Contents lists available at ScienceDirect

Pattern Recognition

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

Restoring vision in hazy weather with hierarchical contrastive learning

Tao Wang [a], Guangpin Tao [a], Wanglong Lu [b], Kaihao Zhang [c], Wenhan Luo [d], Xiaoqin Zhang [e], Tong Lu [a,*]

[a] *National Key Laboratory for Novel Software Technology, Nanjing University, China*
[b] *Memorial University of Newfoundland, Canada*
[c] *Australian National University, Australia*
[d] *Sun Yat-sen University, China*
[e] *Wenzhou University, China*

ARTICLE INFO

Keywords:
Image dehazing
Hierarchical contrastive loss
Feature fusion
Contrastive learning

ABSTRACT

Image restoration under hazy weather condition, which is called single image dehazing, has been of significant interest for various computer vision applications. In recent years, deep learning-based methods have achieved success. However, existing image dehazing methods typically neglect the hierarchy of features in the neural network and fail to exploit their relationships fully. To this end, we propose an effective image dehazing method named Hierarchical Contrastive Dehazing (HCD), which is based on feature fusion and contrastive learning strategies. HCD consists of a hierarchical dehazing network (HDN) and a novel hierarchical contrastive loss (HCL). Specifically, the core design in the HDN is a hierarchical interaction module, which utilizes multi-scale activation to revise the feature responses hierarchically. To cooperate with the training of HDN, we propose HCL which performs contrastive learning on hierarchically paired exemplars, facilitating haze removal. Extensive experiments on public datasets, RESIDE, HazeRD, and DENSE-HAZE, demonstrate that HCD quantitatively outperforms the state-of-the-art methods in terms of PSNR, SSIM and achieves better visual quality.

1. Introduction

Single image dehazing aims to recover the latent haze-free image from a given hazy image. Due to its wide range of applications (*e.g.,* autonomous driving and video surveillance), single image dehazing has become a hot topic in the fields of computer vision and image processing.

Traditional image dehazing methods [1] are mostly based on the image prior and the atmosphere scattering model (ASM) [2]. Specifically, the pipeline of these methods is to find some prior information from images to estimate the transmission map t and global atmosphere light A from the haze image I, and then to use the predicted t and A to recover the clear image J according to ASM as $J(x) = (I(x) - A)/t(x) + A$, where x is the pixel position. Unfortunately, traditional methods usually require time-consuming iteration optimization and handcrafted priors. Thus they may not work well in complex haze scenarios.

In recent years, with the rapid development of deep learning techniques and the collection of large-scale synthetic datasets, many data-driven image dehazing approaches have been proposed to achieve haze removal. In the beginning, many works like [3,4] attempt to estimate the transmission map and the atmospheric light through Convolution Neural Networks (CNNs) and then restore the clear image via ASM. However, the inaccurate estimation of the transmission map or atmospheric light may easily lead to their poor dehazing performance. More recently, another class of data-driven approaches [5,6] directly ignores ASM and uses an end-to-end CNN to learn a mapping between the hazy image and the clear image. For example, Jiang et al. [7] design an end-to-end network containing a haze residual attention sub-network and a detail refinement sub-network to directly recover the clear image from hazy input. The Attention mechanism [8,9] is embedded into the end-to-end network for effective image dehazing. Even though the above data-driven methods greatly improve the visual quality of dehazed results, they share the following drawbacks: (1) *They do not fully exploit hierarchical features in CNNs.* As we know, shallow features of CNNs contain more details and spatial information, while deep features focus on higher-level context and semantic information [10]. Both shallow and deep features of CNNs are beneficial for the process of image dehazing. However, existing methods [11,12] do not fully exploit complementary information from these hierarchical features in CNNs, and it is easy to cause color distortion in the recovered images [13]. (2) *They only consider positive-oriented supervision information in the training stage.* Most

[*] Corresponding author.
E-mail address: lutong@nju.edu.cn (T. Lu).

https://doi.org/10.1016/j.patcog.2023.109956
Received 31 January 2023; Received in revised form 17 August 2023; Accepted 6 September 2023
Available online 16 September 2023

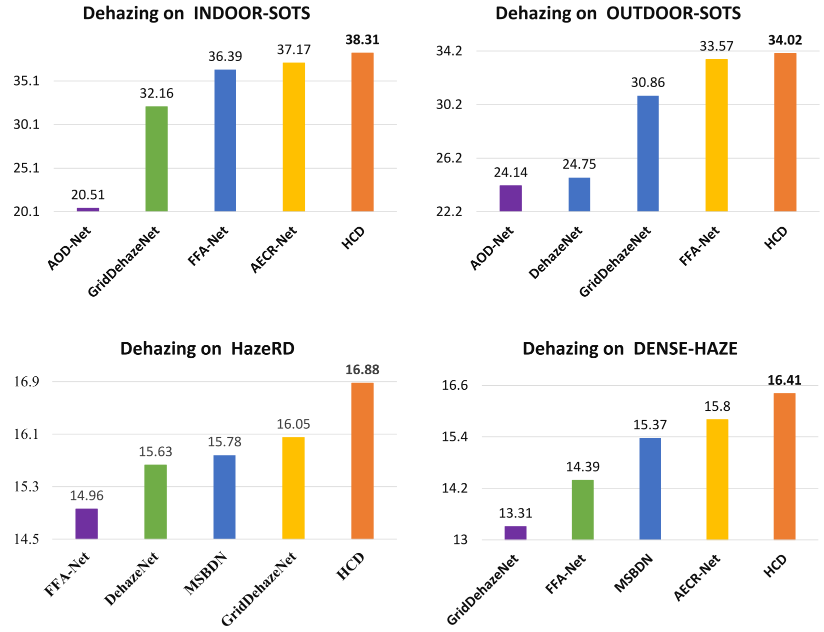


Fig. 1. Performance comparison of the proposed HCD with the state-of-the-art methods on popular image dehazing datasets. Our HCD significantly advances the state-of-the-art image dehazing performance in terms of PSNR. (1) + **1.14** dB on the indoor subset of Synthetic Objective Testing Set (SOTS) [15], (2) + **0.45** dB on the outdoor subset of SOTS [15], (3) + **0.83** dB on the HazeRD dataset [16], and (4) + **0.61** dB on the DENSE-HAZE dataset [16].

data-driven image dehazing methods typically regard haze-free images as positive samples to guide the optimization of the model and do not fully mine the hazy input images (negative samples) in the training stage. Ignoring negative-oriented learning reduces the representation ability of the model to some extent, which results in lower image restoration performance of the model [14].

To address the problems mentioned above, we aim to design an effective feature fusion scheme in the network and a positive and negative-oriented supervision strategy to further supervise the network training. To this end, we propose a novel dehazing method called Hierarchical Contrastive Dehazing (HCD). The proposed HCD consists of a hierarchical dehazing network (HDN) and a hierarchical contrastive loss (HCL). Specifically, HDN includes a hierarchical feature extractor, a hierarchical interaction module (HIM), and a multi-output image reconstruction module. The hierarchical feature extractor extracts hierarchical features (*i.e.,* multi-resolution features) from a hazy input. The features are then fed into the HIM, which allows information flow to exchange efficiently across different branches and improves the dehazing performance. This module first progressively propagates higher-level features into the shallow layers to suppress the noise in lower-level features and then incorporates the detailed information from lower layers into the deep layers. After that, the multi-output image reconstruction module assembles features with three resolutions and reconstructs clean images. Finally, HCL is embedded into our special hierarchical structure dehazing network. It guides the network to exploit a representation by maximizing similarity and dissimilarity over samples that are organized into similar and dissimilar pairs in a hierarchical manner.

The closest image dehazing methods with our proposed HCD are GridDehazeNet [17] and AECR-Net [14]. However, our HCD differs from GridDehazeNet and AECR-Net in several ways. First, our proposed HFB refines features by using differences between features in different branches and hierarchically propagating information from the bottom to the top branch, allowing it to effectively utilize the hierarchical features in the network. In contrast, GridDehazeNet and AECR-Net use channel-wise attention and skip connections, which cannot fully utilize hierarchical features in the network. Second, our proposed hierarchical contrastive loss is applied at multiple scales using negative and positive samples to enhance the feature representation ability of the network. However, GridDehazeNet only considers the information of positive images as an upper bound, and AECR-Net only performs contrastive learning on a single image scale. By employing a hierarchical structure for contrastive learning, we further enhance the network's feature representation ability. Finally, our extensive experiments demonstrate that the proposed HCD significantly outperforms state-of-the-art image dehazing methods, including GridDehazeNet and AECR-Net, as illustrated in Fig. 1.

