

Generalizing to Out-of-Sample Degradations via Model Reprogramming

Runhua Jiang[✉] and Yahong Han[✉], Member, IEEE

Abstract— Existing image restoration models are typically designed for specific tasks and struggle to generalize to out-of-sample degradations not encountered during training. While zero-shot methods can address this limitation by fine-tuning model parameters on testing samples, their effectiveness relies on predefined natural priors and physical models of specific degradations. Nevertheless, determining out-of-sample degradations faced in real-world scenarios is always impractical. As a result, it is more desirable to train restoration models with inherent generalization ability. To this end, this work introduces the Out-of-Sample Restoration (OSR) task, which aims to develop restoration models capable of handling out-of-sample degradations. An intuitive solution involves pre-translating out-of-sample degradations to known degradations of restoration models. However, directly translating them in the image space could lead to complex image translation issues. To address this issue, we propose a model reprogramming framework, which translates out-of-sample degradations by quantum mechanic and wave functions. Specifically, input images are decoupled as wave functions of amplitude and phase terms. The translation of out-of-sample degradation is performed by adapting the phase term. Meanwhile, the image content is maintained and enhanced in the amplitude term. By taking these two terms as inputs, restoration models are able to handle out-of-sample degradations without fine-tuning. Through extensive experiments across multiple evaluation cases, we demonstrate the effectiveness and flexibility of our proposed framework. Our codes are available at <https://github.com/ddghjikle/Out-of-sample-restoration>.

Index Terms— Image restoration, out-of-sample degradation, model reprogramming, image-to-image translation.

I. INTRODUCTION

IMAGE restoration have achieved remarkable success with the rapid development of deep neural networks. Previous researches [1], [2], [3] always characterize specific types of degradations as individual issues and propose dedicated solutions. While this methodology effectively addresses some real-world scenarios, it falls short when considering more complex situations like autonomous driving on rainy days. In such cases, the perceived images can be degraded by a combination of rain, haze, blur, and noise, making it difficult

Manuscript received 26 June 2023; revised 14 December 2023 and 15 February 2024; accepted 6 March 2024. Date of publication 5 April 2024; date of current version 12 April 2024. This work was supported in part by NSFC under Grant 62376186 and Grant 61932009 and in part by the Chinese Association for Artificial Intelligence (CAAI)-Huawei MindSpore Open Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nam Ik Cho. (*Corresponding author: Yahong Han*)

The authors are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: ddghjikle@tju.edu.cn; yahong@tju.edu.cn).

Digital Object Identifier 10.1109/TIP.2024.3378181

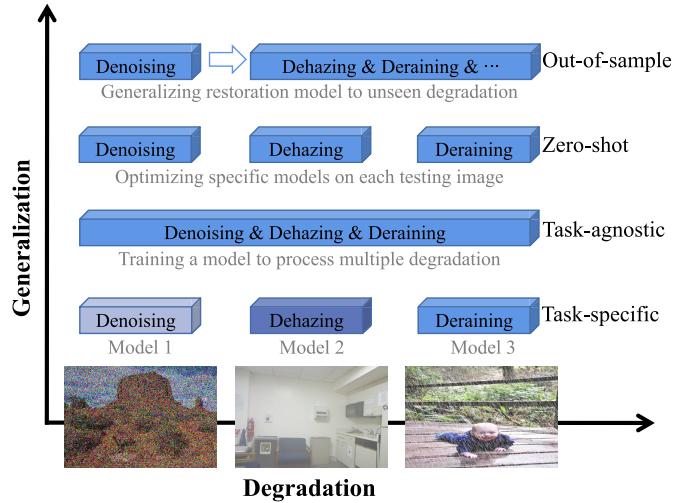


Fig. 1. The introduced out-of-sample restoration task is to develop models with the capability of handling unknown degradations. It extends previous restoration researches as paying more attention to cross-degradation generalization.

to attribute the degradation to a specific form. Moreover, as real-world images can exhibit diverse and unpredictable degradation patterns, enumerating all possible degradations to train restoration networks is practically infeasible [4]. Therefore, it is imperative to develop restoration models that can effectively restore images degraded by various factors, including those not encountered during training.

In contrast to supervised approaches focused on removing specific degradations, some studies explored the challenges of handling multiple degradations [1], [2] or generalizing from synthetic images to real-world scenarios [5]. To name a few, Liang et al. [1] introduced Swin Transformer layers with residual connections to remove low-resolution, noisy, and compression artifacts. The authors of [2] proposed an efficient multi-head attention and feed-forward network for capturing long-range pixel interactions. Their experiments demonstrate that Transformer models, irrespective of degradation categories, can effectively perform when training and testing sets exhibit the same degradation. Conversely, other study [5] highlights the domain gap between synthetic and real-world degraded images, leading to a significant decrease in performance. They assume that both synthetic and real-world domains suffer from the same degradation, and perform unsupervised learning on the real-world domain. Recently, this motivation has been extended by zero-shot methods [6], [7], which assume the domain gap exists between training and

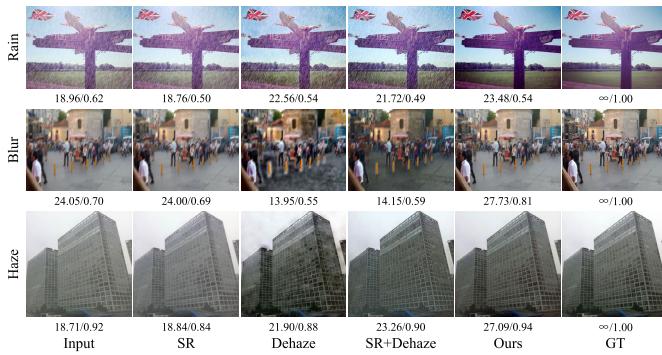


Fig. 2. Visual examples of the out-of-sample degradation issue. These degraded images are restored by Super-resolution (SR), Dehaze, SR+Dehaze networks, and our methods. PSNR and SSIM of each example are provided for better comparison.

testing sets rather than between synthetic and real-world training samples. To address this issue, zero-shot methods require prior knowledge of degradation categories, allowing them to fine-tune models for each testing sample. However, none of existing methods consider the challenging and practical circumstances wherein testing samples are affected by degradations not presented in the training samples.

In this work, we introduce the Out-of-Sample Restoration (OSR) task, which aims to develop restoration models capable of generalizing to unknown degradations. Fig. 1 illustrates differences between this task and previous researches. Specifically, task-specific methods [1], [3], [8] focus on establishing image-to-image translation network for a certain kind of degradation, while task-agnostic methods [4] combine various types of degraded images to train a single restoration network. However, as Fig. 2 shows, while the task-specific network Dehaze effectively eliminates haze degradation, it struggles to address out-of-sample degradations such as blur and rain. Furthermore, the task-agnostic network SR+Dehaze, trained on datasets encompassing super-resolution and dehazing networks, fails to exhibit improved performance on rainy examples. Although zero-shot researches [6], [7] can address this issue by fine-tuning restoration models on testing samples with out-of-sample degradations, they require prior knowledge of degradation categories. In contrast, the OSR task aims to learn generalizable models from a limited set of training samples, making it different and complementary to previous researches.

While the OSR task shares similar motivation with domain adaptation task, which addresses the challenge of multi-domain discrepancy, existing domain adaptation methods prove ineffective in this context. The reason is that current domain adaptation methods are specifically designed for semantic-related tasks, such as image classification [9], [10] and semantic segmentation [11]. For example, Long et al. [9] embedded semantic representations of all task-specific layers into a kernel Hilbert space. Saito et al. [10] proposed to align the distributions of source and target domains by leveraging the semantic information within the classification boundaries. Subsequent studies, such as those by Li et al. [12] and Han et al. [13], divided feature space of classification networks according to their semantic discriminability or

domain invariance. However, compared with semantic information, image content is more susceptible to degradation, resulting in complex discrepancies across different degradations.

