ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing

Zhongze Wang Haitao Zhao* Jingchao Peng Lujian Yao Kaijie Zhao
East China University of Science and Technology, Shanghai, China
{zzwang, lujianyao, kjzhao}@mail.ecust.edu.cn, haitaozhao@ecust.edu.cn, starry-sky@outlook.com

Abstract

Unpaired image dehazing (UID) holds significant research importance due to the challenges in acquiring haze/clear image pairs with identical backgrounds. This paper proposes a novel method for UID named Orthogonal Decoupling Contrastive Regularization (ODCR). Our method is grounded in the assumption that an image consists of both haze-related features, which influence the degree of haze, and haze-unrelated features, such as texture and semantic information. ODCR aims to ensure that the haze-related features of the dehazing result closely resemble those of the clear image, while the haze-unrelated features align with the input hazy image. To accomplish the motivation, Orthogonal MLPs optimized geometrically on the Stiefel manifold are proposed, which can project image features into an orthogonal space, thereby reducing the relevance between different features. Furthermore, a task-driven Depth-wise Feature Classifier (DWFC) is proposed, which assigns weights to the orthogonal features based on the contribution of each channel's feature in predicting whether the feature source is hazy or clear in a self-supervised fashion. Finally, a Weighted PatchNCE (WPNCE) loss is introduced to achieve the pulling of haze-related features in the output image toward those of clear images, while bringing haze-unrelated features close to those of the hazy input. Extensive experiments demonstrate the superior performance of our ODCR method on UID.

1. Introduction

Haze is a common atmospheric phenomenon caused by the accumulation of aerosol particles. It can cause severe quality degradation of images, which can affect subsequent computer vision tasks [27, 50]. The degradation of haze effect can be described as the atmospheric scattering model (ASM) [14, 33]:

$$\mathbf{I}(x) = \mathbf{J}(x)t(x) + \mathbf{A}(1 - t(x)) \tag{1}$$

where $\mathbf{I}(x)$, $\mathbf{J}(x)$, $t(x)$ and $\mathbf{A}$ stand for the hazy image, the clear image, the transmission map (T-map) and the global atmospheric light separately.

To enhance image clarity and detail, numerous image dehazing methods have been introduced. Early image dehazing methods [2, 3, 16, 18, 58] are mainly based on handcrafted priors. These methods conduct statistical analyses on hazy and clear images to acquire prior knowledge for image enhancement, but have limited robustness due to specific assumptions. With the development of deep neural networks, many deep learning dehazing methods [6, 13, 24, 40, 46, 48] have emerged, which train networks in a supervised manner on large-scale synthetic datasets, resulting in significantly improved dehazing performance.

Despite their impressive performance on synthetic data, the real-world applicability of these methods is limited by the challenge of acquiring paired images with identical backgrounds. Consequently, research into training dehazing models on unpaired datasets is gaining traction. Most current unpaired image dehazing (UID) strategies [9, 15, 47] adopt the Cycle-GAN framework [57], which constructs hazy-clear-hazy and clear-hazy-clear conversion cycles. These approaches depend on cycle-consistency loss to ensure content consistency across the dehazing process. However, the cycle-consistency loss assumes a bijective relationship between the two domains [35], which is too strict for image dehazing. In real world, a clear image can correspond to a hazy image with varying degrees of haze within the same scene.

To bypass this bijection limitation, CUT-like methods [32, 35, 44] have been introduced, eschewing the Cycle-GAN architecture for a singular GAN framework. These methods preserve consistency by maximizing mutual information between the features of a query patch in the dehazed output and the corresponding patch in the original hazy input, as depicted in Fig. 1 (a). Nonetheless, this approach incurs a contradiction between maximizing mutual information and attaining effective dehazing, and do not fully utilize the guiding role of clear images for dehazing.

To achieving background consistency and bridging the

*Corresponding author.

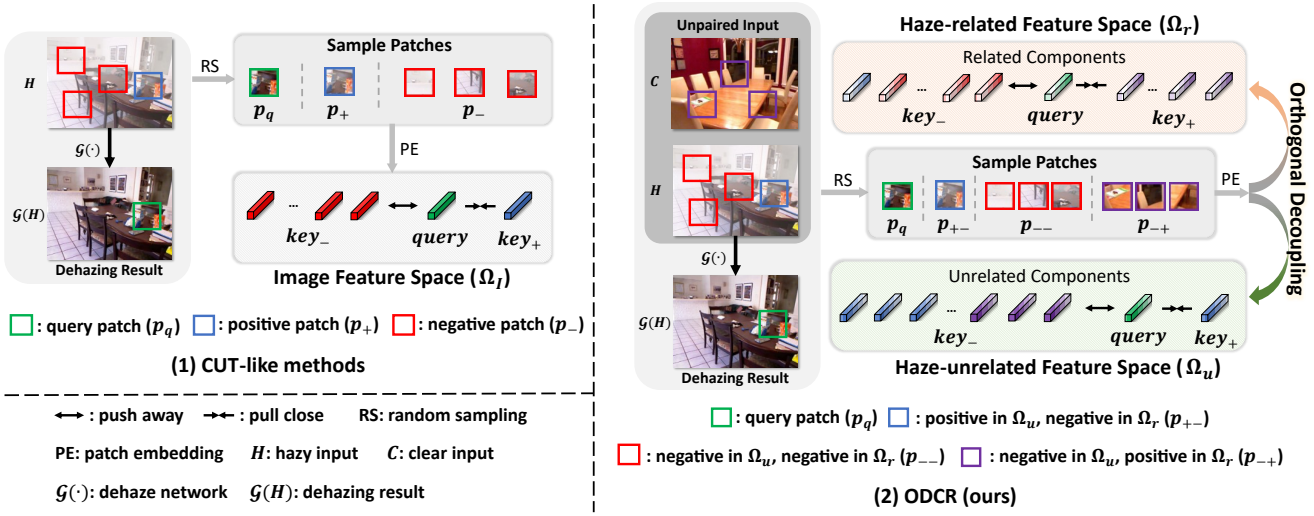


Figure 1. Illustration of how (a) the CUT-like methods and (b) ODCR work. CUT-like methods directly pull features of the query patch and positive patch close, leading to a contradiction in maximizing the mutual information between the two patches and dehazing. In ODCR, orthogonal decoupling are proposed to decouple image features to haze-related (describing haze level) and unrelated (describing non-haze information, such as semantic and texture) components. Then the mutual information between query and the positive patches in different feature spaces are maximized, thus avoiding the above contradiction.

gap between hazy and clear domains, one feasible idea involves the decoupling of query patch features into haze-related components, which quantify the level of haze, and haze-unrelated components, embodying non-haze attributes like semantics and texture. The objective is to maximize the mutual information between corresponding haze-related components of the query patch and clear image patches, as well as between haze-unrelated components of the query patch and the hazy patch at the same location.

