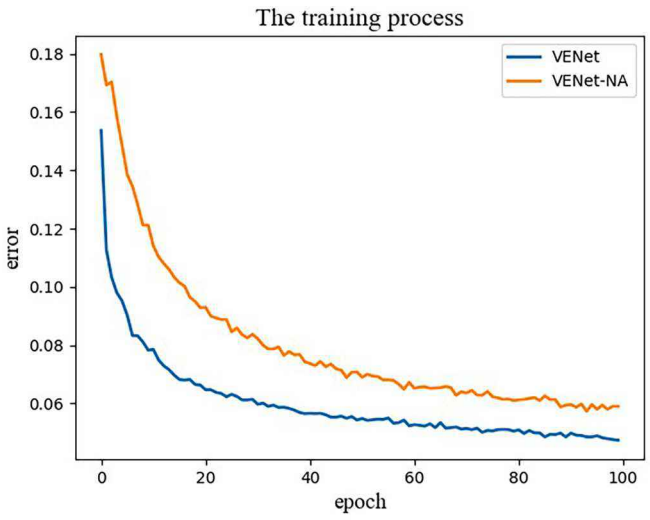

Fig. 11. Comparison of the training processes between VENet-NA and VENet; in particular, the ordinate represents the absolute error calculated by (16).

7.5. Engineering application and verification

In Yunnan Province, the Xiangli Expressway is routed along high altitude ridges, and along the line, there are abundant water systems such as the Jinsha River and Chongjiang River. During rainy seasons in winter or summer, due to differences in the distribution of temperature and humidity inside and outside the tunnels, there are frequent fogs in the short connection sections between tunnels and the tunnel entrance

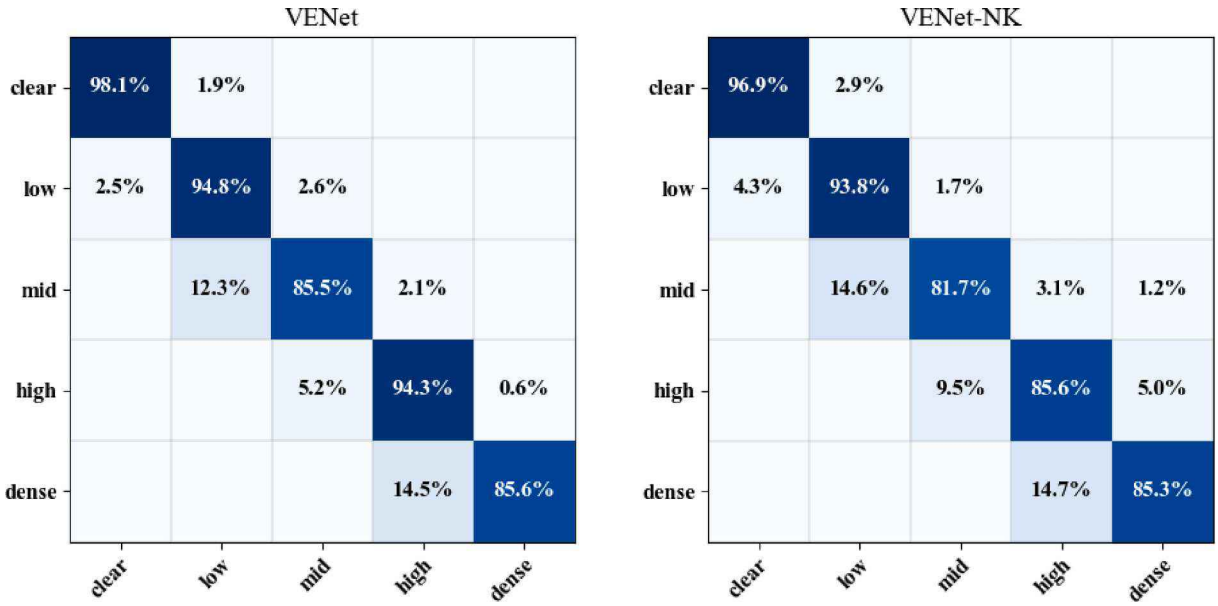

Fig. 10. The confusion matrix of VENet with key local feature (VENet) and VENet without key local feature (VENet-NK).

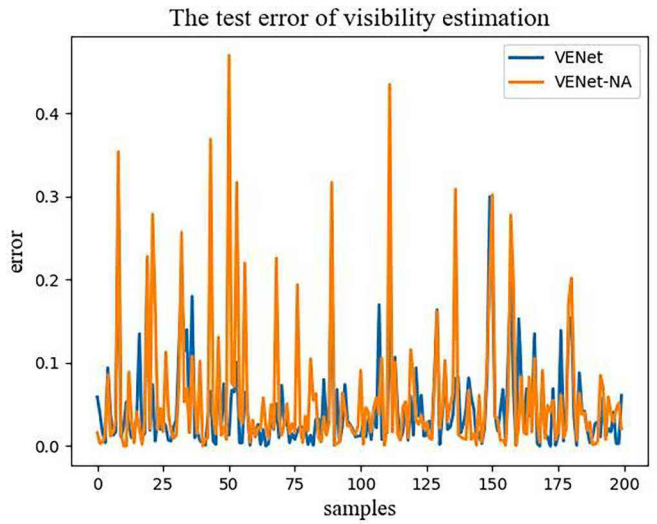

Fig. 12. Comparison of the test results between VENet and VENet-NA, where the ordinate represents the absolute error, and the abscissa represents the sample number.