An intuitive solution of the OSR task is to translate the out-of-sample degradations into degradations recognizable by restoration models, with the expectation that these models can then generate clear outputs. However, performing this translation in image space could raise complex image-to-image translation issues. To address this issue, we propose a novel approach by reprogramming restoration models using quantum mechanim. By this way, images are represented as wave functions comprising amplitude and phase terms. The amplitude encapsulates real-value features that represent the image content, while the phase modulates style vectors within a continuous space [14], [15]. These representations allow for efficient translation by aligning continuous phase vectors, addressing the complexities inherent in image-to-image translation. Consequently, our proposed framework is implemented with input and output transform functions. The input transform function utilizes phase-aware vision MLPs to disentangle input images into wave functions and align the phase intensity. Reprogrammed models can then efficiently translate the recognizable degradations into the style of clear images, and enhance content details within the amplitude component. Finally, the output transform function remaps these wave functions back into the image space. Through extensive experiments, it is demonstrated that our framework outperforms previous domain adaptation and image restoration methods in terms of generalization ability and restoration effectiveness.

Main contributions of this work can be summarized as follows:

- We introduce a new task, *i.e.*, Out-of-Sample Restoration (OSR), which requires models to handle a wide range of degradations, including those not encountered during training. This task complements previous restoration tasks and focuses on developing models with cross-degradation generalization ability.
- To tackle the OSR task, we propose a model reprogramming framework that represents images as wave functions incorporating both amplitude and phase. This framework effectively adapts restoration models to out-of-sample degradations by enhancing the amplitude element and transforming the phase term. To the best of our knowledge, this is the first attempt to combine model reprogramming with image restoration studies.
- Extensive experiments are conducted by using noisy, hazy, blurry, rainy, and low-resolution images. By comparisons with existing domain adaptation and image restoration methods, effectiveness and flexibility of the proposed framework are comprehensively verified.

II. RELATED WORK

In this section, we first provide an overview of image restoration. Then, the concepts of domain adaptation and model reprogramming are introduced as both of them are relevant to this work.

A. Image Restoration

Image restoration is to remove degradation artifacts from input images. Previous works [16], [17] characterize image restoration as different tasks due to the complexity of real-world scenarios. For example, the image dehazing task aims at removing the haze degradation and enhance illumination appearances [3], [18]. The image deblurring task is to recover sharp image details from blurry artifacts. Learning-based models, such as convolution neural networks [3], [19], [20], [21] and Transformers [1], [2], have achieved remarkable success in addressing specific restoration tasks. Recent researches have also explored multi-task restoration, *i.e.*, recovering multiple degradations through a single network [4]. However, both task-specific and task-agnostic methods often struggle with degradations that are unseen during training, highlighting the need for more research on generalizable restoration models. Consequently, the zero-shot restoration task has been defined to transfer restoration models directly on testing images [6], [22]. However, few attention has been paid to train generalizable restoration models with limited types of training images. In addition, zero-shot restoration does not fully consider the influences caused by the transferred models and the target degradation. To address these issues, we introduce the Out-of-Sample Restoration (OSR) task, which provides a comprehensive understanding of the generalization ability of restoration models.

B. Domain Adaptation

Domain adaptation aims at improving the performance of networks on target domains by transferring knowledge from different but related source domains. One popular methodology is to learn effective representations that are robust to domain discrepancy [9], [10]. To name a few, Chen et al. [25] introduced the concepts of feature transferability and discriminability through spectral analyses, demonstrating that eigenvectors with large singular values contribute to robustness against domain discrepancies. Latter, authors of [12] found that feature alignment should concentrate on semantic information for avoiding negative influences from semantic-irrelevant features. In contrast, Han et al. [13] considered the parameter space rather than the feature space. The authors demonstrated that different parts of network parameters are respectively related to transferability and discriminability. Thus, transferability-related parameters should be emphasized for learning domain-invariant features. However, existing domain adaptation methods are primarily designed for visual understanding tasks, *e.g.*, image classification and semantic segmentation. As a result, they tend to perform poorly when it comes to learning transferable restoration models, motivating our exploration of the OSR task and its solutions.

C. Model Reprogramming

Model reprogramming refers to repurposing a pre-trained model from a source domain to solve tasks in a target domain without fine-tuning the model [26]. It is originally cast as adversarial reprogramming due to the implication that the task in the target domain can be different from pre-trained

tasks [27]. To achieve this, an input transformation module is needed to map target-domain inputs to the input space of the pre-trained source model. Tsai et al. [28] introduced the zero-padding operation with positional index for input transformation. During reprogramming, only the padded pixels can be changed. However, the performance of the repurposed model is influenced by the spatial size of the padded pixels. To alleviate this limitation, methods such as [29] employ trainable encoders as the input transformation module. The trainable input transformation modules also enable cross-modal reprogramming. For example, Neekhara et al. [30] achieved cross-modal transformation using a set of trained tokens, successfully repurposing ImageNet classifiers for topic classification. Similarly, the Voice2Series proposed in [31] reprograms acoustic models for time series classification. Compared to the input transformation module, the output mapping layer is simpler due to that previous methods most consider classification tasks. Therefore, the output mapping can be achieved by manually defined strategies [28] or trainable linear layers [30]. In contrast to above methods, this work extends the model reprogramming community by introducing model reprogramming to the OSR task and defining the transform functions as wave-inspired disentanglement and fusion operations.

III. PROPOSED APPROACH

A. Task Formulation

As discussed previously, the out-of-sample restoration task aims to develop restoration models that can effectively handle degradations beyond training samples. To formally define this task, we consider five widely studied degradations: noise [32], haze [18], blur [33], [34], [35], rain [36], and low-resolution (LR) [37], [38]. Then, the introduced OSR task can be specified as obtaining models capable of effectively addressing all five degradations without training on each of them individually.

In general, there are two possible solutions for tackling this task. On the one hand, neural networks, equipped with abundant parameters, possess the capacity to learn substantial knowledge from the available training samples. However, naively optimizing these over-parameterized networks easily over-fits training samples [13]. In addition, due to distribution discrepancies shown in Fig. 3, the augmentation of available degradations within training samples does not necessarily guarantee robust generalization to out-of-sample degradations. On the other hand, restoration models exhibit the capability to formulate mapping functions between specific types of degraded and clear images. Therefore, handling out-of-sample degradations becomes feasible after translating them into degradations that are recognizable by the restoration models. Nevertheless, the direct translation of degraded images introduces intricate image-to-image translation issues, a challenge compounded by the diverse discrepancies across multiple degradations. Essentially, this method involves the reprogramming of restoration models through pre-processing operations, *i.e.*, image-to-image translation. To navigate this translation challenge, we introduce the quantum mechanism and wave function, which are illustrated in the subsequent sections.

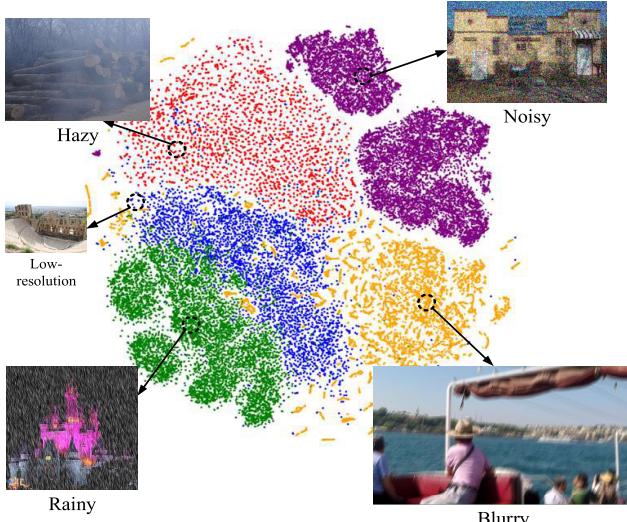


Fig. 3. Distributions of images suffered from haze, noise, blur, rain, or low-resolution. These visualization results are obtained by leveraging t-SNE [23] to visualize features generated by the perceptual function [24].

B. Overall Framework

To overcome the image-to-image translation challenge in reprogramming restoration models, we propose to represent input images by quantum mechanics and wave functions. Fig. 4 presents an overview of our proposed framework. First of all, the input transform module is designed following quantum mechanism, where entities are represented by wave functions comprising both amplitude and phase components [14], [39], [40]. The amplitude corresponds to a real-valued feature that represents the maximum intensity of the wave, while the phase term modulates intensity patterns by indicating locations of each point. This design allows for the decoupling of input images into continuous vectors of content and style, with the style representation aligned to recognizable degradations. Consequently, it is possible to utilize the reprogrammed model to approximate clear outputs. Second, the reprogrammed model aims to remove the degradations based on the input transform module. This involves mapping the style representation and enhancing the content details. Since no existing methods study the problem of reprogramming restoration models, two kinds of restoration models, *i.e.*, randomly initialized and specifically trained, are explored in subsequent experiments. Finally, after processing these two terms, the output transform function formulates wave functions and remaps them into the original image space to yield clear outputs. It is worth noting that our framework distinguishes itself from existing wave-based restoration methods [14], [15], which directly extract feature maps from wave-based representations. In contrast, the proposed framework directly leverages these representations.