To summarize, the contributions of our work are as follows:

- We propose a novel Hierarchical Contrastive Dehazing (HCD) method, which employs a hierarchical feature fusion technique and a contrastive learning strategy to effectively enhance the feature representation ability of the model.
- The implementation of a hierarchical interaction module in a hierarchical structure network allows information flow to exchange efficiently across different branches and improves the dehazing performance.
- By considering both positive and negative-oriented supervision, the proposed hierarchical contrastive loss effectively guides the model to learn the valid features for the image dehazing.
- Extensive experimental results on benchmarks demonstrate that HCD performs favorably against state-of-the-art approaches.

The remainder of this paper is organized as follows. Section 2 presents the related work. Section 3 introduces our proposed method. Section 4 reports experimental results. Section 5 provides a conclusion of this paper.

2. Related work

The proposed method is related to image dehazing and contrastive learning, which are reviewed in the following.

2.1. Single image dehazing

Image dehazing aims to recover a clear image from the hazy image, which is a popular research topic in the computer vision community.

A wide range of methods has been proposed in the literature to address this problem. They are approximately categorized into traditional prior-based methods and deep learning-based methods.

Traditional prior-based methods mainly focus on exploring the statistical properties of images (*i.e.,* statistical prior) to estimate the atmospheric light and transmission map and then recover the clear image by ASM [18]. Tan [2] proposes an image dehazing method by maximizing the local contrast of hazy images, which is based on the statistical observation that clear images have more contrast than hazy images. He et al. [19] propose a haze removal approach utilizing the dark channel prior. This prior is motivated by the assumption that the dark channels of clear images are close to zero. In [20], the color-line prior is employed to achieve image dehazing. The color-line prior hypothesizes that pixels in small image patches have the characteristic of a one-dimensional distribution in RGB space. Zhu et al. [21] employ a linear model to estimate the depth information of images based on the color attenuation prior for image haze removal. Yuan et al. [22] propose a new unified framework for image dehazing, which aims to effectively use several existing priors to obtain clear images from the hazy image. Although the prior-based methods have achieved impressive results, the representation ability of these hand-crafted priors is limited, especially for highly complex hazy scenes.

In recent years, with the rapid development of deep learning, deep learning-based fog removal methods have been extensively studied. Deep learning-based dehazing methods can be approximately divided into supervised methods, semi-supervised methods, and unsupervised methods [23]. For supervised dehazing methods, some methods are designed based on ASM model. For instance, Cai et al. [4] first employ a convolutional neural network to estimate the transmission map from hazy images and then restore dehazed images based on the atmospheric scattering model. Li et al. [3] first reformulate the scattering model, which translates the problem of estimating the transmission and atmospheric light into estimating an intermediate parameter. Then, based on the re-formulated scattering model, they propose an AOD-Net model to achieve image dehazing. On the other hand, some approaches directly learn the mapping from hazy and clear images through convolutional neural networks to achieve image dehazing. For example, Qin et al. [11] design a deep FFA-Net for the dehazing task. The core component in the FFA-Net is the feature attention module, which includes a pixel attention block, a channel attention block, and a residual operation. In [24], Yin et al. propose a parallel attention network for image dehazing, which mainly designs a parallel spatial/channel-wise attention block to capture more informative spatial and channel-wise features. As for semi-supervised dehazing methods, representative approaches such as PSD [25] and SSDT [26] first utilize a backbone network for pre-training purposes to acquire a basic network that fits synthetic data. Then, the unsupervised fine-tuning process on real-world domains is applied to the network to improve the capability of the network to deal with real word hazy images. The supervised and semi-supervised dehazing methods rely on paired data in the training process, which limits their application. Therefore, some unsupervised dehazing methods are proposed. For example, inspired by CycleGAN [27], Cycle-Dehaze [28] and CDNet [29] achieve image dehazing by unsupervised domain translation.

2.2. Contrastive learning

Recently, contrastive learning has been widely used in self-supervised representation learning. The goal of contrastive learning is to learn an invariant representation from the data in the training dataset. The key step of the contrastive learning technique is to design an effective strategy to maximize the complementary information over data samples. Some contrastive learning methods improve representation ability by designing a contrastive loss, such as triplet loss and InfoNCE loss. The contrastive loss is used to push an exemplar close to similar samples while pushing it far away from dissimilar samples. For example, Park et al. [30] use the InfoNCE loss to train the network for unpaired image-to-image translation and demonstrate that contrastive learning techniques can significantly improve the performance of models in conditional image synthesis tasks. With the rapid development of contrastive learning techniques, several low-level vision tasks have employed contrastive loss and achieved promising performance. Zhang et al. [31] employ the contrastive learning technique to solve blind super-resolution in real-world scenery. They design contrastive decoupling encoding for learning resolution-invariant features and use these learned features to obtain high-resolution images. Recently, Wu et al. [14] propose a novel pixel-wise contrastive loss and regard it as a regularization term to train a network for image dehazing. Unlike Wu et al. [14] only perform contrastive learning on a single image scale, we employ a hierarchical structure for contrastive learning to further enhance the network's feature representation ability, resulting in a significant improvement in haze removal performance.

3. Proposed method

In this section, we first introduce an overview of the proposed HCD and then detail each component within it respectively. The loss function to optimize the network is introduced in the end.

3.1. Method overview

We propose the HCD method for the image dehazing task, which can fully exploit hierarchical representation from the hazy image. The overall architecture of HCD is illustrated in Fig. 2. Our HCD includes a hierarchical dehazing network (HDN) and a hierarchical contrastive loss (HCL). HDN is used to dehaze the input image, and HCL utilizes the information of positive and negative samples to guide the HDN training. These two components are working together to produce a good performance in image dehazing. Specifically, as shown in Fig. 2, given a hazy input image, a hierarchical feature extractor (HFE) first extracts hierarchical visual features, then a hierarchical interaction module (HIM) fuses these hierarchical features alternately and hierarchically. After that, a multi-output image reconstruction module (MOIRM) reconstructs the output features of HIM and generates different multi-scale dehazed images. Finally, HDN performs contrastive learning in a hierarchical manner via HCL to further improve the feature learning ability. In the following subsections, we detail each component of our HCD, *i.e.,* HDN and HCL.

3.2. Hierarchical dehazing network

Hierarchical Feature Extractor. Feature representation plays an essential role in the computer vision community, and the representation directly affects the performance of a deep learning method [32]. As discussed in [32], a good feature representation should have the following characteristics. One is that it can capture multiple configurations from the input. Another is that it should organize the explanatory factors of the input data as a hierarchy, where more abstract concepts are at a higher level. To this end, we propose a hierarchical feature extractor (HFE) in the network to effectively extract features from a hazy input image.

HFE module is designed to extract multi-scale and hierarchical features from input images, which are then used for subsequent feature fusion and dehazing. In HFE, we improve the feature extraction capability of the network via deformable convolution to expand the receptive field with an adaptive shape. Specifically, as shown in Fig. 2, HFE is designed under three parallel branches to produce hierarchical features with different resolutions and depths. Each branch in HFE is composed of a 3 × 3 convolution Conv and a deformable convolution DCN [33]. The convolution is used to transform the resolution and depth of the input feature, and the deformable convolution is employed to extract abundant features. We experimentally demonstrate that DCN increases

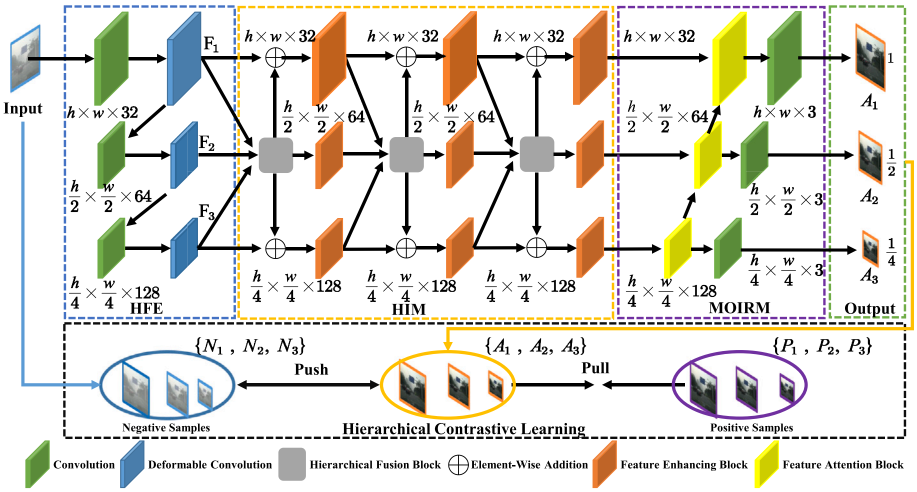


Fig. 2. The architecture of our HCD. It includes a hierarchical dehazing network (HDN) shown at the top and a hierarchical contrastive loss (HCL) shown at the bottom. HDN consists of a hierarchical feature extractor (HFE), a hierarchical interaction module (HIM), and a multi-output image reconstruction module (MOIRM). In HIM, the hierarchical fusion block(HFB) is the core component, which is shown in Fig. 3. The proposed HCD employs contrastive constraints in a hierarchical structure manner to perform feature representation learning, which can better help haze removal.

the dehazing performance of the model. In particular, for a hazy input image N, the output feature resolution of the upper branch is the same as N, and for the other two branches, the resolutions of features are decreased by factors of $\frac{1}{2}$ and $\frac{1}{4}$ respectively. In addition, the depths of hierarchical features $\mathbf{F_1}, \mathbf{F_2}, \mathbf{F_3}$ outputted from the three branches are 32, 64, and 128, respectively.