The above feature decoupling faces two challenges. The first is **how to decouple the image features to components with low relevance.** The inherent blending of haze-related and unrelated features underlies the conflict between maximizing mutual information and effective dehazing. Decoupling into components of reduced relevance is crucial to resolving this conflict. According to the ASM [14, 33], hazy images are captured by deep coupling of multiple physical quantities, making it difficult to completely realize the decoupling of features. And the second challenge is **how to assign the decoupled features as haze-related/-unrelated components.** Without the guidance of ground truth images, it becomes difficult for networks to distinguish between features pertaining to haze and those that do not.

To address above challenges, a novel Orthogonal Decoupling Contrastive Regularization (ODCR) is proposed for UID in this paper. In ODCR, we first repartition the samples according to different feature spaces as shown in Fig. 1 (b). To solve the first challenge, we propose to introduce orthogonal constraint, which is widely used in traditional machine learning [5, 45, 51] and deep learning [12, 22, 30, 36], to decouple of image features into components with low relevance. To address the second challenge, a self-supervised Depth-wise Feature Classifier (DWFC) for mapping image features to hazy or clear labels is introduced. DWFC yields heat vectors that reflect the significance of each channel in discerning whether the feature is extracted from a hazy or clear image can be obtained. Based on the sample repartition and heat vectors, a Weighted PatchNCE (WPNCE) loss is proposed to realize the pulling of related/unrelated features in different feature space.

In summary, our main contributions are as follows:

- The proposed ODCR projects image features to orthogonal space by Orthogonal-MLPs which are geometrically optimized on the Stiefel manifold to reduce the relevance between features.
- A self supervised DWFC is proposed to assign orthogonal features to haze-related and unrelated components, which provides weights indicating the degree of relevance of each channel to haze is proposed.
- A Weighted PatchNCE is proposed to maximize the mutual information between the corresponding components of query and positive samples in different feature spaces.

2. Related Works

Unpaired Image Dehazing. Considering the difficulty of obtaining large-scale data for supervised training, some methods [9, 10, 15, 47, 53] focus on learning mappings that restore hazy images to clear images from unpaired data. Zhu *et al.* [57] first propose Cycle-GAN network based on cycle-consistency loss to solve the unpaired image-to-image (i2i) problem. Engin *et al.* [15] proposed a Cycle-GAN-

like method, which combines cycle-consistency and perceptual loss for UID. Afterwards, some Cycle-GAN-like UID methods [9, 10, 47, 53] are proposed, among which D^4 proposed by Yang *et al*. [47] realizes density-awareness by decomposing the transmission map in ASM into haze density and background depth, and then achieves excellent results on multiple datasets. Chen *et al*. [9] propose CDD-Net combining adversarial contrastive loss and cycle-consistency loss for extracting task-relevant and -irrelevant information. However, these Cycle-GAN-like methods are all based on the bijection assumption between the haze and clear domains [35], which is overly strict as a scene may correspond to multiple levels of haze.

To avoid the problems associated with the bijection assumption, Park *et al*. [35] propose CUT, which maintains consistency by maximizing the mutual information between patches at the same location in the input and output. However, in image dehazing, the contradiction arises between maximizing the mutual information between the hazy-clear patch pair and the requirement that the output be a haze-free image. Subsequent CUT-like methods [32, 44] do not focus on this contradiction either. Unlike these methods, our ODCR mitigates the contradiction by decoupling features into haze-related and unrelated components, individually aligning them with the corresponding components of the clear image and the original input, respectively.

Orthogonal Constraint. Orthogonal constraints play a pivotal role in diminishing feature relevance and curtailing redundant information in traditional machine learning [5, 45, 51] and deep learning [12, 22, 30, 36]. While reduced rank Procrustes rotation [52] and eigen decomposition [29] address orthogonal constraint problems in traditional machine learning, these methods are not applicable within deep learning frameworks. For deep learning methods with orthogonal constraint, one way is to convert the problem to an unconstrained one using Lagrange multipliers. However, the above method views the problem as a "black box" and it is hard to take advantage of orthogonal spaces [34]. Some methods [12, 30] include orthogonal regularizations in loss function, which do not guarantee that the parameters are in orthogonal space. In contrast to the aforementioned methods, our ODCR proposes to solve the orthogonal constraint problem using geometric optimization on the Stiefel manifold and thus perform a strict orthogonal decoupling of image features.

Contrastive Learning for Image Dehazing. Contrastive learning [7, 8, 17, 19–21, 49] has shown power in high-level self-supervised representation learning tasks. In recent years, contrastive learning are also applied to low-level image enhancement tasks [11, 42, 43, 46, 55, 59]. Wu *et al*. [46] first used contrastive learning in image dehazing and the proposed AECR aligns the features of the generated query image with those of the ground truths at pixel level. Zheng *et al*. [55] propose to additionally bound the solution space with results of other methods, considering that the lower bound of the solution space is always far away from positive samples. The above methods are applications of contrastive learning on supervised image dehazing. Chen *et al*. [9] introduce adversarial contrastive learning in Cycle-GAN network to disentangle task-relevant and -irrelevant factors for UID. However, the combination of GAN and adversarial contrastive learning makes the training process unstable. CUT-like methods [32, 35, 44] are based on PatchNCE, which is a contrastive regularization that pulls the query patch in the output close to the patch at the same location in the hazy input, leading background consistency. But it produce a contradiction of pulling the query to hazy or clear. In ODCR, we propose a Weighted PatchNCE (WPNCE), which avoids the contradiction by maximizing the mutual information of haze-related/-unrelated components of query and key sample features, respectively.