C. Input Transform Module

According to quantum mechanism [14], [39], [40], an image can be represented as a wave \hat{z} with both amplitude and phase terms:

$$\hat{z} = |z| \odot e^{\theta i}, \quad (1)$$

where $|\cdot|$ and \odot are absolute value operation and element-wise multiplication. i denotes the imaginary unit satisfying $i^2 = -1$. The amplitude $|z|$ is a real-value feature representing image content, and the phase θ indicates distributions in the periodic function $e^{\theta i}$. In fact, the difference between amplitude $|z|$ and real-value feature in neural networks is the absolute operation $|\cdot|$. Therefore, this operation can be integrated into the phase term as follows:

$$|z| \odot e^{\theta i} = \begin{cases} z_t \odot e^{\theta_t i}, & z_t > 0, \\ z_t \odot e^{(\theta_t + \pi)i}, & \text{otherwise,} \end{cases} \quad (2)$$

where t denotes the position of each pixel. Therefore, the absolute operation can be removed and the amplitude can be obtained by a plain channel-FC operation with multi-channel inputs $[x_1, x_2, \dots, x_n]$:

$$\begin{aligned} z &= \text{Channel-FC}(x_i, W^c), \\ &= W^c x_i, i = 1, 2, \dots, n, \end{aligned} \quad (3)$$

where W^c is the weight with learnable parameters. Regarding the phase term θ , which indicates the style distribution of the input images, various strategies can be employed to estimate it. One straightforward method is to use trainable vectors to represent it. However, this method does not consider diversity between images with different degradations. Therefore, following [14], we also adopt a channel-FC layer to estimate θ .

Although above operations provide representations of the amplitude and phase terms, computations in the complex domain and the periodic function are challenging. Therefore, we further process these two terms using Euler's formulation and represent \hat{z} with real part and imaginary part:

$$\hat{z} = |z| \odot \cos\theta + i|z| \odot \sin\theta. \quad (4)$$

By doing so, a complex-value wave is represented as two real-value vectors, indicating the real and imaginary parts. As both of them are put through the reprogrammed model, these two vectors are defined with three channels and same spatial sizes as input images. Eq. 4 not only alleviates the computational difficulties but also relaxes the requirement of accurately decomposing the phase and amplitude components, allowing our framework to leverage their complementary relationships [3].

D. Output Transform Function

With the input transform module decouples and adapts style representations of input images. The reprogrammed model is able to remove recognizable degradation and recover content details. Given outputs of the reprogrammed model, output transform function aims to remap wave-like representations into image space. To this end, the output transform function is designed as shown in Fig. 4. First, the real and imaginary terms restored by reprogrammed model are taken as inputs. To leverage their complementary relationships, these two terms are put through Eq. 4, resulting in wave representations that approximate clear images. By considering that the reprogrammed models tend to recover these two terms by different degrees, both of them and the wave representation are dynamically fused by a fully-connected layer. Owing to the matrix

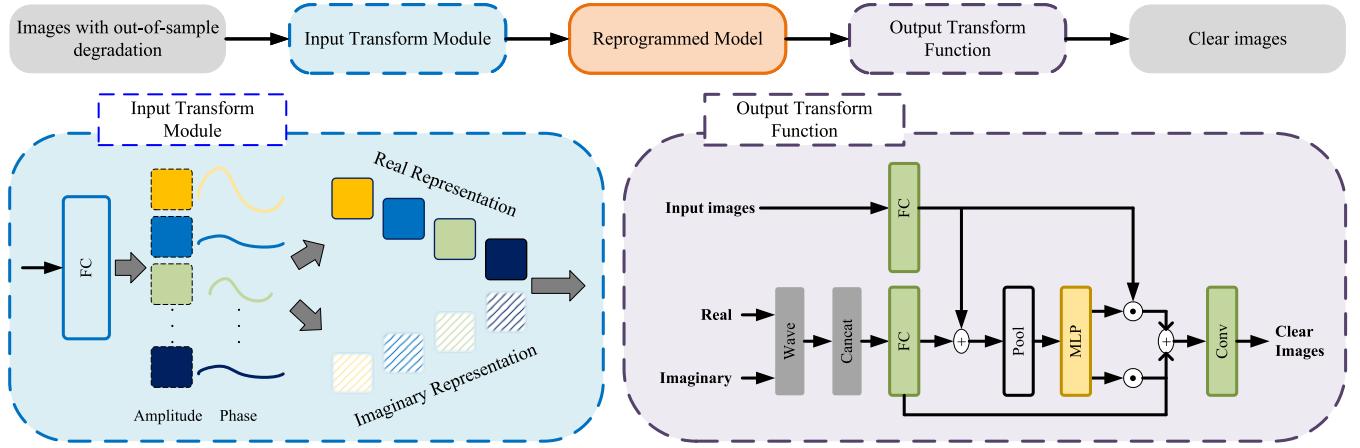


Fig. 4. Overview of the proposed model reprogramming framework. The input transform module represents input images as amplitude and phase terms. The amplitude term is real-value feature representing image content, and the phase term modulates image styles. A restoration model is used to enhance the amplitude term and align the phase term. Finally, wave functions of processed amplitude and phase are formed and mapped to image space. Dashed rectangles in the top part represent trainable modules.

multiplication with learnable parameters, the wave representation can be further enhanced with more details and realistic colors. After that, the wave representation is dynamically added with input images to form residual information between degraded inputs and clear ground-truth. The added results are first processed by a fully-connected layer for channel-wise aggregation. Then, output features are spatially aggregated by MLP modules with two token-FC operations [14]. Formally, the token-FC operation can be defined as

$$\begin{aligned} o &= \text{token-}FC(x_i, W^t), \\ &= W^t \odot x_i, i = 1, 2, \dots, n, \end{aligned} \quad (5)$$

where o indicates the output features. Above aggregation estimates pixel-wise consistencies between wave representation and input images. Therefore, results of the MLP modules are leveraged as dynamic weights to fuse wave representation and input images. Except the lightweight architecture, above functions are designed based on the quantum mechanism, leading to more effective restoration for out-of-sample degradations.

For optimizing the input transform module and output transform function, the smooth L_1 loss and the perceptual loss are computed:

$$\mathcal{L} = L_S + \lambda L_P, \quad (6)$$

where λ equals to 0.04 as [19] does. The smooth L_1 loss provides a quantitative measure of the difference between output images \hat{x} and clear images X . It can be defined as:

$$\begin{aligned} L_S &= \frac{1}{N} \sum_{x=1}^N \sum_{i=1}^3 F_S(J_i(\hat{x}) - J_i(X)), \\ F_S(e) &= \begin{cases} 0.5e^2, & \text{if } |e| < 1, \\ |e| - 0.5, & \text{otherwise,} \end{cases} \end{aligned} \quad (7)$$

where $J_i(\cdot)$ denote values of the i th channel of pixels in the output and clear images. N is the total number of pixels. The perceptual loss [24] quantifies the visual difference between

output images and their clear versions:

$$L_P = \sum_{j=1}^3 \frac{1}{C_j H_j W_j} \|\phi_j(\hat{x}) - \phi_j(X)\|_2^2, \quad (8)$$

where $\phi_j(\cdot)$ denotes three feature maps from VGG16 network [24]. C_j , H_j , and W_j specify the dimensions of $\phi_j(\cdot)$.

IV. EXPERIMENT

This section provides experimental analyses and comparison results. To be specific, we first illustrate datasets used for LR, rain, noise, blur, and haze degradations. Then, implementation details and evaluation protocols are introduced. After that, experiments about the OSR task and the proposed framework are discussed. Finally, the proposed framework is compared to existing domain adaptation and image restoration methods to validate its effectiveness.

A. Dataset

In this work, we form the out-of-sample restoration task by five widely researched degradations: low-resolution [37], [38], rain [36], noise [32], blur [33], [34], [35], and haze [18].