Hierarchical Interaction Module. As discussed in previous work, conventional networks are susceptible to bottleneck effects [17] because the progressive down-sampling operations in the feature extractor stage cause the feature information loss problem. Therefore, we propose a plug-and-play hierarchical interaction module (HIM), as shown in Fig. 2, to let information flow exchange effectively across different branches in the network. HIM is located between the hierarchical feature extractor and the multi-output image reconstruction module, which contains three identical sub-modules. More specifically, each sub-module consists of a hierarchical fusion block (HFB) and three feature enhancement blocks (FEB). The features $\{\mathbf{F_1}, \mathbf{F_2}, \mathbf{F_3}\}$ from HFE are firstly enhanced by HFB and are then refined by FEB. For simplicity, we introduce one sub-module in the following.

The input hierarchical features $\{\mathbf{F_1}, \mathbf{F_2}, \mathbf{F_3}\}$ are first processed by HFB. As shown in Fig. 3, to fully exploit hierarchical features from non-adjacent stages, HFB is divided into two core steps: *Bottom-Up Fusion:* HFB firstly hierarchically propagates information from the bottom branch to the top branch. This process contains two fusion stages. In the first stage, feature $\mathbf{F_2}$ from the second branch is progressively forwarded by a ReLU non-linear activation function and a 3 × 3 convolution to produce the enhanced feature F_{21}. Inspired by [34], we then compute the difference between F_{21} and $\mathbf{F_3}$ and update the feature $\mathbf{F_2}$ with the computed difference. Again, in the second stage, feature $\mathbf{F_1}$ from the top branch is processed by a ReLU and a convolution. We then compute the difference between the output of the first stage and $\mathbf{F_1}$, and obtain the updated feature $\mathbf{F}_1'$. In this way, the high-level context information from the bottom branch is propagated to the top branch. This process is formulated as:

$$\begin{aligned} F_{21} &= \mathbf{F_3} - \mathrm{Conv}(\mathrm{ReLU}(\mathbf{F_2})), \\ F_{22} &= \mathrm{TransConv}(\mathrm{ReLU}(F_{21})) + \mathbf{F_2}, \\ F_{11} &= F_{22} - \mathrm{Conv}(\mathrm{ReLU}(\mathbf{F_1})), \\ \mathbf{F}_1' &= \mathrm{TransConv}(\mathrm{ReLU}(F_{11})) + \mathbf{F_1}, \end{aligned} \tag{1}$$

where $\mathbf{F}_1'$ is the updated version of $\mathbf{F_1}$, Conv refers to a 3×3 convolution with a stride of 2, and TransConv denotes deconvolution that is applied to transform the shapes of features so that the features from different scales can be used.

Top-Down Fusion: To further fuse the hierarchical features, we design a symmetric hierarchical top-down fusion structure in HFB. As illustrated in Fig. 3, similar to hierarchical bottom-up fusion, hierarchical top-down fusion has two fusion stages. In the first fusion stage, a ReLU activation function and a deconvolution are employed to transform shapes of feature F_{22} to the same as $\mathbf{F}_1'$. Then, the differences between F_{22} and $\mathbf{F}_1'$ are used to refine the feature F_{22}. After that, in the second fusion stage, the refined feature $\mathbf{F}_2'$ is fed into the bottom branch to further refine $\mathbf{F_3}$. The process of the top-down fusion can be presented as:

$$\begin{aligned} F_{23} &= \mathbf{F}_1' - \mathrm{TransConv}(\mathrm{ReLU}(F_{22})), \\ \mathbf{F}_2' &= \mathrm{Conv}(\mathrm{ReLU}(F_{23})) + F_{22}, \\ F_{31} &= \mathbf{F}_2' - \mathrm{TransConv}(\mathrm{ReLU}(\mathbf{F_3})), \\ \mathbf{F}_3' &= \mathrm{Conv}(\mathrm{ReLU}(F_{31})) + \mathbf{F_3}, \end{aligned} \tag{2}$$

where $\mathbf{F}_1', \mathbf{F}_2', \mathbf{F}_3'$ are the outputs of HFB. In the end, as shown in Fig. 2, the outputs of HFB are further strengthened via parallel residual connection and FEB, where FEB is the Residual Dense Block in [17]. As shown in Fig. 4, FEB block consists of five densely connected convolutional layers and a residual skip connection. The first four layers use a kernel size of 3 to increase the number of feature maps. Subsequently, the last layer utilizes a kernel size of 1 to fuse these feature maps. Finally, a skip connection is employed to combine the output of this block with its input. Our HIM module can thus exploit multi-scale hierarchical features to improve the dehazing performance.

Multi-output Image Reconstruction Module. In HIE, different output branches produce feature maps with different resolutions. We consider that these hierarchical feature maps with different characteristics can be used to produce different samples for subsequent contrastive learning. Thus, we design a multi-output image reconstruction module in the tail of the network. As illustrated in Fig. 2, in each branch, we first apply a feature attention block (FAB) to refine the feature, and then use a single convolution layer to reconstruct the image. The image reconstruction in each branch can be formulated as follows:

$$A_n = \begin{cases} \mathrm{Conv}\left(\mathrm{FAB}_n\left(\left(\mathrm{FAB}_{n+1}^{\mathrm{out}}\right)^{\uparrow}; \mathrm{HIM}_n^{\mathrm{out}}\right)\right), & n = 1,2 \\ \mathrm{Conv}\left(\mathrm{FAB}_n\left(\mathrm{HIM}_{n+1}^{\mathrm{out}}\right)\right), & n = 3 \end{cases} \tag{3}$$

where $\mathrm{HIM}_n^{\mathrm{out}}$, $\mathrm{FAB}_n^{\mathrm{out}}$ are the outputs of the nth branch HIM and FAB. Conv is a 3 × 3 convolution. Up-sampling ↑ is used such that the features from different scales can be fused.

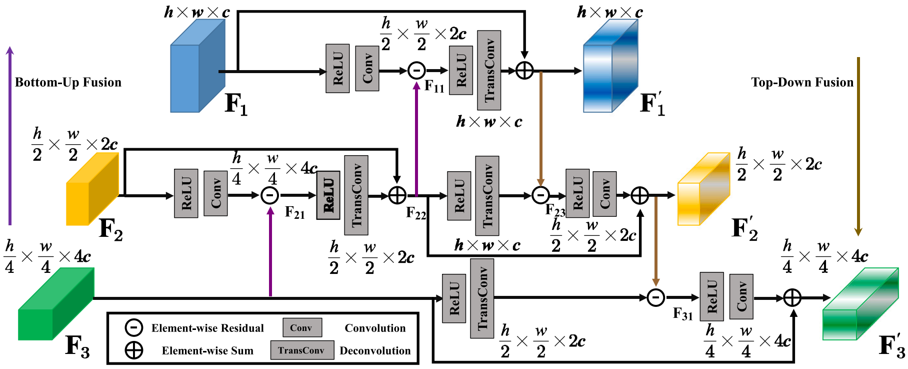


Fig. 3. Illustration of the hierarchical fusion block (HFB). It contains two steps: Bottom-Up Fusion and Top-Down Fusion. $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ are input hierarchical features, and $\mathbf{F}_1', \mathbf{F}_2', \mathbf{F}_3'$ are corresponding updated versions by HFB. Element-wise Residual operation refers to obtaining the residual feature by subtracting the corresponding elements of two features. The Bottom-Up Fusion and Top-Down Fusion stages can be formulated as in Eqs. (1) and (2) respectively.

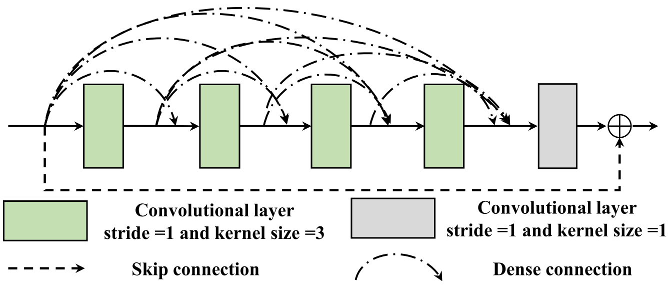


Fig. 4. Illustration of the feature enhancement block (FEB). FEB block consists of five densely connected convolutional layers and a residual skip connection.