3. Orthogonal Decoupling Contrastive Regularization

In training process, given two unpaired clear image set $\mathcal{X}_C = \{C_i\}_{i=1}^{N_C}$ and hazy image set $\mathcal{X}_H = \{H_j\}_{j=1}^{N_H}$, a couple of unpaired images $\{C, H\}$ is input and we aim to train a generator $\mathcal{G}$ that can output a clear image $\mathcal{G}(H)$, which has a haze degree that converges to a clear image and keeps the haze-unrelated information such as image texture and semantics consistent with H.

3.1. Sample Repartition

In the CUT-like methods depicted in Fig. 1, a query patch p_q from the generated dehazed image is paired with a corresponding patch p_+ at the same location in the input hazy image as a positive sample, while other patches in the input image serve as negative samples p_-. This approach to sample partitioning, however, exhibits two critical limitations. First, it overlooks the influence of the clear domain, which is essential for restoring the haze level in the output to that of a clear image. Second, it creates an inherent conflict in determining whether to align the haze level of p_q close to that of p_+, when attempting to increase the mutual information between the positive sample pairs.

To overcome the identified limitations, we propose a refined strategy for partitioning the positivity and negativity of sample patches. We assume that the features of a patch contain both haze-related features describing the haze level and haze-unrelated features containing image texture, semantics. For any given patch, its haze-related and unrelated components are separately classified as positive or negative within their respective feature spaces, as illustrated in Fig. 1 (b). We adopt a dual subscript system to categorize the nature of samples. The first subscript indicates positivity or negativity in terms of haze-unrelated features: a patch at

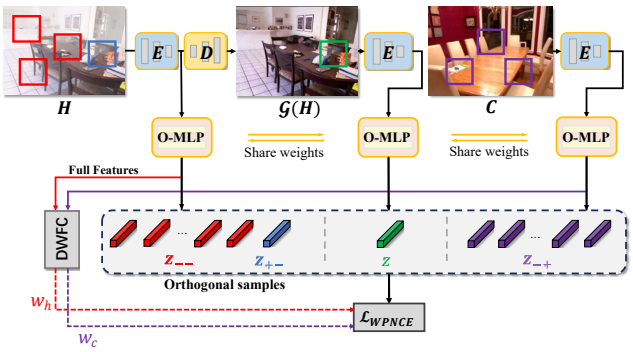


Figure 2. The pipeline of the proposed ODCR.

the same location as the query patch in the hazy domain H is deemed positive, while all others are negative. The second subscript signifies positivity or negativity concerning hazy or clear: patches in H are negative, whereas those in the clear domain C are positive. This methodology results in distinct notational representations for various patches, as elaborated below:

- p_{+-}: the patch in H with the same position as p_q;
- p_{-+}: all patches in C;
- p_{--}: all patches other than p_{+-} in H.

3.2. Orthogonal Decoupling

In this subsection, we provide an introduction to how ODCR achieves orthogonal decoupling and solve the two chanllenges mentioned in Sec. 1.

3.2.1 Orthogonal Projection

Orthogonal MLP. To achieve decoupling of haze-related and unrelated features, the relevance between the two kind of features needs to be reduced. Therefore, a MLP with orthogonal constraints is proposed to project the image features into the orthogonal space to reduce the relevance between the features:

$$z_k = \mathcal{H}_\Theta(\mathcal{G}^i_{enc}(p_k)), s.t. \Theta^T\Theta = I \tag{2}$$

where $\mathcal{H}_\Theta(\cdot)$ stands for the MLP with orthogonal constraint and Θ stands for its parameter matrix. $\mathcal{G}^i_{enc}(\cdot)$ represents the feature of the i-th encoder layer in the generator $\mathcal{G}$.

To solve the problem with orthogonal constraints, one way is to convert it to an unconstrained problem using Lagrange multipliers [4]. However, the method views the problem as a "black box" and it is hard to take advantage of orthogonal spaces [34]. Some methods [12, 30] include orthogonal regularizations in loss functions, which cannot guarantee that the parameters are in orthogonal space. Therefore, we propose to solve the orthogonal constraint problem using geometric optimization on the Stiefel manifold and thus perform a strict orthogonal decomposition of the features.

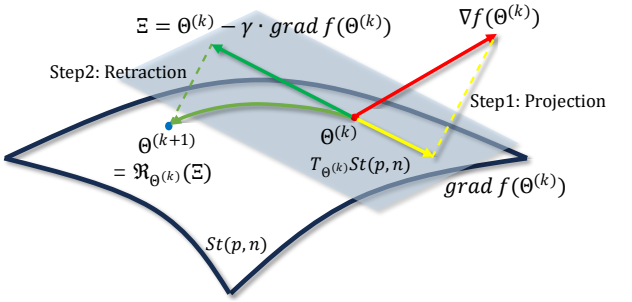


Figure 3. The illustration of the geometric optimization on the Stiefel manifold.

Geometric Optimization on the Stiefel Manifold. A Stiefel manifold is a set containing all orthogonal matrices in a specified space, i.e. $St(p,n) \triangleq \{\Theta \in \mathbb{R}^{n\times p} : \Theta^T\Theta = I_p\}$. Its tangent space at a point Θ can be defined as: $T_\Theta St(p,n) \triangleq \{Z \in \mathbb{R}^{n\times p} : \Theta^T Z + Z^T\Theta = 0\}$. For our orthogonal decomposition problem, the ideal approach is to find the optimal solution of $\mathcal{H}_\Theta$ on the Stiefel manifold.