1) *Low-Resolution*: Following previous researches [2], [37], we consider three upscaling factors, *i.e.*, $\times 2$, $\times 3$, and $\times 4$. Training samples are obtained by down-sampling images from the DIV2K [38] and Flickr2K [37] datasets, resulting in 10,650 training samples. For evaluation, we use the B100 [41] datasets, which have been processed with the same upscaling factors.

2) *Rain*: We utilize the Rain1200 [36] dataset, which contains images with three different levels of rain: heavy, medium, and light. This dataset includes 12,000 training images and 1,200 testing images. As these images have complex distributions across the different levels, we directly use these three kinds of images for both training and evaluation.

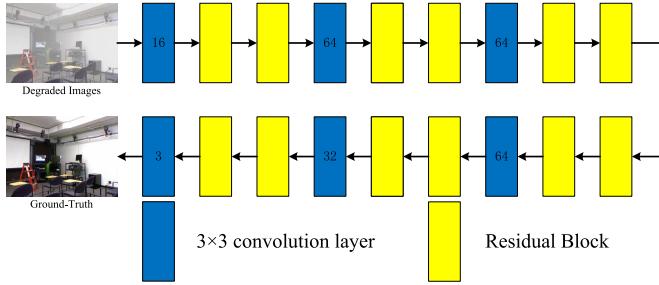


Fig. 5. Architecture of the Res12 model. Numbers in blue rectangles denote channels of output features.

3) *Noise*: To generate noisy images, we employ MIT-Adobe FiveK [32] dataset following the experimental setup in [42]. Specifically, we add additive Gaussian noise with standard deviations (σ) of 15, 25, and 50 to images in this dataset. As most compared methods do not take the standard deviation as input, noisy images with different σ are not separated. For comparison, we obtain testing samples by adding Gaussian noise with same standard variation to CBSD68 [43] datasets.

4) *Blur*: We utilize three popular datasets. The GoPro dataset [33] consists of 3,214 pairs of blurred images and ground-truth, captured from 33 sequences at a resolution of 1280×720 . It is divided into 2,103 pairs for training and 1,111 pairs for testing. The VideoDeblurring dataset [35] is a large-scale video deblurring dataset generated by different devices such as GoPro Hero 4 and iPhone 6s. Each video contains approximately 100 frames, and the frame size is 1280×720 . The HIDE dataset [34] includes two parts: HIDE I (1,304 long-shot pictures) and HIDE II (7,118 close-up pictures). We combine these three datasets to obtain 14,408 training samples. For quantitative evaluation, we compute metrics on the testing sets of the GoPro dataset and RealBlur [45] datasets.

5) *Haze*: The RESIDE dataset [18] contains both synthesized and real-world image pairs captured in indoor and outdoor scenarios. It includes five subsets: 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). We use 13,990 images from ITS as training samples. Additionally, we leverage 500 indoor and 500 outdoor images from the SOTS as testing samples.

B. Implementation

For the reprogrammed model, we adopt a cascade CNN termed as Res12. As Fig. 5 shows, it consists of six 3×3 convolution layers and 12 residual blocks. Each residual block comprises two convolution layers and a ReLU layer. The convolution layers map the input features to new dimensions such as 16, 64, or 32. By default, all convolution layers have a kernel size of 3×3 , and the stride and padding sizes are set to 1. Transform functions in the proposed framework are trained using RGB image patches of size 120×120 . For optimization, the Adam optimizer [46] is used with a batch size of 8 for training samples. The initial learning rate is set to 0.001, and reduced by half every 20 epochs. All trainable

TABLE I
PSNR METRICS OF RES12 AND SWINIR [1] TRAINED ON DIFFERENT KINDS OF DEGRADED IMAGES. THE BEST RESULTS ARE DENOTED BY BOLD FONTS

Train	Test	Res12					SwinIR [1]						
		LR	Rain	Noise	Blur	Haze	Avg	LR	Rain	Noise	Blur	Haze	Avg
	LR	22.40	20.67	18.89	25.35	12.12	19.89	22.73	19.82	18.89	25.26	12.01	19.74
	Rain	23.13	32.21	19.93	25.40	12.42	22.62	23.32	32.29	20.01	25.51	12.33	22.69
	Noise	23.60	22.63	19.58	25.43	12.36	20.72	23.52	22.38	20.01	25.47	11.99	20.67
	Blur	23.19	22.15	19.92	27.56	12.06	20.98	22.90	22.04	19.94	27.23	12.10	20.84
	Haze	15.94	21.49	16.64	17.07	27.83	19.79	17.90	22.46	17.68	19.76	29.03	21.37

parameters are optimized for 300 epochs, and all experiments are carried out on an NVIDIA GeForce GTX 2080Ti GPU via PyTorch and MindSpore [44] frameworks.

Due to degradation in the training set has important influences on model generalization, two evaluation protocols are leveraged in this work. First, the training set suffers from only one kind of degradation, and compared methods are evaluated on testing sets of all five degradations. Second, two kinds of degraded images are combined to obtain training samples, and all testing sets are leveraged to make comparison. To be more specific, according to baseline performance, 10,650 low-resolution images in DIV2K and Flickr2K, along with 13,990 images from the RESIDE dataset, are leveraged as training samples.

C. Baseline Performance of OSR

Since the OSR task is introduced in this work, our initial investigation focuses on assessing the performance of current CNN and Transformer models.

1) *Protocol 1*: First of all, we evaluate the baseline Res12 and popular Transformer model SwinIR [1], following the first evaluation protocol. This protocol employs only one kind of degraded images for training. As shown in Table I, metrics calculated on out-of-sample degradation are significantly worse than metrics on the diagonal. These results demonstrate that current CNN and Transformer both suffer from discrepancies among various degradations. In addition, although SwinIR outperforms Res12 when training and testing sets share the same degradation, their average performances remain comparable. This suggests that the abundant parameters within SwinIR does not necessarily guarantee better generalization ability. Similar results can be observed from Table II, where Res12 with different widths and depths tend to have similar performance. We also explore to implement Res12 by the MindSpore [44] framework. Corresponding experiments provide comparable performance than the PyTorch framework, e.g., achieving 2.58 dB improvement on hazy samples. Overall, above experiments highlight two findings: 1) discrepancies across multiple degradations considerably influence evaluation results, and 2) over-parameterized restoration networks do not consistently exhibit superior generalization abilities.

2) *Protocol 2*: As discussed earlier, training restoration models with multiple types of degraded images seems to be a plausible solution for OSR. Therefore, we explore whether diverse training samples contribute to improve average performance. To this end, we combine two out of five degradations as training samples and evaluate the baseline model Res12.

TABLE II

INFLUENCES CAUSED BY ARCHITECTURES OF RES12. FOR FAIR COMPARISON, TRAINING SAMPLES ARE TAKEN FORM THE RAIN DEGRADATION. THE BEST RESULTS ARE DENOTED BY BOLD FONTS

Width	LR	Rain	Noise	Blur	Haze	Avg
128	22.95	30.14	19.68	25.02	12.44	22.05
64	23.13	32.21	19.93	25.40	12.42	22.62
32	23.05	31.54	19.85	25.39	12.51	22.47
Depth	LR	Rain	Noise	Blur	Haze	Avg
24	19.78	32.03	16.96	22.19	14.55	21.10
12	23.13	32.21	19.93	25.40	12.42	22.62
6	23.09	31.26	19.81	25.35	12.42	22.39
Framework	LR	Rain	Noise	Blur	Haze	Avg
MindSpore [44]	23.13	32.28	19.93	25.40	15.00	23.15
Pytorch	23.13	32.21	19.93	25.40	12.42	22.62

TABLE III

BASELINE PERFORMANCE OF RES12 MODEL TRAINED WITH VARIOUS COMBINATIONS OF TRAINING IMAGES. THE BEST RESULTS ARE DENOTED BY BOLD FONTS

Test \ Train	LR	Rain	Noise	Blur	Haze	Avg
LR & Rain	23.07	31.77	19.90	25.51	12.16	22.48
LR & Noise	23.33	21.78	19.86	25.43	11.99	20.48
LR & Blur	23.01	22.12	19.81	27.24	12.04	20.84
LR & Haze	19.46	19.72	17.03	17.70	21.91	19.16
Rain & Noise	23.12	31.89	19.91	25.42	12.46	22.56
Rain & Blur	23.10	31.59	19.94	27.30	12.16	22.82
Rain & Haze	21.52	30.33	19.00	19.03	24.71	22.92
Noise & Blur	23.11	22.21	19.92	27.07	12.15	20.89
Noise & Haze	18.19	20.80	16.59	16.91	24.85	19.47
Blur & Haze	22.20	21.40	19.28	23.52	20.77	21.43

The results presented in Table III indicate that above operation does not ensure performance improvement. Notably, combining LR and Haze as training samples results in an average PSNR decrease of approximately 4% across five degradations. However, the best performance is observed when combining images affected by rain and haze, leading to a notable performance increase. Similar improvements are evident when Rain and Haze degradations are included in the training samples. These findings align with the earlier discussion on the OSR task, emphasizing the crucial effect of diverse discrepancies among different degradations in addressing this task.