Table 4
Performance comparison among the single task networks (FLC, FVE), multitask network (VENet-NC), and multitask network cascade (VENet).

Model	Precision (%)	Recall (%)	Error (%)
FLC	91.6	90.3	–
FVE	–	–	6.0
VENet-NC	92.1	91.3	6.3
VENet	92.3	90.8	4.6

and exit area, leading to prominent traffic safety issues. To address this issue, the technology proposed in the paper has been applied and deployed at the entrance and exit of two tunnels of Xiangli Expressway surveillance (see Fig. 2: two images on the left), and the detection accuracy of fog visibility has reached over 90%. Fig. 13 is the real-time visibility analysis and warning system interface based on video images for Yunnan Expressway, the system can display real-time tunnel entrance and exit scene images per minute, and output fog levels and visibility values.

The system has been in operation since June 2021, and has been connected to the identification data of the visibility sensors. When there is a significant error between the system identification results and the visibility sensors identification results, manual verification is conducted. Fig. 14 is a comparison diagram of the system, visibility sensors, and manual verification results for 8000 pieces of data. Based on the results of manual verification, the accuracy rate of system recognition reached 92.1%. Due to the good operation of the system, the proposed VENet has been popularized and applied at the entrance and exit of a tunnel on Yunnan Mazhao Expressway surveillance (see Fig. 2: the image on the right)

8. Conclusion

The advantage of the proposed approach compared to similar schemes are threefold, firstly, a deep learning based multi visual feature fusion network is proposed, named VENet for fog visibility estimation from a single image. Secondly, a multitask deep learning network cascade is constructed, consisting of a fog level classification network and a fog visibility estimation network. In particular, an anchor-based regression method is proposed that can help the network achieve fast convergence and accurate predictions. A special feature extractor is also introduced to obtain the discriminative features from fog-scene images. Thirdly, it is also worth emphasizing that a standard Fog Visibility Estimation Image (FVEI) dataset is constructed, which greatly compensates for the lack of data in the field of vision-based visibility estimation and can provide significant support for future research. The results of extensive experiments have demonstrated that the proposed VENet can achieve excellent performance for both fog level classification and fog visibility estimation. In addition, the proposed VENet has been applied on Yunnan Xiangli and Mazhao Expressway surveillance,

Fig. 13. The real-time visibility analysis and warning system interface based on video images for Yunnan Expressway.

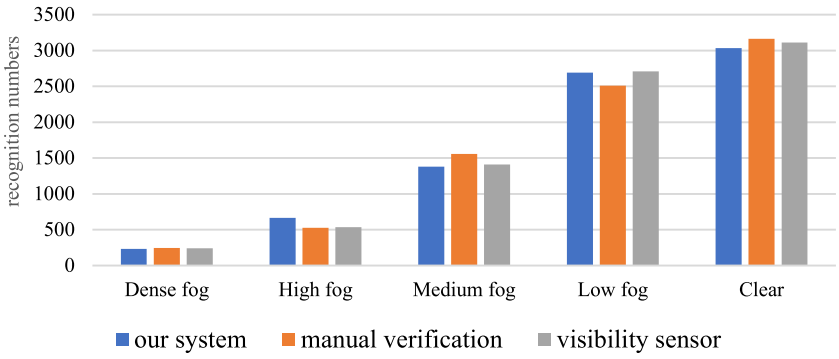

Fig. 14. Comparison of the system, visibility sensors, and manual verification results for 8000 pieces of data.

and has achieved promising application results.

More interesting future work can be extended from this study such as the incorporation of meta learning and transfer learning into the feature fusion task in the proposed VENet. For now, though the proposed model is able to achieve a high estimation accuracy on the existing data, a limitation of the proposed approach is that there may be overfitting issues. One way to solve the issue is to optimize the model from the perspective of improving the robustness of the algorithm and exploring more ways to enhance the data (for example, through the GAN (Generative Adversarial Networks) to generate more fog-scene images). It is pertinent to expand the data set to include a wider variety of scenarios.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (Grant No. 2022YFC3002601), the General Program of Key Science and Technology in transportation, the Ministry of Transport, PRC (Grant No. 2018-MS4-102 & ZL-2018-04), the Science and Technology Demonstration project of Ministry of Transport, PRC (Grant No. 2017-09), the Science and Technology Innovation Program of the Department of Transportation, Yunnan province, China (No. 2021-6 and 2020-75), and Yunnan Key Laboratory of Digital Communications (grant NO. 202205AG070008).

References

Asery, R., Sunkaria, R. K., Sharma, L. D., & Kumar, A. (2016). June). Fog detection using GLCM based features and SVM. In *2016 Conference on Advances in Signal Processing*. https://doi.org/10.1109/CASP.2016.7746209

Bronte, S., Bergasa, L. M., & Alcantarilla, P. F. (2009, November). Fog detection system based on computer vision techniques. *In 2009 12th International IEEE conference on intelligent transportation systems*, Missouri.