Assuming that $f(\Theta)$ is a loss function defined in the Euclidean space and $\nabla f(\Theta)$ is its gradient in the Euclidean space, it cannot be optimized directly using optimizers such as SGD [38] and ADAM [23], but requires an additional two process. Denote the Riemannian gradient $grad\ f(\Theta)$ as the tangent vector gradient of $f(\cdot)$ on the tangent space at point Θ:

$$grad\, f(\Theta) = \nabla f(\Theta) - \frac{1}{2}\Theta\Theta^T\nabla f(\Theta) - \frac{1}{2}\Theta\nabla f(\Theta)^T\Theta \tag{3}$$

which points to the direction where the loss function $f(\cdot)$ on the Stiefel manifold ascends steepest and it can be proved by the following theorem:

Theorem 1. Given $\nabla f(\Theta)$ in Euclidean space and $grad\ f(\Theta)$ defined by Eq. 3, $grad\ f(\Theta)$ is the orthogonal projection of $\nabla f(\Theta)$ onto the tangent space of the Stiefel manifold. And the first step is:

Step 1: Project the gradient of the Euclidean space onto the tangent space the Stiefel manifold. And the updating process on the tangent space $T_{\Theta^{(k)}}St(p,n)$ at iteration k+1 is:

$$\Xi = \Theta^{(k)} - \gamma grad\, f(\Theta^{(k)}) \tag{4}$$

where γ is the gradient update step size. After that, the point Ξ on the tangent space need to be remapped onto the Stiefel manifold based on the following theorem:

Theorem 2. Given a point $\Xi = \Theta^{(k)} - \gamma grad\ f(\Theta^{(k)})$ on the tangent space $T_{\Theta^{(k)}}St(p,n)$ of a point on a Stiefel manifold, there is a retraction operation:

$$\mathfrak{R}_{\Theta^{(k)}}(\Xi) = (\Theta^{(k)} + \Xi)(I + \Xi^T\Xi)^{-\frac{1}{2}} \tag{5}$$

$$\mathfrak{R}_{\Theta^{(k)}}(\Xi)^T\mathfrak{R}_{\Theta^{(k)}}(\Xi) = I \tag{6}$$

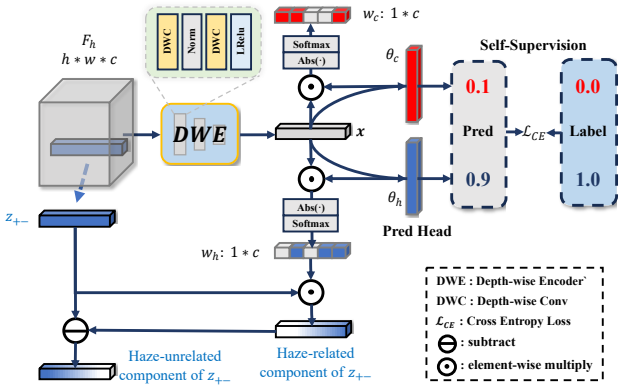


Figure 4. The structure of DWFC and the procedure for obtaining the heat-vector describing the haze relevance of features.

i.e., $\mathfrak{R}_{\Theta^{(k)}}(\Xi)$ is on $St(p,n)$ and the second step is:

Step 2: Map Ξ to $St(p,n)$ via Eq. 5. Then the updated parameter matrix under orthogonal constraints is:

$$\Theta^{(k+1)} = \mathfrak{R}_{\Theta^{(k)}}(\Theta^{(k)} - \gamma grad\, f(\Theta^{(k)})) \tag{7}$$

Fig. 3 illustrates the process of updating the parameters on the manifold. The proofs of Theorems 1 and 2 are detailed in the Supplementary Material.

3.2.2 Depth-wise Feature Classifier

Provided the features projected into orthogonal space, it is not yet known which are haze-related and which are haze-unrelated. To solve this problem, we introduce a Depth-wise Feature Classifier (DWFC). DWFC takes the image features of H or C extracted by $\mathcal{G}_{enc}$ as input to predict whether the feature source is a hazy or clear image. With this self-supervised approach, one channel-wise heat-vector can be obtained for each set of input features.

Fig. 4 illustrates the structure of DWFC. Given an orthogonal feature, it is input into the Depth-wise Encoder (DWE) and processed through 3 convolutional blocks. The processed feature is globally average pooled (GAP) to obtain a one-dimensional feature vector. The feature vector is fed to a fully connection (FC) layer to get the final classification prediction probability. Note that we use depth-wise convolution to avoid information exchange between channels to ensure that each value of the feature vector is only relevant to the corresponding channel.

Since the source of the feature is known, we use it as a label and take cross-entropy loss as a objective function to optimize the DWFC:

$$\begin{aligned}\mathcal{L}_{CE} =& y_h log(\theta_h^T x) + (1-y_h) log(1-\theta_h^T x) \\ &+ y_c log(\theta_c^T x) + (1-y_c) log(1-\theta_c^T x)\end{aligned} \tag{8}$$

where y_h and y_c denote the labels of feature source. If the source of input feature is hazy image, then $y_h = 1$ and $y_c = 0$. θ_h and θ_c represent the weights of the fully connection layer (pred head) of DWFC, and x represents the 1-D feature vector. Thus $\theta_h^T x$ and $\theta_c^T x$ stand for the prediction that the source of the feature is a hazy or clear image.

Inspired by visualization methods in high-level computer vision tasks [39, 41, 56], we argue that the absolute values in the results of the element-wise multiplication between x and θ_h (or θ_c) reflect the magnitude of the role played by the feature of the corresponding channel in the network's decision that the source of the features is hazy (or clear). Thus the heat-vectors describing the haze (or clear) relevance can be formulated as:

$$w_h = softmax(abs(\theta_h \odot x)) \tag{9}$$

$$w_c = softmax(abs(\theta_c \odot x)) \tag{10}$$

where $abs(\cdot)$ denotes a function taking absolute values for all elements in the input vector, and $softmax(\cdot)$ stands for the softmax function.