D. Reprogramming Framework

With the second evaluation protocol and above baseline performances, here we validate effectiveness of the proposed framework.

1) *Translation Space*: As discussed in Sec III-A, although translating out-of-sample degradation into recognizable categories can empower restoration models handle out-of-sample degradation, conducting such translation in the image space inevitably introduces additional challenges. To answer this issue, our reprogramming framework decouples and translates out-of-sample degradation by quantum mechanism. Here we validate above discussion by experiments in Table IV. Firstly, following previous reprogramming research [30], we introduce an additional noise perturbation to translate input images with out-of-sample degradation. Compared to the baseline Res12, this translation method results in an average PSNR decrease

TABLE IV

RESULTS OF TAKING DIFFERENT TRANSLATION MODULES

Translation	Res12	Perturbation [30]	Convolution	ResBlock
Avg. PSNR	19.16	14.86	12.16	7.67

TABLE V

RESULTS OF LEVERAGING DIFFERENT COMPONENTS OF WAVE FUNCTIONS. THE BEST RESULTS ARE DENOTED BY BOLD FONTS

Component	Amplitude	Phase	Amplitude&Phase
Avg. PSNR	20.39	18.93	17.60
Componen	Imaginary	Real	Imaginary&Real
Avg. PSNR	18.50	17.70	20.84

TABLE VI

RESULTS OF TAKING DIFFERENT INPUT TRANSFORM MODULES

Methods	w/o Transform	ResBlock	WaveBlock [14]	Channel-FC
Avg. PSNR	18.21	18.91	19.77	20.84

of 4.3 dB. Moreover, cascade convolutions and residual blocks are implemented to translate out-of-sample degradations in the image space, denoted as the Convolution and ResBlock methods. Notably, the image-to-image translation of out-of-sample degradation significantly decreases overall performance, aligning with our earlier discussion. On the other hand, different components of the wave function are put through the reprogrammed model. As Table V shows, it is evident that solely utilizing imaginary or real representations yields inferior performance compared to the Imaginary&Real method. Moreover, the direct utilization of amplitude and phase components results in a degradation in performance compared to the Imaginary&Real approach, owing to the intricate computations involved in the complex domain and periodic functions. This trend persists even when taking only amplitude or phase as inputs. Furthermore, a noteworthy observation is the superior performance of amplitude over the phase term. This underscores the fact that the amplitude contains abundant content information from the ground-truth, whereas the phase term primarily captures image styles. Overall, above experimental analyses demonstrate the pivotal role of amplitude in encoding the majority of the ground-truth content, while the phase term is more focused on encoding image styles.

2) *Input Transform Module*: After demonstrating effectiveness of the imaginary and real representations, we proceed to verify the efficacy of our input transform module. To this end, we conduct experiments by removing this module or replacing it with residual blocks and wave blocks [14]. As Table VI indicates, the w/o Transform method, obtained by removing this module, results in worse performance than baseline Res12. The reason is that 1) the output transform function cannot perform effectively without real and imaginary representations provided by this module and 2) the reprogrammed model suffer from the cross-degradation discrepancies. Moreover, it is observed that both residual block and wave block fail to improve the average PSNR. Although both wave block and our module aim to obtain real and imaginary representations, the wave block further mixes these two representations as

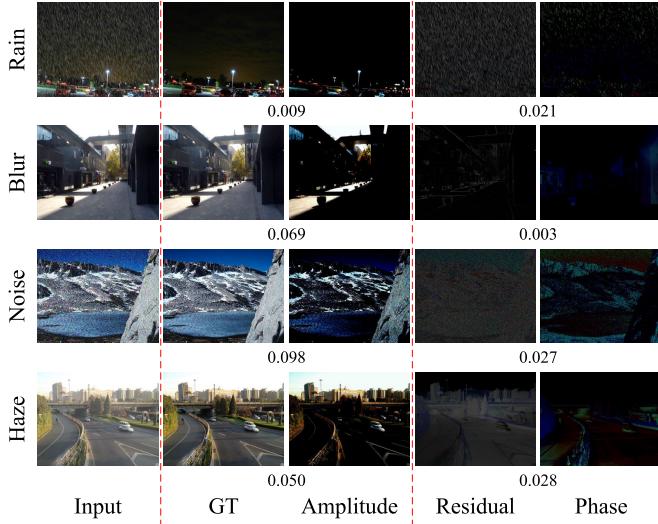


Fig. 6. Visualizations of the amplitude and phase terms. The mean square error (MSE) is provided to demonstrate that amplitude and phase terms reflect image content and degraded styles, respectively.

output features, presenting challenges in recovering image content and style from corresponding representations.

We further validate the input transform module through visualization results. Since the input images are generated by degrading the image content within the ground-truth (GT) using factors such as rain or blur, we utilize the GT and residual images, obtained by comparing the input and GT images, as benchmarks for assessing image content and degraded appearances. For clear comparison, we further measure the mean square error (MSE) between ground-truth and amplitude to verify their content consistency. On the other hand, since the residual image reflects influence caused by degradation factors, we evaluate the MSE between this residual image and the phase term to ascertain whether the phase term effectively captures the styles of different degradation. As Fig. 6 shows, the extracted amplitude and phase terms exhibit subtle distinctions and maintain a similar appearance to the GT and residual images. Therefore, it is verified that amplitude encodes the image content and phase encodes the image style. To demonstrate that our methods can adapt the phase term and enhance the amplitude term, we provide Fig. 8 to illustrate distributions of these two terms. By examining the ranges of axes in these sub-figures, it can be found that our methods generate phase and amplitude terms with significantly tighter distributions. This observation underscores the effectiveness of our approach in recovering out-of-sample degradations.

3) *Output Transform Function*: Next, we validate effectiveness of the output transform function. As can be observed from Table VII, removing the imaginary or real components both degrades effectiveness of the output transform function. In addition, results in the right part indicate that the average performance can be improved by employing more MLP layers. By repeating two MLP modules [14], the average performance increases to 21.38. However, when numbers of the MLP modules become 4, the average performance becomes worse than employing single MLP.

TABLE VII
RESULTS OF TAKING DIFFERENT OUTPUT TRANSFORM FUNCTIONS.
MLP * n MEANS n MLP MODULES ARE IMPLEMENTED
IN THIS FUNCTION

Components	w/o Imaginary	w/o Real	MLP	MLP * 2	MLP * 4
Avg. PSNR	19.05	17.90	20.84	21.38	20.24

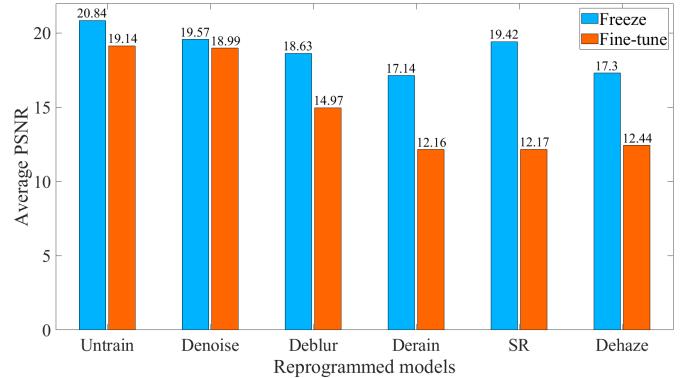


Fig. 7. Reprogramming untrained or trained Res12 by fine-tuning or freezing model parameters.