3.3. Hierarchical contrastive learning

The proposed HDN recovers the images in a hierarchical structure. Therefore, it is natural that we consider joint optimization of all the dehazed images output by HDN. To this end, we resort to contrastive learning, which is a discriminant-based technique that pulls similar samples closer while pushing dissimilar samples away [14,30]. Though contrastive learning has shown effectiveness in many high-level vision tasks, its potential for single image dehazing has not been fully explored. We thus propose a novel hierarchical contrastive loss to guide our HDN to remove complex haze components. Two aspects are required to consider when constructing our loss: one is to build the appropriate positive and negative pairs of samples, and the other is to find a suitable latent space to differentiate sample distribution. For the first aspect, benefiting from the hierarchical structure of HDN, we can collect abundant contrastive samples in different resolutions. We label hazy input images in three resolutions as negative samples, and regard their corresponding ground truth images as positive samples. For the second aspect, inspired by [14], we employ a pre-trained VGG-19 network [35] to obtain feature embedding for measuring feature similarity. The VGG-19 network [35] is trained on ImageNet dataset [36] for image classification.

As shown in Fig. 5, to build the positive and negative pairs, we pair the restored images $\{A_1, A_2, A_3\}$ by HDN and their corresponding ground truth images $\{P_1, P_2, P_3\}$ (positive samples) in different resolutions as positive-oriented supervision to guide HDN to recover haze-free images. We take the hazy input images $\{N_1, N_2, N_3\}$ together with the restored images as negative pairs to enforce HDN to focus more on learning the complex haze components. The hierarchical contrastive loss is represented as:

$$\mathcal{L}_{\mathrm{hcl}} = \sum_{i=1}^{3} \left(\sum_{j=1}^{3} \left\| \hat{A}_i - \hat{P}_j \right\|_1 \right) \left(\sum_{k=1}^{3} \frac{1}{\left\| \hat{A}_i - \hat{N}_k \right\|_1} \right), \tag{4}$$

where $\hat{N}, \hat{A}, \hat{P}$ denote the extracted features from the VGG-19 network and $\|.\|_1$ is L_1 norm to measure the similarity between two extracted features. Such a special design is expected to capture the relationship between recovered images in different resolutions. In the implementation, we extract the features from the 1st, 3rd, 5th, 9th, and 13th layers of the pre-trained VGG-19 and set the corresponding coefficients as $\frac{1}{32}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, and 1, respectively, which have been shown to be effective in image dehazing [14,17]. These features contain high-level semantic information and can effectively represent the global and local features of the image. In addition, when dealing with the cross-scale images, we interpolate images with different scales to the middle scale to calculate the loss. Our hierarchical contrastive loss aims to use the information from ground-truth images as an upper bound and treat hazy images as a lower bound to leverage the information from positive and negative images for image dehazing. Therefore, it is important to note that the reference embeddings in our loss are the generated images, not the ground-truth images.

3.4. Loss function

To train our proposed HDN, we design a loss function combining the Charbonnier loss [37] and the proposed hierarchical contrastive loss. We regard the Charbonnier loss as a pixel-wise loss, which is used between the recovered images and the ground truth images at each scale. The Charbonnier loss is defined as:

$$\mathcal{L}_{\mathrm{char}} = \frac{1}{3} \sum_{k=1}^{3} \sqrt{\|A_k - P_k\|^2 + \epsilon^2}, \tag{5}$$

where A_k and P_k represent the dehazed image and ground-truth image respectively. k is the index of the image scale level in the model. The constant ϵ is empirically set to 10^{-3}. With this constant, the network trained with Charbonnier loss can better handle outliers and improve performance over $\mathcal{L}_2$ loss function. In addition, the network can converge faster [37]. The final loss function $\mathcal{L}_{\mathrm{total}}$ to train our proposed HDN is determined as follows:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{char}} + \lambda \mathcal{L}_{\mathrm{hcl}}, \tag{6}$$

where $\mathcal{L}_{\mathrm{char}}$ is the Charbonnier loss, $\mathcal{L}_{\mathrm{hcl}}$ is the proposed hierarchical contrastive loss. λ is a hyper-parameter to balance these two loss terms, which is empirically set to 0.1.

4. Experiments

In this part, we first explain the implementation details of the proposed method. Then, we show the image dehazing results of our approach and the comparison with the state-of-the-art methods. Finally, we conduct extensive ablation studies to verify the effectiveness of modules in the proposed method.

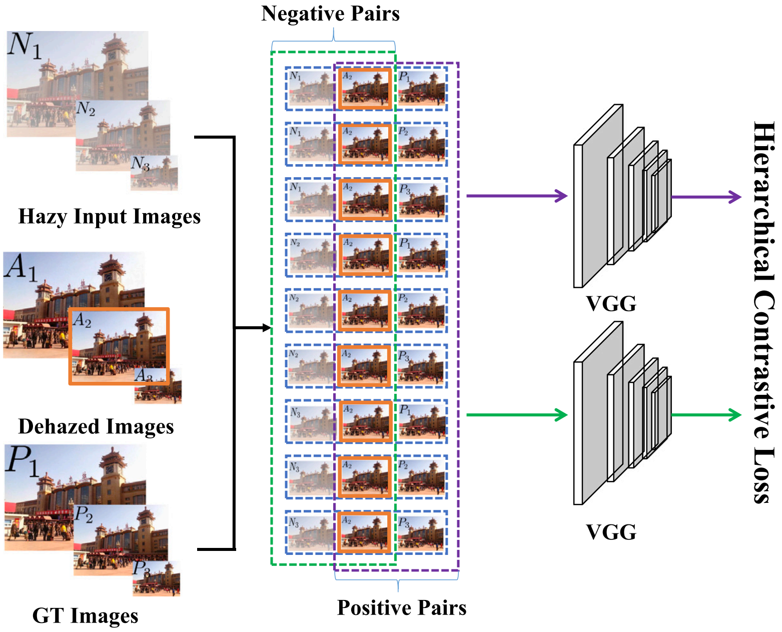


Fig. 5. Illustration of the proposed hierarchical contrastive learning. Taking the recovered image A_2 as an example, we construct different types of positive and negative pairs. The image A_2, positive samples and negative samples are encoded into features by a pre-trained VGG-19 network. The hierarchical contrastive loss compares these features and guides the network to learn more useful information for image dehazing.

Table 1
Comparison results with the state-of-the-art image dehazing approaches on the benchmark datasets. The best and the second best performances are highlighted and underlined respectively. The proposed method achieves the best performance compared with previous state-of-the-arts.

Methods	SOTS (indoor/outdoor)		SOTS-average		HazeRD		DENSE-HAZE	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DCP [19]	15.16/13.85	0.8546/0.5416	13.85	0.6516	15.22	0.7737	10.06	0.3856
DehazeNet [4]	19.82/24.75	0.8209/0.9269	22.29	0.8739	15.63	0.7517	13.84	0.4252
AOD-Net [3]	20.51/24.14	0.8162/0.9198	22.33	0.8680	15.54	0.7449	13.14	0.4144
GridDehazeNet [17]	32.16/30.86	0.9836/0.9819	31.51	0.9828	16.05	0.7932	13.31	0.3681
FFA-Net [11]	36.39/33.57	0.9886/0.9840	34.98	0.9698	14.96	0.7654	14.39	0.4524
MSBDN [34]	−/−	−/−	33.79	0.9840	15.78	0.7982	15.37	0.4858
AECR-Net [14]	37.17/−	0.9901/−	−	−	−	−	15.80	0.4660
HCD	**38.31/34.02**	**0.9954/0.9936**	**36.16**	**0.9945**	**16.88**	**0.8088**	**16.41**	**0.5662**

4.1. Implementation details

Datasets. We evaluate the proposed method on synthetic and real-world datasets. RESIDE [15] is a large-scale synthetic dataset including indoor and outdoor scenes. This dataset contains five subsets, *i.e.,* Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objective Testing Set (SOTS), Real World task-driven Testing Set (RTTS), and Hybrid Subjective Testing Set (HSTS). ITS contains 13,990 hazy images and 1,399 clear images. SOTS consists of 500 indoor and outdoor paired images respectively. Following the previous works [11,17,34], we adopt ITS and OTS to train the proposed method respectively, and use SOTS and RTTS for performance evaluation. Furthermore, we test the proposed method on more challenging datasets, including HazeRD [16] and DENSE-HAZE [38], which are collected in real-world scenarios.

Experimental Details. In the experiment, we augment the training data with random rotations of 90, 180, and 270. We randomly crop a 240 × 240 patch in the training stage. The batch size is set to 16 and all weights of the models are initialized by the Xavier method. The learning rate is 2×10^{-4}, which is steadily decreased to 1×10^{-6} by the cosine annealing strategy. We train all models for 400 epochs. And models are optimized by the Adam optimizer, where β_1 and β_2 are set to 0.9 and 0.999 respectively. In addition, we adopt the Pytorch framework to perform all experiments on the NVIDIA Tesla V100 GPU.