For example, if the absolute value of an element in $\theta_h \odot x$ is large, it can be assumed that the feature of the corresponding channel prompt (or inhibit) the network's judgment that the feature source is a hazy image, i.e., the feature of the channel is inclined to be a hazy-related (or unrelated) feature. Specifically, for features from hazy (or clear) images, we assign them with w_h (or w_c).

3.3. Weighted PatchNCE

Based on the sample partition in Sec. 3.1 and the heat-vectors in Sec. 3.2.2, Weighted PatchNCE (WPNCE) for UID is proposed. WPNCE is a loss function based on mutual information between features, and we first give the definition of the weighted mutual information of two feature vectors:

$$l(w, z_1, z_2) = exp(w \odot {z_2}^T \times z_1/\tau) \tag{11}$$

where w is the weight, τ is the temperature coefficient and z_1 and z_2 represent the two feature vectors for computing the mutual information. For WPNCE, the mutual information between the query patch and the positive components of all other key patches are desired to be maximized, which can be denoted as:

$$\mathcal{P} = l(w_h, z, z_{+-}) + \sum_{n=1}^{N_{-+}} l(w_c, z, {z_{-+}}^n) \tag{12}$$

and minimize the mutual information with the negative components:

$$\begin{aligned}\mathcal{N} =& l((\mathbf{1}-w_h), z, z_{+-}) + \sum_{n=1}^{N_{--}} l(\mathbf{1}, z, {z_{--}}^n) \\ &+ \sum_{n=1}^{N_{-+}} l((\mathbf{1}-w_c), z, {z_{-+}}^n)\end{aligned} \tag{13}$$

Table 1. Quantitative comparison of ODCR with the state-of-the-art image dehazing methods on several datasets. Best results are **bolded** and second best results are underlined. Cells where results are not available are replaced by "-". The latency is measured on 256×256 images using a single RTX 4090 GPU.

	Method	SOTS-indoor [25]		SOTS-outdoor [25]		NH-HAZE 2 [1]		Overhead	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	#Param (M)	Latency (ms)
Paired	DehazeNet [6]	19.82	0.818	24.75	0.927	10.62	0.521	0.009	0.919
	AOD-Net [24]	20.51	0.816	24.14	0.920	12.33	0.631	0.002	0.390
	MSCNN [37]	19.84	0.833	14.62	0.908	11.74	0.566	0.008	0.619
	GDN [31]	**32.16**	**0.983**	17.69	0.841	12.04	0.557	0.956	9.905
Unpaired	DCP [18]	13.10	0.699	19.13	0.815	14.90	0.668	-	-
	CycleGAN [57]	21.34	0.898	20.55	0.856	13.95	0.689	11.38	10.22
	CycleDehaze [15]	20.11	0.854	21.31	0.899	14.12	0.701	11.38	10.19
	YOLY [26]	15.84	0.819	14.75	0.857	13.38	0.595	32.00	-
	USID-Net [28]	21.41	0.894	23.89	0.919	15.62	0.740	3.780	31.01
	RefineDNet [54]	24.36	0.939	19.84	0.853	14.20	0.754	65.80	248.5
	D^4 [47]	25.42	0.932	25.83	0.956	14.52	0.709	10.70	28.08
	CUT [35]	24.30	0.911	23.67	0.904	15.92	0.758	11.38	10.06
	ODCR (ours)	26.32	0.945	**26.16**	**0.960**	**17.56**	**0.766**	11.38	10.14

where $\mathbf{1}$ is a vector of the same shape as w_c or w_h with all elements 1. Finally, we integrate them into a loss function in the form of InfoNCE.

$$\mathcal{L}_{WPNCE} = -log(\frac{\mathcal{P}}{\mathcal{P}+\mathcal{N}}) \tag{14}$$

And the full objective is as follows:

$$\mathcal{L} = \mathcal{L}_{GAN} + \mathcal{L}_{WPNCE} + \mathcal{L}_{CE} + \mathcal{L}_{idt} \tag{15}$$

where $\mathcal{L}_{GAN}$ and $\mathcal{L}_{idt}$ stand for the GAN loss and identity loss in CUT [35].

4. Experiments

4.1. Datasets and Metrics

We conduct experiments on several datasets to evaluate the performance of our method on UID. The datasets include RESIDE [25], NH-HAZE 2 [1] and Fattal's [16]. The test sets cover synthetic, artificial, and real-world images.

RESIDE is a widely used image dehazing dataset containing several subsets. We use ITS (13990 pairs of indoor images) from RESIDE as the training set and SOTS (500 indoor and 500 outdoor image pairs) as the test set. NH-HAZE 2 is an artificial dataset for the NTIRE 2021 competition, which consists of 25 pairs of non-homogeneous hazy images and clear images. And Fattal's dataset is a real-world dataset that includes 41 real hazy images in various scenes. Commonly used image quality evaluation metrics: PSNR (dB) and SSIM are employed to evaluate the dehazing performance of ODCR.

4.2. Performance Evaluation

We compare ODCR with several SOTA image dehazing methods. Some of them are trained using paired data, including DehazeNet [6], AOD-Net [24], MSCNN [37] and GDN [31]. Others do not require paired data, including DCP [18], CycleGAN [57], CycleDehaze [15], YOLY [26], USID-Net [28], RefineDNet [54], D^4 [47] and CUT [35]. Notably, we follow the evaluation strategy of D^4 [47]. We train our model only on ITS and test on all the test sets and all other methods also follow this strategy for fairness.