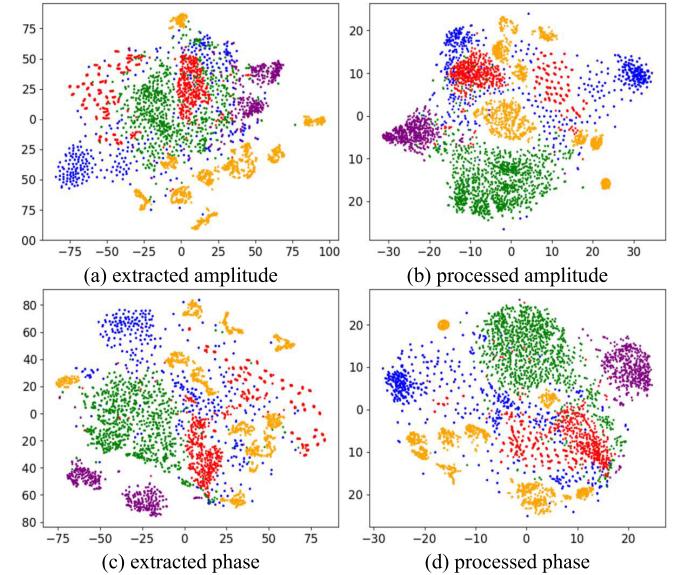


Fig. 8. T-SNE [23] results of amplitude and phase terms extracted from hazy, noisy, blurry, rainy, and low-resolution images. Components of these degraded images are represented by red, purple, gold, green, and blue dots, respectively. Please note that ranges of axes in (b) and (d) are much smaller than that of (a) and (c).

E. Reprogrammed Models

As our framework aims to reprogram restoration models to out-of-sample degradations, here we investigate influences of different reprogrammed models. To this end, two kinds of Res12 are leveraged as the reprogrammed model, *i.e.*, randomly initialized models and trained on one kind of degraded images (presented in Table I).

1) *Fine-tuned v.s. Frozen*: Fine-tuning pre-trained models on downstream tasks is a popular methodology in transfer learning. Therefore, we first validate whether fine-tuning is beneficial for reprogramming restoration models. From Fig. 7, it can be observed that for reprogramming restoration models, freezing model parameters proves to be more effective than

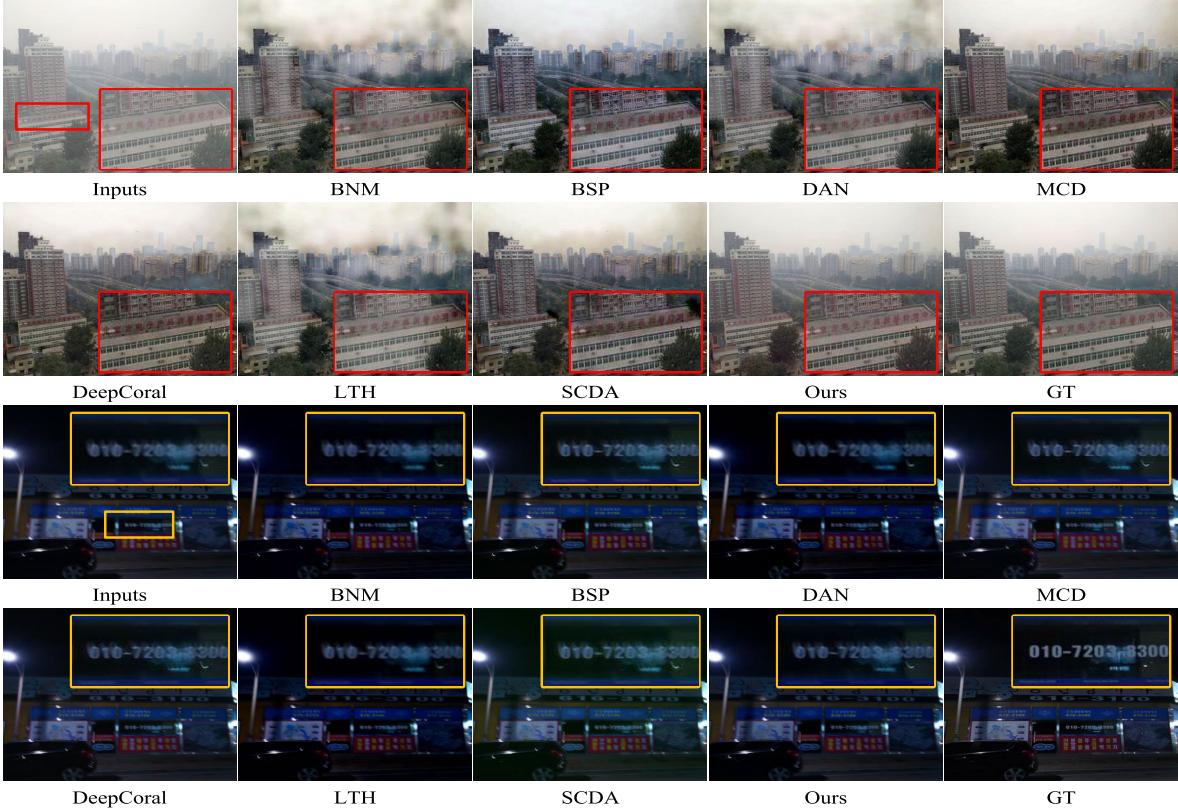


Fig. 9. Visual comparisons between the proposed framework and previous domain adaptation methods [9], [12], [13], [25], [49], [50] on image dehazing and deblurring tasks.

TABLE VIII

RESULTS OF REPROGRAMMING UNTRAINED MODELS WITH DIFFERENT INITIALIZATION METHODS

Initialization Methods	Xavier-Normal [47]	Xavier-Uniform [47]
Avg. PSNR	18.70	17.54
Initialization Methods	Kaiming-Normal [48]	Kaiming-Uniform [48]
Avg. PSNR	17.33	20.84

fine-tuning them. This effectiveness can be attributed to three reasons. First, model reprogramming and fine-tuning are two distinct tasks, with the latter allowing for the updating of model parameters. Second, during reprogramming, optimizing model parameters increases the difficulty in learning effective transformation modules. Third, discrepancies among out-of-sample degradations are challenging for fine-tuning. According to the experimental results, the restoration models are frozen during reprogramming.

2) *Trained v.s. Untrained*: Following previous model reprogramming researches [26], [30], here we explore differences between reprogramming trained and untrained restoration models. As depicted in Fig. 7, reprogramming untrained models yields superior performance. Although this seems incredible, similar results can also be found in [7], [26], and [30]. The underlying reason lies in the mechanism of model reprogramming. Untrained models possess unorganized knowledge that is not specific to any particular task, making them less robust against adversarial noises. Moreover, Table VIII shows untrained models with different initialization methods also have diverge performance. Xavier initialization [47] is derived from the assumption that the singular values of the Jacobian associated with each layer should close 1. Therefore, Glorot and Bengio [47] adopted a properly

TABLE IX

PERFORMANCE COMPARISONS WITH PREVIOUS DOMAIN ADAPTATION METHODS. THE BEST RESULTS ARE DENOTED BY BOLD FONTS

Method	LR	Rain	Noise	Blur	Haze	Avg
Res12	19.46	19.72	17.03	17.70	21.91	19.16
DAN [9]	18.19	22.54	16.69	17.48	27.94	20.57
DeepCoral [50]	20.36	20.33	17.75	18.59	20.09	19.42
MCD [10]	20.23	20.31	18.24	19.51	20.66	19.79
BSP [25]	20.34	20.27	18.30	18.72	20.60	19.65
BNM [49]	18.93	22.85	16.52	17.78	27.12	20.64
CICConv [11]	13.11	11.64	11.83	12.47	13.32	12.47
SCDA [12]	20.12	19.18	17.23	18.87	21.59	19.40
LTH [13]	18.49	22.52	15.95	17.81	26.32	20.22
Ours	22.04	22.47	19.41	22.51	17.75	20.84

scaled distribution. He et al. [48] find that this assumption is not suitable for non-linear layers. They thus proposed a robust initialization method that particularly considers the rectifier nonlinearities.