Comparison Methods. We compare performance of our method with seven state-of-the-art image dehazing methods: DCP [19], DehazeNet [4], AOD-Net [3], GridDehazeNet [17], FFA-Net [11], MSBDN [34], and AECR-Net [14].

4.2. Evaluation on synthetic datasets

We evaluate our method on two synthetic dehazing datasets (*i.e.,* SOTS and HazeRD). The 2nd and 3rd columns in Table 1 show quantitative results on SOTS dataset in terms of PSNR and SSIM. As reported in the Table, DCP [19], DehazeNet and AOD-Net present low values of PSNR and SSIM in both indoor and outdoor subsets, indicating that the dehazing results with low quality are produced. Compared with the previous methods, recent end-to-end methods (GridDehazNet [17], FFA-Net [11], MSBDN [34], AECR-Net [14], and HCD) obtain better performance. Among them, the dehazing performance values of our proposed method rank first in both indoor and outdoor subsets of SOTS regarding PSNR and SSIM. For the indoor subset, our model surpasses the second best method (AECR-Net [14]) over 1.14 dB, 0.0053 on PSNR and SSIM respectively. For the outdoor subset, our model achieves 0.45 dB and 0.0096 improvement in terms of PSNR and SSIM. As for a more challenging dataset HazeRD, the evaluation results are shown in the 4th column of Table 1. We find that the performance of

Table 2
Quantitative comparison on SOTS-Indoor and DENSE-HAZE datasets. The best and second best performances are bold and underlined respectively. To ensure the fairness of comparison, we calculate MACs based on 256 × 256 color images.

Methods	SOTS (indoor)		DENSE-HAZE		Overhead	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	#Param	MACs
AECR-Net [14]	37.17	0.9901	15.80	0.4660	2.611 M	52.20 G
AECR-Net (large)	37.96	0.9941	16.02	0.5592	8.570 M	105.41 G
HCD (tiny)	37.62	0.9940	15.86	0.5603	2.580 M	51.04 G
HCD	**38.31**	**0.9954**	**16.41**	**0.5662**	5.580 M	104.03 G

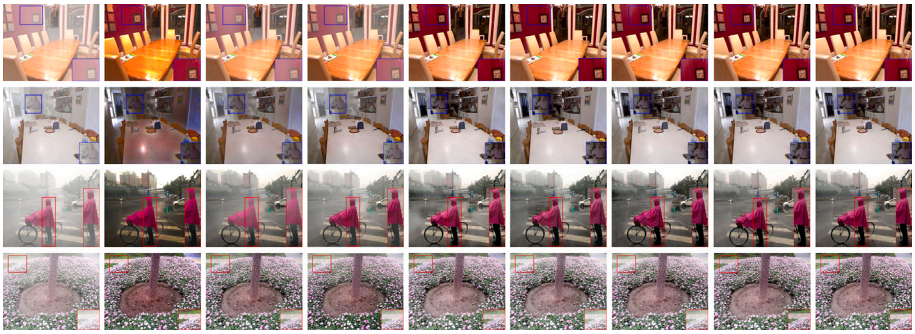


Fig. 6. Visual comparison with the state-of-the-art image dehazing methods on the SOTS dataset. The top two rows are images from the indoor subset, and the bottom two rows are images from the outdoor subset. **Zoom in for details**.

some learning-based methods is worse than that of the SOTS dataset. For example, FFA-Net [11] obtains PSNR of 14.96 dB and SSIM of 0.7654, which is even worse than DCP [19]. This may be attributed to that, the number of samples in the HazeRD dataset is not sufficient enough to train the models well. Among all comparison methods, our HCD achieves the best dehazing performance in terms of PSNR and SSIM. The comparison reveals that our method achieves the highest performance for image dehazing.

To compare with the state-of-the-art AECR-Net, we conduct experiments from the following two aspects: (1) Increasing the complexity of AECR-Net (called AECR-Net (large)) to make it close to the proposed HCD for performance comparison; (2) Reducing the complexity of our proposed model (named HCD (tiny)), and then compares the performance with AECR-Net. The results are shown in Table 2 demonstrate that the proposed HCD achieves better performance and complexity trade-off (*i.e.,* 36.16 dB, 5.58 M, 104.03 G) than the state-of-the-art AECR-Net.

We also demonstrate visual comparisons of the dehazed results from different methods. As shown in Fig. 6, we present the haze removal results for all comparison methods in indoor and outdoor hazy images. In Fig. 6, the top two rows are images from the indoor testing subset, and the bottom two rows correspond to the outdoor testing subset. From Fig. 6, we find that DCP [19] and DehazeNet [4] suffer from the color distortion problem so their dehazed results seem unrealistic. The recovered images of AOD-Net [3] are darker than the ground truth images in some cases (*e.g.,* the dining table in the second column of Fig. 6). Though GridDehazeNet [17], MSBDN [34], and FFA-Net [11] perform well, the artifacts of incomplete haze removal still exist in the restored images, *e.g.,* the flower in the images of Fig. 6. Compared with these methods, our method produces images with rich details and color information, and there are rare artifacts in the dehazed images. Overall, our method achieves the best performance among the comparative methods from both quantitative and qualitative aspects.

4.3. Evaluation on real world datasets

We further evaluate the proposed method quantitatively and qualitatively on real-world images. Specifically, we conduct quantitative experiments on the DENSE-HAZE dataset, in which hazy images are captured in natural scenes. As shown in Table 1, our HCD achieves the best performance in terms of PSNR and SSIM. Our HCD significantly outperforms the second best method. To be specific, compared with the state-of-the-art AECR-Net [14], our HCD achieves an advance of 0.61 dB in terms of PSNR and surpasses MSBDN [34] by 0.08 with regard to SSIM.

Table 3
Ablation study of individual components. The values are average dehazed results of different variants on the indoor subset of the SOTS dataset. The best and second best performances are bold and underlined respectively.

Models	Components			SOTS-indoor		Params.
	DCN	HFB	HCL	PSNR ↑	SSIM ↑	
baseline	✗	✗	✗	34.18	0.9926	**2.34** M
baseline+DCN	✔	✗	✗	34.79	0.9921	4.04 M
baseline+DCN+HFB	✔	✔	✗	36.69	0.9931	5.58 M
full model (HCD)	✔	✔	✔	**38.31**	**0.9954**	5.58 M

As for qualitative comparison, we compare the visual results of different methods on real-world hazy images from the RTTS dataset. As illustrated in Fig. 7, prior-based method DCP [19] suffers from the color distortion problem (*e.g.,* the sky and street in the images of the second column in Fig. 7). DehazeNet [4] and AOD-Net [3] cause color distortion in some scenery, and GridDehazeNet [17] encounters dark-area artifacts (*e.g.,* see the third image in the fifth column of Fig. 7). Compared to the above methods, MSBAN [34] and FFA-Net [11] produce better dehazing results. However, they still have the incomplete haze removal problem and produce some remaining haze artifacts in the restored images. In contrast, our model can work well and produce more visually plausible dehazing results in real-world scenery.

4.4. Ablation study

We conduct extensive ablation studies to validate each component of our HCD. For a fair comparison, all models are trained on ITS with the same settings and are tested on the indoor subsets of the SOTS dataset.

Individual components. To better verify each component of our HCD, we conduct an ablation study by considering three factors: the

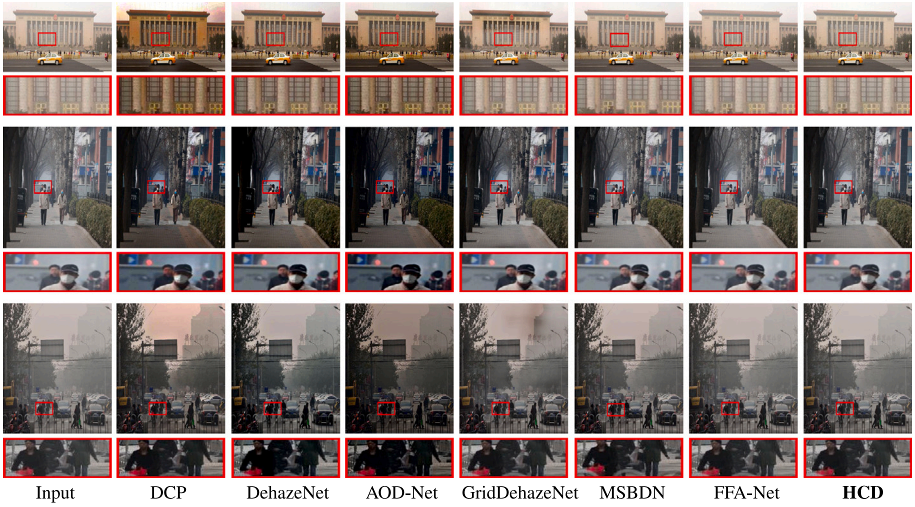


Fig. 7. Visual comparison on real-world hazy images from the RTTS dataset. **Zoom in for details**.