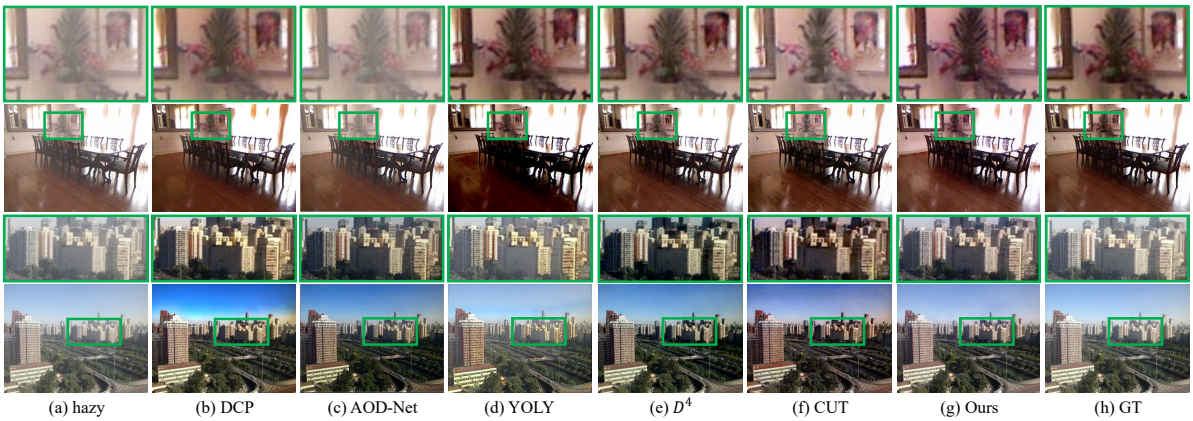
Quantitative Evaluation. The quantitative comparison of different methods is recorded in Table 1. Our ODCR obtains the second-best results on the SOTS-indoor dataset, and the best results are obtained by GDNet, which is a supervised method. However, testing the same model on the SOTS-outdoor and NH-HAZE 2 datasets, which are not trained accordingly, ODCR outperforms all paired and unpaired data-based methods, including GDNet. The results above demonstrate that ODCR is able to learn the patterns of haze-free images with excellent generalization by decoupling the image features into haze-related and unrelated components and performing contrastive learning separately.

Qualitative Evaluation. The qualitative comparison of synthetic and artificial datasets between various methods is displayed in Fig. 5. The supervised method AOD-Net [24] fails to dehaze in local area. DCP [18], YOLY [26] and D^4 [47] suffer from color distortion. Compared with CUT [35], which only performs contrastive learning on the hazy patches, ODCR shows improved overall dehazing performance while maintaining haze-unrelated information.

In addition, the visual comparison of the different methods on the real-world dataset Fattal's is shown in Fig. 6. It obviously illustrates that ODCR better removes haze from the boxed region compared to other methods and has minimal image quality degradation including artifacts, loss of structural details, and color distortion, verifying the dehazing effectiveness of ODCR on real-world images.

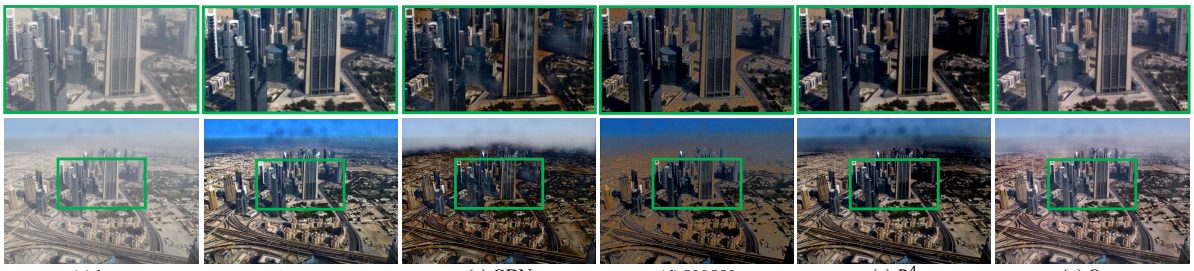
4.3. Ablation Studies and Discussion

Effectiveness of Orthogonal Decoupling. We propose an O-MLP with orthogonal constraint to minimize feature em-



(a) hazy (b) DCP (c) AOD-Net (d) YOLY (e) D^4 (f) CUT (g) Ours (h) GT

Figure 5. Visual comparison of various dehazing methods on SOTS-indoor [25] and SOTS-outdoor [25]. Areas where our method works better are boxed out and zoomed in, or you can zoom in by yourself to get a better view.



(a) hazy (b) DCP (c) GDN (d) YOLY (e) D^4 (g) Ours

Figure 6. Visual comparison of various dehazing methods on Fattal's dataset [16]. Areas where our method works better are boxed out and zoomed in, or you can zoom in by yourself to get a better view.

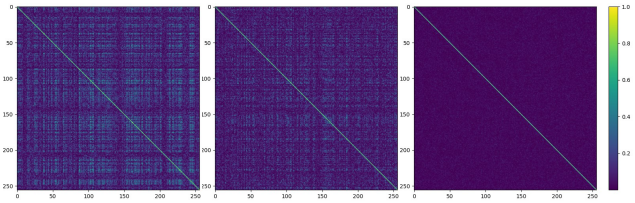
bedding relevance, effectively decoupling haze-related and unrelated features. Our quantitative analysis contrasts three distinct cases: the absence of Orthogonal Decoupling (OD), the imposition of an orthogonal loss function, and optimization conducted on the Stiefel manifold. The comparative results, as detailed in Table 2, affirm that the Stiefel manifold optimization outperforms the other approaches in dehazing performance.

To further substantiate the O-MLP's capacity to attenuate feature relevance, we examine the cosine similarity matrices of the feature embeddings. Fig. 7 delineates the inter-channel cosine similarity matrices derived from the feature projections in each case. It indicates a substantial redundancy in features without OD. While the orthogonal loss function offers a reduction in feature relevance, it does not achieve optimal separation. In stark contrast, O-MLP, when optimized on the Stiefel manifold, demonstrates a marked minimization in feature relevance, which we attribute to its strict orthogonality. This geometric optimization is pivotal in achieving the desired feature decoupling necessary for effective dehazing.