F. Comparisons

1) *Domain Adaptation Methods*: For evaluating generalization ability of the proposed framework, we first compare it with domain adaptation methods. Most of these methods are proposed as specific loss functions, which need access to both source and target samples. Therefore, these functions together with Eq. 6 are used to train the baseline Res12. Table IX presents corresponding results. It is evident that our framework is more suitable for restoration models. Compared with the latest method LTH [13], the proposed framework achieves better performance on out-of-sample degradations such as noise and blur. Fig. 9 further illustrates that previous

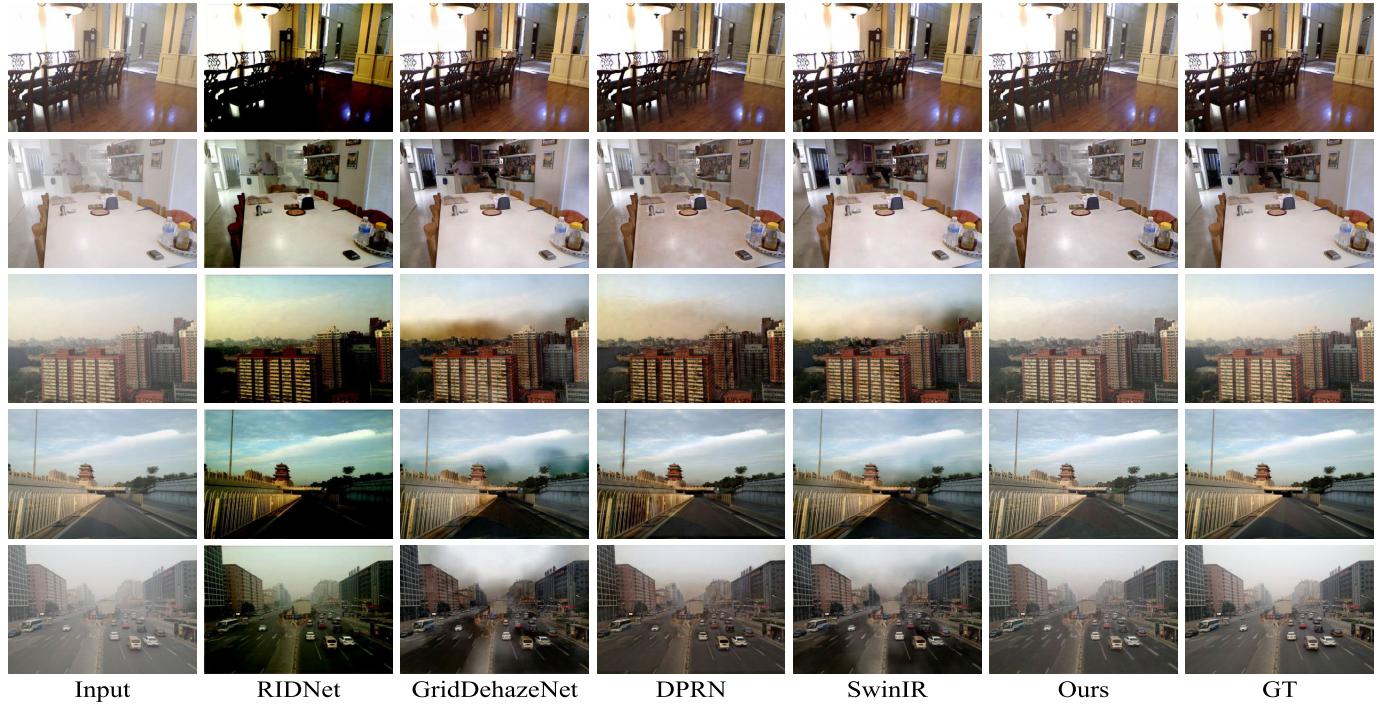


Fig. 10. Visual comparisons on the image dehazing task [1], [3], [19], [51]. The top two examples are indoor hazy images, while other examples are outdoor hazy images. It is worth noting that these outdoor hazy scenarios do not exist in training sets.

TABLE X

PERFORMANCE COMPARISONS WITH PREVIOUS METHODS FOLLOWING THE PROTOCOL 2, i.e., TRAINING ON HAZY AND LR IMAGES BUT EVALUATING ON FIVE KINDS OF DEGRADED IMAGES. THE BEST RESULTS ARE DENOTED BY BOLD FONTS. PARAM INDICATES TRAINABLE PARAMETERS ($\times 10^5$) OF EACH METHOD

Taxonomy	Method	LR	Rain	Noise	Blur	Haze	Avg	Param
CNN	Res12	19.46	19.72	17.03	17.70	21.91	19.16	7.39
	MPRNet [52]	16.36	19.17	14.58	16.46	21.91	17.70	157.41
	MIRNet [53]	15.63	22.87	16.67	18.30	28.68	20.43	317.88
Transformer	SwinIR [1]	16.63	22.69	17.32	18.52	29.56	20.94	19.05
	Restormer [2]	17.64	22.26	15.65	21.66	16.90	18.82	117.40
Attention	GridDehazeNet [19]	15.45	22.26	16.45	20.94	30.73	21.17	9.58
	RIDNet [51]	5.99	8.59	5.79	6.43	12.17	7.79	15.00
Formulation	DPRN [3]	17.60	19.55	16.07	17.62	19.45	18.06	85.44
	DerainRLNet [17]	18.33	21.81	15.57	19.44	21.81	19.39	58.22
	Syn2Real [54]	14.50	22.80	13.97	15.54	22.07	17.78	26.18
Unsupervised	CVF-SID [21]	15.85	21.22	24.16	25.37	11.96	19.71	11.86
	D4 [55]	7.85	12.27	9.61	9.66	15.86	11.05	284.29
	YOLY [7]	13.66	19.17	15.00	16.20	19.38	16.68	236.58
Diffusion	WeatherDiffusion [56]	13.01	11.92	11.96	12.21	24.04	14.63	109.74
Reprogramming	Res12	22.04	22.47	19.41	22.51	17.75	20.84	7.39
	GridDehazeNet	17.32	23.01	16.93	18.24	31.05	21.31	9.58
	SwinIR	16.52	22.74	17.35	18.50	29.54	20.93	19.05
	MPRNet	17.80	20.18	16.20	16.59	20.10	18.17	157.41
	MIRNet	18.63	23.56	18.08	19.25	27.39	21.38	317.88

domain adaptation methods are ineffective to learn transferable restoration models. As a result, these compared methods cannot recover haze degradation and out-of-sample blur artifacts.

2) *Image Restoration Methods*: The proposed framework is further compared with image restoration methods from two aspects. On the one hand, Table X presents the comparison with Transformer-based [1], [2] and Diffusion-based [56] methods, all trained on haze and LR degradations but tested

on five restoration circumstances. By reprogramming Res12, although our method may not surpass over-parameterized methods like SwinIR [1] and GridDehazeNet [19], it significantly improves the generalization ability of the Res12. In addition, when the reprogrammed model is replaced by strong baselines such as GridDehazeNet, our framework achieves better performance. It can be observed that our methods can effectively enhance the generalization ability of MPRNet [52] and MMIRNet [53], all without significantly causing additional parameters. Specifically, upon reprogramming, the PSNR metrics of MPRNet witness an increase of approximately 7.8% across LR, rain, noise, and blur images. Similarly, leveraging our reprogramming framework results in 9.0% improvement for the strong baseline MIRNet across the same set of images. These results underscore the efficacy of our methods in augmenting the performance of various restoration models. Furthermore, we validate our framework by generalizing from synthetic images to real-world scenarios. As training sets of the haze degradation contains only synthetic indoor images, the outdoor hazy images in SOTS are taken for evaluation. As shown in Table XI and Fig. 10, our methods outperform all dehazing methods in term of PSNR and realness.