Branwen, G., & Gokaslan, A. (2019). Danbooru2019: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset.

Belaroussi, R. & Gruyer, D. (2014, June). Impact of reduced visibility from fog on traffic sign, In IEEE Intelligent Vehicles Symposium Dearborn, Michigan.

Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009, 09796*. https://doi.org/10.48550/arXiv.2009.09796.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., & Houlsby, N. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv: 2010.11929*. https://doi.org/10.48550/arXiv.2010.11929.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Li, F. F. (2009, June). Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition*, Florida.

Dai, J., He, K., & Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. IEEE *Conference on Computer Vision and Pattern Recognition*, Las Vegas. https://doi.org/10.1109/CVPR.2016.343.

Georghiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(6), 643–660. https://doi.org/10.1109/34.927464

Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset.

Hautiere, N., Tarel, J. P., Lavenant, J., & Aubert, D. (2006). Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine vision and applications, 17*(1), 8–20. https://doi.org/10.1007/s00138-005-0011-1

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998, April). Coding facial expressions with gabor wavelets. *In Proceedings Third IEEE international conference on automatic face and gesture recognition*, Nara.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, C., Lu, X., Tong, C., & Zeng, W. (2014, December). A fog level detection method based on grayscale features. *In 2014 Seventh International Symposium on Computational Intelligence and Design*, Hangzhou.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Zitnick, C. L. (2014, *September*). Microsoft coco: Common objects in context. *In Computer Vision–ECCV 2014: 13th European Conference,* Switzerland.

Pavlić, M., Belzner, H., Rigoll, G., & Ilić, S. (2012). Image based fog detection in vehicles. In *In 2012 IEEE Intelligent Vehicles Symposium*. https://doi.org/10.1109/IVS.2012.6232256

Palvanov, A., & Cho, Y. I. (2019). Visnet: Deep convolutional neural networks for forecasting atmospheric visibility. *Sensors, 19*(6), 1343.

Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., & Mura, A. Y. N. (2018). Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*. https://doi.org/10.48550/arXiv.1712.06957.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556.

Tarel, J.-P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H., & Gruyer, D. (2012). Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems, 4*(2), 6–20.

Tarel, J.-P., Hautière, N., Cord, A., Gruyer, D. & Halmaoui, H. (2010, June). Improved visibility of road scene images under heterogeneous Fog. *In Proceedings of IEEE Intelligent Vehicles Symposium (IV'10), San Diego, CA*.

Teichmann123, M., Weber, M., Zöllner, M., Cipolla, R., & Urtasun, R. (2018 October). MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving. *In 2018 IEEE Intelligent Vehicles Symposium*, Suzhou.

Traffic Administration Bureau of the Ministry of Public (2021). Statistical annual report of road traffic accidents of the People's Republic of China (2016-2020).

Wang, Y., Jia, L., Li, X., Lu, Y., & Hua, D. (2020). A measurement method for slant visibility with slant path scattered radiance correction by lidar and the SBDART model. *Optics Express, 29*(2), 837–853. https://doi.org/10.1364/OE.409309

Xian, J., Han, Y., Huang, S., Sun, D., & Li, X. (2018). Novel lidar algorithm for horizontal visibility measurement and sea fog monitoring. *Optics Express, 26*(2), 34853–34863. https://doi.org/10.1364/OE.26.034853

Xie, K., Huang, L., Zhang, W., Qin, Q., & Lyu, L. (2022). A CNN-based multi-task framework for weather recognition with multi-scale weather cues. *Expert Systems with Applications, 198*, Article 116689. https://doi.org/10.1016/j.eswa.2022.116689

Xue, Q. W., Xu, J. W., & Du, Z. G. (2022). A Study on the correlation between vehicle control behaviors and rear-end collision risk under foggy conditions. *Journal of Transport Information and Safety, 40*(1), 19–27.

You, Y., Lu, C., Wang, W., & Tang, C.-K. (2018). Relative cnn-rnn: Learning relative atmospheric visibility from images. *IEEE Transactions on Image Processing, 28*, 45–55. https://doi.org/10.1109/TIP.2018.2857219

You, J., Jia, S., Pei, X., & Yao, D. (2022). DMRVisNet: Deep multihead regression network for pixel-wise visibility estimation under foggy weather. *IEEE Transactions on Intelligent Transportation Systems, 23*(11), 22354–22366. https://doi.org/10.48550/arXiv.2112.04278

Further reading

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*. https://doi.org/10.48550/arXiv.1706.05587.

Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*. https://doi.org/10.48550/arXiv.1511.07289.

He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transaction Pattern Analysis and Machine Intelligence, 33*(12), 2341. https://doi.org/10.1109/TPAMI.2010.168

Qin, Z., Yu, F., Liu, C., & Chen, X. (2018). How convolutional neural network see the world-A survey of convolutional neural network visualization methods. *arXiv preprint arXiv:1804.11191*. https://doi.org/10.48550/arXiv.1804.11191.