Effectiveness of DWFC. We evaluate the effectiveness of the proposed DWFC. DWFC assigns haze-related and unrelated features through a self-supervised mechanism leveraging the heat-vector output. Another feasible way is to statically determine a fixed ratio of orthogonal features as haze-related or unrelated. As Table 3 demonstrates, irrespective of the predetermined ratio of related to unrelated features, the dehazing capability is significantly compromised. Fig. 8 also presents a qualitative comparison of dehazing out-

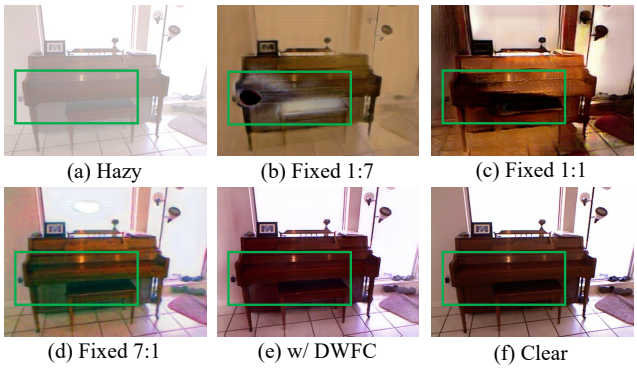
Table 2. Ablation study on the effectiveness of O-MLP.

Setting	Metric	
	PSNR (dB)	SSIM
w/o OD	22.30	0.887
w/ orthogonal loss	24.96	0.928
w/ O-MLP (ours)	**26.32** (+1.36)	**0.945** (+0.017)



(a) w/o OD (b) w/ orthogonal loss function (c) Ours

Figure 7. Visualization of the similarity matrix with different settings.



(a) Hazy (b) Fixed 1:7 (c) Fixed 1:1

(d) Fixed 7:1 (e) w/ DWFC (f) Clear

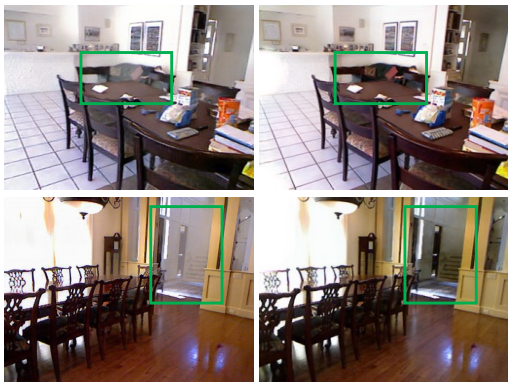
Figure 8. Qualitative comparison on the assignment of related/unrelated features. (a) is the hazy input. (b)-(d) are the dehazing results of the model with different ratio of assignment. (e) is the result of the default network. And (f) is the corresponding clear image.

comes under different configurations. Fixed feature ratios lead to a notable decline in image quality. The degradations above stem from the absence of the DWFC's self-supervision, causing features unrelated to haze that should be mapped to the hazy domain to be mistakenly aligned with the clear domain during contrastive learning. Consequently, this misalignment distorts information channels, such as texture and semantics, critical for accurate dehazing. In addition, the gradual improvement of the dehazing performance as the percentage of haze-unrelated feature increases in Table 3 and Fig. 8 also proves the above view.

Effectiveness of WPNCE. Experiments are conducted on the loss functions to verify the effectiveness of our proposed WPNCE. Compared to PatchNCE, WPNCE exploits the haze-related component in clear images to drive the generated images close to being clear. The quantitative and qualitative comparison between the results of using PatchNCE and WPNCE is recorded in Table 4 and Fig. 9, respectively. Better dehazing results are obtained using WPNCE compared to using PatchNCE, due to the fact that WPNCE treats haze-related and unrelated features differently and avoids the contradiction of dehazing and maximizing the mutual information between the generated image features and the haze patch features.

Table 3. Ablation study on the feature assignment method. Note that Ratio in the table represents assigning orthogonal features in a related:unrelated ratio. For example, when ratio is 1:7, the first 32 channels of a 256-channel orthogonal feature are assigned as related features, and the last 224 channels are unrelated features.

Method	Ratio	Metric	
		PSNR (dB)	SSIM
Fixed	1:7	13.10	0.599
	1:3	13.93	0.574
	1:1	15.33	0.621
	3:1	14.30	0.660
	7:1	18.28	0.726
DWFC (ours)	N/A	**26.32** (+8.04)	**0.945** (+0.219)



(a) w/ $\mathcal{L}_{PatchNCE}$ (b) w/ $\mathcal{L}_{WPNCE}$

Figure 9. Qualitative comparison of PatchNCE and WPNCE.

Table 4. Ablation study on the effectiveness of $\mathcal{L}_{WPNCE}$.

Setting	Metric	
	PSNR (dB)	SSIM
$\mathcal{L}_{PatchNCE}$ [35]	24.30	0.911
$\mathcal{L}_{WPNCE}$ (ours)	**26.32** (+2.02)	**0.945** (+0.034)

5. Conclusion

We propose Orthogonal Decoupling Contrastive Regularization for UID by decoupling image feature into haze-related/-unrelated components. In particular, we repartition the patch samples in terms of haze-related/-unrelated. Afterward, an orthogonal MLP geometrically optimized on the Stiefel manifold is introduced to reduce the relevance between the features by projecting the features of each sample into the orthogonal space. Furthermore, a Depth-wise Feature Classifier for assigning the projected features of each channel as haze-related/-unrelated is proposed. Finally, a novel Weighted PatchNCE is designed to maximize the mutual information between the corresponding components of query and positive samples in different feature spaces. Experiments conducted on synthetic and real-world datasets validate our proposal and analysis. Performance improvements are observed when comparing other SOTA methods for UID.

Acknowledgement.
This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62173143.