In addition to performance improvements, Table X also highlights two important observations. First, among the 14 compared methods, only Restormer, DPRN, DerainRLNet, Syn2Real, and CVF-SID exhibit inferior performance on the haze degradation. DerainRLNet and Syn2Real are specialized for image deraining, while CVF-SID is designed for image denoising, thereby resulting in poorer performance in haze removal compared to their respective specialized tasks. In contrast, DPRN's design, based on physical models of rainy and

TABLE XI
GENERALIZATION COMPARISONS OF RECOVERING OUTDOOR
HAZY IMAGES. ALL METHODS ARE TRAINED ON LR AND
INDOOR HAZY IMAGES

Taxonomy	Method	SOTS-outdoor	
		PSNR	SSIM
Supervised	AOD-Net [8]	22.71	0.91
	DehazeNet [57]	22.57	0.86
	GFN [58]	21.25	0.84
	EPDN [16]	21.55	0.91
Unsupervised	ReineDNet [5]	20.61	0.88
	Cycle-Dehaze [59]	19.95	0.88
	ZID [6]	13.52	0.63
	Ours	23.18	0.90

hazy images, yields similar performance levels for rain and haze degradations. The spatially enriched depth-wise convolution layers make Restormer more sensitive to depth-wise and spatially-variant degradations, *i.e.*, rain and blur. These results indicate that model architectures are related to the generalization ability. Second, upon reprogramming, the Res12, MPRNet, and MIRNet tend to suffer from sacrificed performance in dehazing. Since the parameters of these baselines remain unchanged, applying the baselines directly can yield better performance on haze and LR degradations. Such results underscore the presence of a trade-off between generalization and existing performance. Consequently, it is imperative to investigate how the parameter space of restoration models influences performance on both within-training-set and out-of-sample degradations.

V. CONCLUSION

This work addresses the challenging task of out-of-sample restoration, aiming to develop restoration models capable of generalizing to degradations not encountered during training. To tackle this task, we propose a model reprogramming framework inspired by image-to-image translation and quantum mechanics. In this framework, images are transformed into wave functions represented by amplitude and phase terms, which capture degradation-invariant image content and degradation-specific style distributions, respectively. The process of recovering out-of-sample restoration involves enhancing the amplitude term and aligning the phase component. Instead of training restoration networks from scratch, our framework leverages existing restoration models and adapts them using two transform functions. Extensive experimental results under various circumstances demonstrate effectiveness and flexibility of the proposed framework.

REFERENCES

- [1] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van G., and R. Timofte, “SwinIR: Image restoration using Swin transformer,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops*, Oct. 2021, pp. 1833–1844.
- [2] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient transformer for high-resolution image restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5728–5739.
- [3] X. Zhang, R. Jiang, T. Wang, and W. Luo, “Single image dehazing via dual-path recurrent network,” *IEEE Trans. Image Process.*, vol. 30, pp. 5211–5222, 2021.
- [4] L. Liu et al., “TAPE: Task-agnostic prior embedding for image restoration,” 2022, *arXiv:2203.06074*.
- [5] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, “RefineDNet: A weakly supervised refinement framework for single image dehazing,” *IEEE Trans. Image Process.*, vol. 30, pp. 3391–3404, 2021.
- [6] B. Li, Y. Gou, J. Z. Liu, H. Zhu, and J. T. Zhou, “Zero-shot image dehazing,” *IEEE Trans. Image Process.*, vol. 29, pp. 8457–8466, 2020.
- [7] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, “You only look yourself: Unsupervised and untrained single image dehazing neural network,” *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1754–1767, 2021.
- [8] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “AOD-Net: All-in-one dehazing network,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 4770–4778.
- [9] M. Long, Y. Cao, J. Wang, and M. Jordan, “Learning transferable features with deep adaptation networks,” in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 97–105.
- [10] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, “Maximum classifier discrepancy for unsupervised domain adaptation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3723–3732.
- [11] A. Lengyel, S. Garg, M. Milford, and J. C. van Gemert, “Zero-shot day-night domain adaptation with a physics prior,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4399–4409.
- [12] S. Li et al., “Semantic concentration for domain adaptation,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9102–9111.
- [13] Z. Han, H. Sun, and Y. Yin, “Learning transferable parameters for unsupervised domain adaptation,” *IEEE Trans. Image Process.*, vol. 31, pp. 6424–6439, 2022.
- [14] Y. Tang et al., “An image patch is a wave: Phase-aware vision MLP,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10935–10944.
- [15] J.-J. Huang and P. L. Dragotti, “WINNet: Wavelet-inspired invertible network for image denoising,” *IEEE Trans. Image Process.*, vol. 31, pp. 4377–4392, 2022.
- [16] Y. Qu, Y. Chen, J. Huang, and Y. Xie, “Enhanced pix2pix dehazing network,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.
- [17] C. Chen and H. Li, “Robust representation learning with feedback for single image deraining,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7742–7751.
- [18] B. Li et al., “Benchmarking single-image dehazing and beyond,” *IEEE Trans. Image Process.*, vol. 28, pp. 492–505, 2018.
- [19] X. Liu, Y. Ma, Z. Shi, and J. Chen, “GridDehazeNet: Attention-based multi-scale network for image dehazing,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7314–7323.
- [20] S. Zhuo, Z. Jin, W. Zou, and X. Li, “RIDNet: Recursive information distillation network for color image denoising,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3896–3903.
- [21] R. Neshatavar, M. Yavartanoo, S. Son, and K. M. Lee, “CVF-SID: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17562–17570.
- [22] Y. Wang, J. Yu, and J. Zhang, “Zero-shot image restoration using denoising diffusion null-space model,” 2022, *arXiv:2212.00490*.
- [23] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [24] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
- [25] X. Chen, S. Wang, M. Long, and J. Wang, “Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation,” in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1081–1090.
- [26] P.-Y. Chen, “Model reprogramming: Resource-efficient cross-domain machine learning,” 2022, *arXiv:2202.10629*.
- [27] G. F. Elsayed, I. Goodfellow, and J. Sohl-Dickstein, “Adversarial reprogramming of neural networks,” 2018, *arXiv:1806.11146*.
- [28] Y. Tsai, P.-Y. Chen, and T.-Y. Ho, “Transfer learning without knowing: Reprogramming black-box machine learning models with scarce data and limited resources,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 9614–9624.
- [29] I. Melnyk et al., “Reprogramming pretrained language models for antibody sequence infilling,” 2022, *arXiv:2210.07144*.
- [30] P. Neekhara, S. Hussain, J. Du, S. Dubnov, F. Koushanfar, and J. McAuley, “Cross-modal adversarial reprogramming,” in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 2898–2906.

- [31] C.-H. H. Yang, Y.-Y. Tsai, and P.-Y. Chen, "Voice2Series: Reprogramming acoustic models for time series classification," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11808–11819.
- [32] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 97–104.
- [33] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.
- [34] Z. Shen et al., "Human-aware motion deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5572–5581.
- [35] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1279–1288.
- [36] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704.
- [37] R. Timofte et al., "NTIRE 2018 challenge on single image super-resolution: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 852–863.
- [38] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2017, pp. 1–10.
- [39] M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, and A. Zeilinger, "Wave-particle duality of C₆₀ molecules," *Nature*, vol. 401, no. 6754, pp. 680–682, Oct. 1999.
- [40] E. J. Heller, M. F. Crommie, C. P. Lutz, and D. M. Eigler, "Scattering and absorption of surface electron waves in quantum corrals," *Nature*, vol. 369, no. 6480, pp. 464–466, Jun. 1994.
- [41] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2010.
- [42] S. Gu, Y. Li, L. Van Gool, and R. Timofte, "Self-guided network for fast image denoising," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2511–2520.
- [43] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.
- [44] Huawei. (2020). *Mindspore*. [Online]. Available: <https://www.mindspore.cn/>
- [45] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *Proc. 16th Eur. Conf. Comput. Vision (ECCV)*. Cham, Switzerland: Springer, Aug. 2020, pp. 184–201.
- [46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [47] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [48] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [49] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 3941–3950.
- [50] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 443–450.
- [51] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3155–3164.
- [52] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14821–14831.
- [53] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *Proc. Eur. Conf. Comput. Vis.*, vol. 12370. Cham, Switzerland: Springer, 2020, pp. 492–511.
- [54] R. Yasarla, V. A. Sindagi, and V. M. Patel, "Syn2Real transfer learning for image deraining using Gaussian processes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2726–2736.
- [55] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2037–2046.
- [56] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10346–10357, Aug. 2023.
- [57] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
- [58] W. Ren et al., "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.
- [59] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-Dehaze: Enhanced Cycle-GAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 825–833.



Runhua Jiang received the M.S. degree from the College of Computer Science and Artificial Intelligence, Wenzhou University, China. He is currently pursuing the Ph.D. degree majoring in computer science and technology with the College of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include image/video restoration, domain adaptation, and large vision models.



Yahong Han (Member, IEEE) received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2012. He is currently a Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. From November 2014 to November 2015, he visited Prof. Bin Yu's Group, UC Berkeley, as a Visiting Scholar. His current research interests include multimedia analysis, computer vision, and machine learning.