Table 4
Ablation study of our HCL. $\mathcal{L}_{\text{char}}$ indicates using pixel-wise loss in a multi-scale manner, and $\mathcal{L}_{\text{hchar}}$ represents using pixel-wise loss in a hierarchical structure. $\mathcal{L}_{\text{hcl}}$ refers to the proposed contrastive loss. The best and second best performances are bold and underlined respectively.

Models	Components			SOTS-indoor	
	$\mathcal{L}_{\text{char}}$	$\mathcal{L}_{\text{hchar}}$	$\mathcal{L}_{\text{hcl}}$	PSNR ↑	SSIM ↑
Model1	✔	✗	✗	36.69	0.9931
Model2	✗	✔	✗	37.01	0.9932
HCD	✔	✗	✔	**38.31**	**0.9954**

Table 5
Comparison results of the state-of-the-art image deraining approaches on the benchmark datasets. The best and second best performances are bold and underlined respectively.

Methods	Test100		Rain100L	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DerainNet [39]	22.77	0.810	27.03	0.884
SEMI [40]	22.35	0.788	25.03	0.842
DIDMDN [41]	22.56	0.818	25.23	0.741
UMRL [42]	24.41	0.829	29.18	0.923
RESCAN [43]	25.00	0.835	29.80	0.881
PReNet [44]	24.81	0.851	32.44	0.950
MSPFN [45]	27.50	0.876	32.40	0.933
HCD	**29.43**	**0.892**	**35.01**	**0.956**

first is the deformable convolution DCN in the hierarchical feature extractor. The second is the hierarchical fusion block (HFB) in the hierarchical interaction module and the last one is the hierarchical contrastive loss (HCL). Regarding these factors, we design the following variants to realize the ablation analysis: (1) baseline: it is a baseline network, which does not adopt any of the above components (*i.e.,* the deformable convolution DCN in the hierarchical feature extractor, HFB, and HCL). It means the network cannot benefit from the deformable convolution in the stage of feature extraction, the hierarchical feature fusion by HFB, and the contrastive learning. (2) baseline+DCN: Add the deformable convolution into the hierarchical feature extractor of the baseline network. It means the network can extract more abundant features for image dehazing. (3) baseline+DCN+HFB: This model adds the HFB based on baseline+DCN. (4) HCD: it is our complete model, which is optimized by the proposed HCL and Charbonnier loss. HCL allows the network to use both negative and positive samples for training. In addition, baseline, baseline+DCN, and baseline+DCN+HFB are trained only by the Charbonnier loss. The detailed configuration of these models can be found in Table 3.

The values of PSNR and SSIM are represented in Table 3. The performance of our complete model HCD shows great superiority over its incomplete counterparts, including baseline+DCN+HFB (removing HCL component), baseline+DCN (removing HCL and HFB), and baseline (removing HCL, HFB, and DCN). Comparing baseline and baseline+DCN, the results show that embedding DCN in the hierarchical feature extractor leads to better performance. Moreover, for baseline+DCN+HFB, it surpasses baseline+DCN by 1.89 dB, 0.001 in terms of PSNR and SSIM. The result verifies that designing a reasonable fusion module in the model is important. Finally, HCD gains an evident improvement regarding baseline+DCN+HFB, proving that the proposed hierarchical contrastive learning effectively guides the network to enhance feature representation for image dehazing. In conclusion, each component of this work contributes to improving the dehazing performance.

Ablation study about our HCL. We design a pixel-wise loss in a hierarchical manner like contrastive loss and train the proposed network with the designed hierarchical pixel-wise loss. The results of the derived Model2 are shown in Table 4. Model2 is trained with only the hierarchical pixel-wise loss, and there is only a slight performance improvement compared to Model1 (trained using pixel-wise loss in a multi-scale manner). The comparison results show that the performance improvement of dehazing can indeed be attributed to using the proposed contrastive loss.

4.5. Generalization on other tasks

Image Deraining. To explore the generalization of the proposed method, we consider adopting our model for the task of image deraining, which is similar to image dehazing. Following the previous work [45], we use synthetic paired rainy images to train our model and test the performance on both Test100 [47] and Rain100L [46] testing sets. The training strategy is the same as the image dehazing task. We adopt PSNR and SSIM metrics to evaluate the deraining performance. In this task, PSNR and SSIM are calculated on the Y channel in the YCbCr color space like [45]. We choose seven state-of-the-art image deraining methods for comparison: (1) deraining network (DerainNet) [39], (2) semi-supervised deraining method (SEMI) [40], (3) density aware multi-stream densely connected convolutional neural network (DIDMDN) [41], (4) uncertainty guided multi-scale residual learning network (UMRL) [42], (5) recurrent squeeze-and-excitation context aggregation network (RESCAN) [43], (6) progressive recurrent

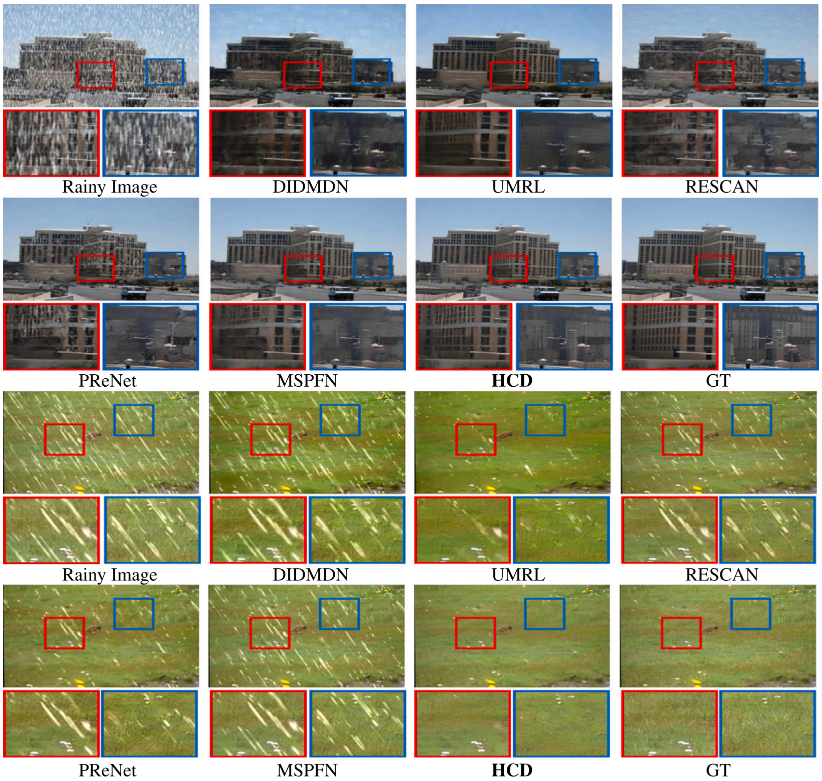


Fig. 8. Visual comparison for the image deraining task on the Test100 and Rain100L datasets [46]. The top two rows of images are from the Test100 dataset, and the bottom two rows are images from the Rain100L dataset. The images generated by our HCD have almost no rainy drops, which are clear and more similar to the GT images.

Table 6
Comparison results for the nighttime dehazing task. The best and second best performances are bold and underlined respectively.

Methods	NHR		
	PSNR ↑	SSIM ↑	LPIPS ↓
NDIM [48]	14.31	0.5256	–
GS [49]	17.32	0.6285	–
MRPF [50]	16.95	0.6674	–
MPR [50]	19.93	0.7772	0.3072
OSFD [51]	21.32	0.8035	0.2910
HCD (Ours)	**23.43**	**0.9533**	**0.0729**

Table 7
Quantitative results when applying the proposed contrastive loss into the existing state-of-the-art dehazing methods.

Methods	SOTS (indoor)	
	PSNR	SSIM
GridDehazeNet [17]	33.24 (1.08 ↑)	0.9887 (0.0051 ↑)
FFA-Net [11]	36.96 (0.57 ↑)	0.9908 (0.0022 ↑)
MSBDN [34]	34.66 (0.87 ↑)	0.9889 (0.0049 ↑)
AECR-Net [14]	37.83 (0.66 ↑)	0.9912 (0.0011 ↑)

network (PReNet) [44], and (7) multi-scale progressive fusion network (MSPFN) [45].

The quantitative results on Test100 and Rain100L are reported in Table 5. Our HCD achieves remarkable performance on both Test100 and Rain100L datasets. Especially, on the Rain100L dataset, our HCD surpasses MSPFN by 2.61 dB and 0.023 in terms of PSNR and SSIM. We also show the visual comparison in Fig. 8. We note that the state-of-the-art deraining methods (*e.g.,* PReNet and MSPFN) do not remove the rain streaks well (see patches of restored images in Fig. 8). In contrast, our HCD can remove the rain well and recover high-quality images with truthful details compared to other methods. The comparison results demonstrate our HCD generalizes well on image deraining, though our HCD is specifically designed for image dehazing.