References

[1] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, and Radu Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 627–646, 2021. 6

[2] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016. 1

[3] Dana Berman, Tali Treibitz, and Shai Avidan. Single image dehazing using haze-lines. *IEEE transactions on pattern analysis and machine intelligence*, 42(3):720–734, 2018. 1

[4] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014. 4

[5] Rasmus Bro and Age K Smilde. Principal component analysis. *Analytical methods*, 6(9):2812–2831, 2014. 2, 3

[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 1, 6

[7] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 132–149, 2018. 3

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3

[9] Xiang Chen, Zhentao Fan, Pengpeng Li, Longgang Dai, Caihua Kong, Zhuoran Zheng, Yufeng Huang, and Yufeng Li. Unpaired deep image dehazing using contrastive disentanglement learning. In *European Conference on Computer Vision*, pages 632–648. Springer, 2022. 1, 2, 3

[10] Xiang Chen, Yufeng Li, Caihua Kong, and Longgang Dai. Unpaired image dehazing with physical-guided restoration and depth-guided refinement. *IEEE Signal Processing Letters*, 29:587–591, 2022. 2, 3

[11] Xiang Chen, Jinshan Pan, Kui Jiang, Yufeng Li, Yufeng Huang, Caihua Kong, Longgang Dai, and Zhentao Fan. Unpaired deep image deraining using dual contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2017–2026, 2022. 3

[12] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask aet with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2553–2562, 2021. 2, 3, 4

[13] Qili Deng, Ziling Huang, Chung-Chi Tsai, and Chia-Wen Lin. Hardgan: A haze-aware representation distillation gan for single image dehazing. In *European Conference on Computer Vision*, pages 722–738. Springer, 2020. 1

[14] RM Drake and JE Gordon. Mie scattering. *American Journal of Physics*, 53(10):955–962, 1985. 1, 2

[15] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018. 1, 2, 6

[16] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):1–14, 2014. 1, 6, 7

[17] Tengda Han, Weidi Xie, and Andrew Zisserman. Self-supervised co-training for video representation learning. *Advances in Neural Information Processing Systems*, 33:5679–5690, 2020. 3

[18] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 1, 6

[19] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927, 2019. 3

[20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[21] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670*, 2018. 3

[22] Zhengwei Hu, Jingchao Peng, and Haitao Zhao. Dynamic neural orthogonal mapping for fault detection. *International Journal of Machine Learning and Cybernetics*, 12:1501–1516, 2021. 2, 3

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[24] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 1, 6

[25] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 6, 7

[26] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Zitao Liu, Joey Tianyi Zhou, and Xi Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129:1754–1767, 2021. 6

[27] Chengyang Li, Heng Zhou, Yang Liu, Caidong Yang, Yongqiang Xie, Zhongbo Li, and Liping Zhu. Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

[28] Jiafeng Li, Yaopeng Li, Li Zhuo, Lingyan Kuang, and Tianjian Yu. Usid-net: Unsupervised single image dehazing network via disentangled representations. *IEEE transactions on multimedia*, 2022. 6

[29] Pei Li, Wenlin Zhang, Chengjun Lu, Rui Zhang, and Xuelong Li. Robust kernel principal component analysis with optimal mean. *Neural Networks*, 152:347–352, 2022. 3

[30] Sun-Ao Liu, Yiheng Zhang, Zhaofan Qiu, Hongtao Xie, Yongdong Zhang, and Ting Yao. Learning orthogonal prototypes for generalized few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11319–11328, 2023. 2, 3, 4

[31] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 6

[32] Xiaotong Luo, Wenjin Yang, Yuan Xie, and Yanyun Qu. Farewell to cyclegan: Single gan with decoupled constraint for unpaired image dehazing. *Available at SSRN 4620491*. 1, 3

[33] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York*, 1976. 1, 2

[34] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999. 3, 4

[35] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 1, 3, 6, 8

[36] Jingchao Peng, Haitao Zhao, and Zhengwei Hu. Second-order component analysis for fault detection. *Journal of Process Control*, 108:25–39, 2021. 2, 3

[37] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016. 6

[38] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016. 4

[39] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 5

[40] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 1

[41] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020. 5

[42] Tao Wang, Guangpin Tao, Wanglong Lu, Kaihao Zhang, Wenhan Luo, Xiaoqin Zhang, and Tong Lu. Restoring vision in hazy weather with hierarchical contrastive learning. *Pattern Recognition*, page 109956, 2023. 3

[43] Yongzhen Wang, Jiamei Xiong, Xuefeng Yan, and Mingqiang Wei. Uscformer: Unified transformer with semantically contrastive learning for image dehazing. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 3

[44] Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. Ucl-dehaze: Towards real-world image dehazing via unsupervised contrastive learning. *IEEE Transactions on Image Processing*, 2024. 1, 3

[45] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. 2, 3

[46] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 1, 3

[47] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2037–2046, 2022. 1, 2, 3, 6

[48] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. *European Conference on Computer Vision*, 2022. 1

[49] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021. 3

[50] Zhengxi Zhang, Liang Zhao, Yunan Liu, Shanshan Zhang, and Jian Yang. Unified density-aware image dehazing and object detection in real-world hazy scenes. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 1

[51] Haitao Zhao, Pong Chi Yuen, and James T Kwok. A novel incremental principal component analysis and its application for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(4):873–886, 2006. 2, 3

[52] Haitao Zhao, Zhihui Lai, Henry Leung, Xianyi Zhang, Haitao Zhao, Zhihui Lai, Henry Leung, and Xianyi Zhang. Sparse feature learning. *Feature Learning and Understanding: Algorithms and Applications*, pages 103–133, 2020. 3

[53] Jingming Zhao, Juan Zhang, Zhi Li, Jenq-Neng Hwang, Yongbin Gao, Zhijun Fang, Xiaoyan Jiang, and Bo Huang. Dd-cyclegan: Unpaired image dehazing via double-discriminator cycle-consistent generative adversarial network. *Engineering Applications of Artificial Intelligence*, 82: 263–271, 2019. 2, 3

[54] Shiyu Zhao, Lin Zhang, Ying Shen, and Yicong Zhou. Refinednet: A weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, 30:3391–3404, 2021. 6

[55] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5785–5794, 2023. 3

[56] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 5

[57] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 1, 2, 6

[58] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11):3522–3533, 2015. 1

[59] Yunhao Zou and Ying Fu. Estimating fine-grained noise model via contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12682–12691, 2022. 3