Nighttime Dehazing. We further investigate the potential of our proposed method on the nighttime dehazing task. To be specific, we adopt the NHR dataset [51] to train and test our proposed method. NHR contains 17,940 pairs of images. We choose 1,794 pairs of images for evaluation and other samples for training, following the previous work [51]. For comparison, we select five representative nighttime image dehazing methods, including NDIM [48], GS [49], MRPF [50], MPR [50], and OSFD [51].

The quantitative results are reported in Table 6. We use both pixel-wise (PSNR and SSIM) and perceptual (LPIPS) metrics to evaluate the performance. As listed in Table 6, HCD achieves the best performance in terms of PSNR, SSIM, and LPIPS, respectively. Especially, the advance of HCD is 2.11 dB and 0.1498 in terms of PSNR and SSIM compared to OSFD [51] that is the best method among the comparison methods. The perceptual metric LPIPS further demonstrates the superiority of our HCD. Fig. 9 provides a visual comparison. Compared with state-of-the-art methods (MPR [50] and OSFD [51]), the proposed HCD can effectively remove the haze and recover images with better brightness at the same time, and it introduces fewer artifacts. Quantitative and qualitative results indicate that the proposed HCD has a strong capability to process hazy images under low-light conditions.

4.6. Discussion about the proposed HCL

To further test the universality of the proposed hierarchical loss, we add the HCL into the existing state-of-the-art methods including GridDehazeNet [17], FFA-Net [11], MSBDN [34], and AECR-Net [14] for performance evaluation. As shown in Table 7, our proposed loss can indeed improve the performance of SOTS methods. For example, GridDehazeNet [17] achieves higher PSNR and SSIM with gains of 1.08 and 0.0051, respectively. This universal experimental validation shows

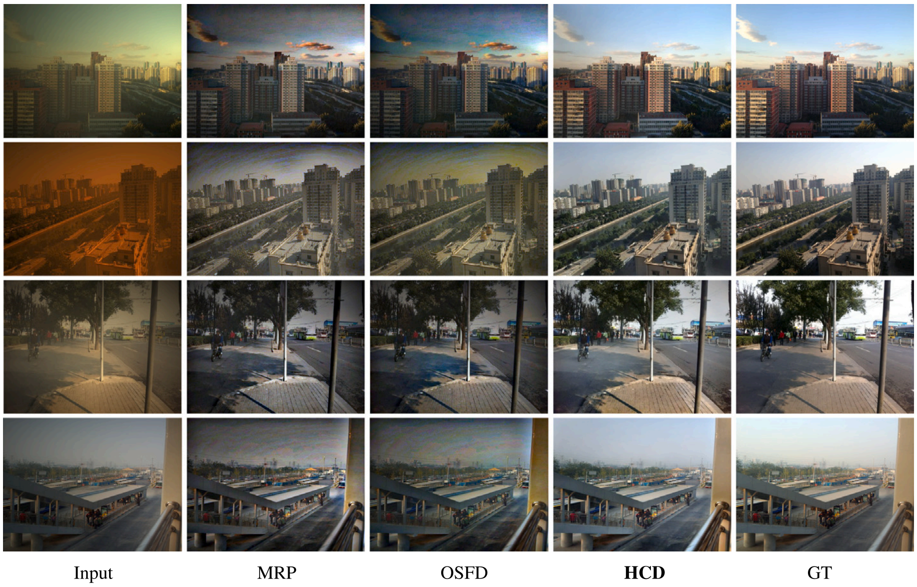
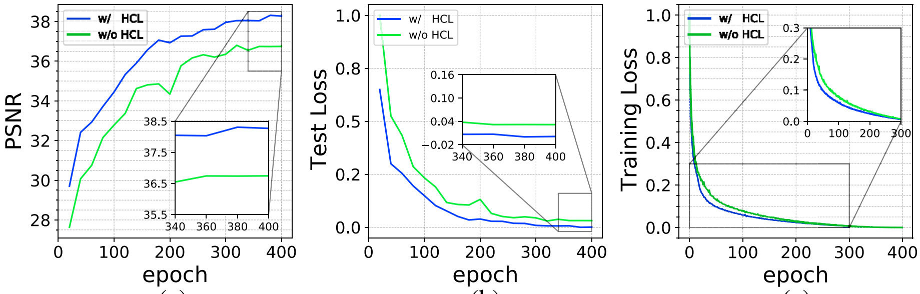


Fig. 9. Visual examples for the nighttime dehazing task on the NHR dataset [51]. Our HCD can successfully remove the haze in low-light conditions and does not introduce artifacts in the recovered images.



(a) (b) (c)

Fig. 10. Quantitative comparison between the proposed models trained with and without HCL. (a) Results in the SOTS-indoor testing set. (b) Test loss curve. (c) Training loss curve. The model trained with HCL achieves better performance.

that the proposed loss does not rely on a particular network and it can train the dehazing network effectively.

To demonstrate the effectiveness of our proposed HCL, we conduct an analysis to assess whether it enhances the model's performance. To validate this, we perform experiments by removing HCL (as shown in Fig. 10(a), (b) and (c)). The results consistently show that while HCL may not significantly accelerate the convergence speed during training, the model trained with HCL consistently achieves higher PSNR values on the testing set compared to the model trained without it. This substantial improvement in PSNR values suggests that our proposed HCL positively impacts the model's performance, enhancing its ability to produce superior image restorations.

5. Conclusion

In this work, we propose a novel hierarchical contrastive dehazing (HCD) method, which consists of a hierarchical dehazing network (HDN) and a hierarchical contrastive loss (HCL). In HDN, we propose a hierarchical interaction module, which effectively fuses the hierarchical features so that the learned features can better facilitate haze removal. To further remove the haze component from the input image, our special hierarchical HDN performs hierarchical contrastive learning by constructing the positive and negative pairs in a hierarchical manner. Extensive experimental results on synthetic benchmarks and real-world images have shown the great superiority of our HCD over state-of-the-art methods.

However, similar to other image dehazing methods, our method still has some limitations under real-world application. Accordingly, there are multiple directions to explore in the future. First, due to our method relying on paired data, we will explore the unsupervised learning strategy, which can help our method easily deal with the image dehazing problem on real-world scenery. Second, we will extend our method to many other relevant image restoration tasks, such as image desnowing, image denoising, image deblurring, and low-light image enhancement.

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 was supported in part by the National Natural Science Foundation of China (Grant No. 62372223), in part by the Fundamental Research Funds for the Central Universities (No. XJ2023000701).

References

[1] U. Ali, J. Choi, K. Min, Y.-K. Choi, M.T. Mahmood, Boundary-constrained robust regularization for single image dehazing, Pattern Recognit. 140 (2023) 109522.

[2] R.T. Tan, Visibility in bad weather from a single image, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[3] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, Aod-net: All-in-one dehazing network, in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 4770–4778.

[4] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.

[5] Z. Li, C. Zheng, H. Shu, S. Wu, Dual-scale single image dehazing via neural augmentation, IEEE Trans. Image Process. 31 (2022) 6213–6223.

[6] C. Lin, X. Rong, X. Yu, Msaff-net: Multi-scale attention feature fusion networks for single image dehazing and beyond, IEEE Trans. Multimed. (2022).

[7] N. Jiang, K. Hu, T. Zhang, W. Chen, Y. Xu, T. Zhao, Deep hybrid model for single image dehazing and detail refinement, Pattern Recognit. 136 (2023) 109227.

[8] Y. Liu, X. Hou, Local multi-scale feature aggregation network for real-time image dehazing, Pattern Recognit. (2023) 109599.

[9] H. Sun, B. Li, Z. Dan, W. Hu, B. Du, W. Yang, J. Wan, Multi-level feature interaction and efficient non-local information enhanced channel attention for image dehazing, Neural Netw. 163 (2023) 10–27.

[10] B. Hariharan, P. Arbeláez, R. Girshick, J. Malik, Hypercolumns for object segmentation and fine-grained localization, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 447–456.

[11] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, FFA-Net: Feature fusion attention network for single image dehazing, in: Proceedings of AAAI Conference on Artificial Intelligence, 2020, pp. 11908–11915.

[12] Y. Qu, Y. Chen, J. Huang, Y. Xie, Enhanced pix2pix dehazing network, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8160–8168.

[13] X. Zhang, T. Wang, W. Luo, P. Huang, Multi-level fusion and attention-guided CNN for image dehazing, IEEE Trans. Circuits Syst. Video Technol. 31 (11) (2020) 4162–4173.

[14] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma, Contrastive learning for compact single image dehazing, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 10551–10560.

[15] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, Z. Wang, Benchmarking single-image dehazing and beyond, IEEE Trans. Image Process. 28 (1) (2018) 492–505.

[16] Y. Zhang, L. Ding, G. Sharma, Hazerd: an outdoor scene dataset and benchmark for single image dehazing, in: Proceedings of IEEE International Conference on Image Processing, 2017, pp. 3205–3209.

[17] X. Liu, Y. Ma, Z. Shi, J. Chen, Griddehazenet: Attention-based multi-scale network for image dehazing, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7314–7323.

[18] L. Mutimbu, A. Robles-Kelly, A factor graph evidence combining approach to image defogging, Pattern Recognit. 82 (2018) 56–67.

[19] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.

[20] R. Fattal, Dehazing using color-lines, ACM Trans. Graph. 34 (1) (2014) 1–14.

[21] Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, IEEE Trans. Image Process. 24 (11) (2015) 3522–3533.

[22] F. Yuan, Y. Zhou, X. Xia, X. Qian, J. Huang, A confidence prior for image dehazing, Pattern Recognit. 119 (2021) 108076.

[23] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, D. Tao, A comprehensive survey and taxonomy on single image dehazing based on deep learning, ACM Comput. Surv. (2022).

[24] S. Yin, Y. Wang, Y.-H. Yang, A novel image-dehazing network with a parallel attention block, Pattern Recognit. 102 (2020) 107255.

[25] Z. Chen, Y. Wang, Y. Yang, D. Liu, PSD: Principled synthetic-to-real dehazing guided by physical priors, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 7180–7189.

[26] K. Zhang, Y. Li, Single image dehazing via semi-supervised domain translation and architecture search, IEEE Signal Process. Lett. 28 (2021) 2127–2131.

[27] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[28] D. Engin, A. Genç, H. Kemal Ekenel, Cycle-dehaze: Enhanced cyclegan for single image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 825–833.

[29] A. Dudhane, S. Murala, Cdnet: Single image de-hazing using unpaired adversarial training, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2019, pp. 1147–1155.

[30] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, in: Proceedings of European Conference on Computer Vision, 2020, pp. 319–345.

[31] J. Zhang, S. Lu, F. Zhan, Y. Yu, Blind image super-resolution via contrastive representation learning, 2021, arXiv preprint arXiv:2107.00708.

[32] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

[33] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 764–773.

[34] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 2157–2167.

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of International Conference on Learning Representations, 2015.

[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[37] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proceedings of IEEE International Conference on Image Processing, 1994, pp. 168–172.

[38] C.O. Ancuti, C. Ancuti, M. Sbert, R. Timofte, Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images, in: Proceedings of IEEE International Conference on Image Processing, 2019, pp. 1014–1018.

[39] X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, Clearing the skies: A deep network architecture for single-image rain removal, IEEE Trans. Image Process. 26 (6) (2017) 2944–2956.

[40] W. Wei, D. Meng, Q. Zhao, Z. Xu, Y. Wu, Semi-supervised transfer learning for image rain removal, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3877–3886.

[41] H. Zhang, V.M. Patel, Density-aware single image de-raining using a multi-stream dense network, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 695–704.

[42] R. Yasarla, V.M. Patel, Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8405–8414.

[43] X. Li, J. Wu, Z. Lin, H. Liu, H. Zha, Recurrent squeeze-and-excitation context aggregation net for single image deraining, in: Proceedings of European Conference on Computer Vision, 2018, pp. 254–269.

[44] D. Ren, W. Zuo, Q. Hu, P. Zhu, D. Meng, Progressive image deraining networks: A better and simpler baseline, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3937–3946.

[45] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, J. Jiang, Multi-scale progressive fusion network for single image deraining, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 8346–8355.

[46] W. Yang, R.T. Tan, J. Feng, J. Liu, Z. Guo, S. Yan, Deep joint rain detection and removal from a single image, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1357–1366.

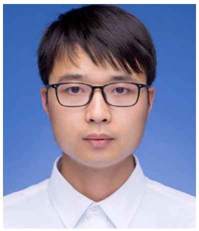
[47] H. Zhang, V. Sindagi, V.M. Patel, Image de-raining using a conditional generative adversarial network, IEEE Trans. Circuits Syst. Video Technol. 30 (11) (2019) 3943–3956.

[48] J. Zhang, Y. Cao, Z. Wang, Nighttime haze removal based on a new imaging model, in: Proceedings of IEEE International Conference on Image Processing, 2014, pp. 4557–4561.

[49] Y. Li, R.T. Tan, M.S. Brown, Nighttime haze removal with glow and multiple light colors, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 226–234.

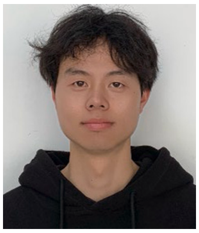
[50] J. Zhang, Y. Cao, S. Fang, Y. Kang, C. Wen Chen, Fast haze removal for nighttime image using maximum reflectance prior, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7418–7426.

[51] J. Zhang, Y. Cao, Z.-J. Zha, D. Tao, Nighttime dehazing with a synthetic benchmark, in: Proceedings of ACM International Conference on Multimedia, 2020, pp. 2355–2363.



Tao Wang is currently a Ph.D. candidate with Department of Computer Science and Technology, Nanjing University, China. He received the B.Sc. degree in information and computing science from Hainan Normal University, China, in 2018. His research interests include several topics in computer vision and machine learning, such as object tracking, image/video quality restoration, adversarial learning, image-to-image translation and reinforcement learning.

Guangpin Tao was a postgraduate candidate in the Department of Computer Science and Technology of Nanjing University during the completion of this work. He obtains a bachelor's degree in computer science and technology from Nankai University in 2009. His research interest is deep-learning based blind image restoration, including image denoising and super-resolution.



Wanglong Lu is a Ph.D. student at Memorial University of Newfoundland. He received his B.Sc. degree in digital media technology from Communication University of Zhejiang in 2018 and his M.Sc. degree in computer software and theory from Wenzhou University in 2021. His research interests include image recognition, Image restoration, image editing and processing.

Kaihao Zhang obtained his Ph.D. degree from the College of Engineering and Computer Science, The Australian National University, Canberra, ACT, Australia. His research interests focus on computer vision and deep learning. He has more than 30 referred publications in international conferences and journals, including CVPR, ICCV, ECCV, NeurIPS, AAAI, ACMMM, TPAMI, IJCV, TIP, TMM, etc.

Wenhan Luo is currently an Associate Professor with Sun Yat-sen University. Prior to that, he worked as a research scientist for Tencent and Amazon. He has published over 40 papers in top conferences and leading journals, including ICML, CVPR, ICCV, ECCV, ACL, AAAI, ICLR, TPAMI, IJCV, AI, TIP, etc. He also has been reviewer, senior PC member and Guest Editor for several prestigious journals and conferences. His research interests include several topics in computer vision and machine learning, such as image/video synthesis, image/video quality restoration, reinforcement learning. He received the Ph.D. degree from Imperial College London, UK, 2016, M.E. degree from Institute of Automation, Chinese Academy of Sciences, China, 2012 and B.E. degree from Huazhong University of Science and Technology, China, 2009.

Xiaoqin Zhang received the B.Sc. degree in electronic information science and technology from Central South University, China, in 2005 and Ph.D. degree in pattern recognition and intelligent system from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, in 2010. He is currently a professor in Wenzhou University, China. His research interests are in pattern recognition, computer vision and machine learning. He has published more than 80 papers in international and national journals, and international conferences, including IEEE T-PAMI, IJCV, IEEE T-IP, IEEE T-IE, IEEE T-C, ICCV, CVPR, NIPS, IJCAI, AAAI, and among others.

Tong Lu received the Ph.D. degree in computer science from Nanjing University in 2005. He received his M.Sc. and B.Sc. degree from the same university in 2002 and 1993, respectively. He served as Associate Professor and Assistant Professor in the Department of Computer Science and Technology at Nanjing University from 2007 and 2005. He is now a full Professor at the same university. He also has served as Visiting Scholar at National University of Singapore and Department of Computer Science and Engineering, Hong Kong University of Science and Technology, respectively. He is also a member of the National Key Laboratory of Novel Software Technology in China. He has published over 60 papers and authored 2 books in his area of interest, and issued more than 20 international or Chinese invention patents. His current interests are in the areas of multimedia, computer vision and pattern recognition algorithms/systems. Dr. Tong Lu was a member of ACM, IAPR, ISAI and a senior member of China Computer Federation (CCF). He is the Youth Associate Editor of Journal on Frontiers of Computer Science (FCS), and has served as the Secretary-general of CAD&CG Committee of Jiangsu Computer Federation in China since 2008. He has been member of the program committee or session chair of more than 10 international scientific conferences, and the Chair of Organization Committee of Youth Scholar Forum of State Key Laboratory for Novel Software Technology since 2010.