Learning from History: Task-agnostic Model Contrastive Learning for Image Restoration

Gang Wu, Junjun Jiang*, Kui Jiang, Xianming Liu

Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China
{gwu, jiangjunjun, jiangkui, csxm}@hit.edu.cn

Abstract

Contrastive learning has emerged as a prevailing paradigm for high-level vision tasks, which, by introducing properly negative samples, has also been exploited for low-level vision tasks to achieve a compact optimization space to account for their ill-posed nature. However, existing methods rely on manually predefined and task-oriented negatives, which often exhibit pronounced task-specific biases. To address this challenge, our paper introduces an innovative method termed 'learning from history', which dynamically generates negative samples from the target model itself. Our approach, named Model Contrastive Learning for Image Restoration (MCLIR), rejuvenates latency models as negative models, making it compatible with diverse image restoration tasks. We propose the Self-Prior guided Negative loss (SPN) to enable it. This approach significantly enhances existing models when retrained with the proposed model contrastive paradigm. The results show significant improvements in image restoration across various tasks and architectures. For example, models retrained with SPN outperform the original FFANet and DehazeFormer by 3.41 and 0.57 dB on the RESIDE indoor dataset for image dehazing. Similarly, they achieve notable improvements of 0.47 dB on SPA-Data over IDT for image deraining and 0.12 dB on Manga109 for a 4× scale super-resolution over lightweight SwinIR, respectively. Code and retrained models are available at https://github.com/Aitical/MCLIR.

Introduction

Image restoration, aiming at recovering a high-quality image from the degraded one, is a fundamental problem in the fields of image processing and computer vision (Yang et al. 2021; Wang, Chen, and Hoi 2021a; Jiang et al. 2023). Deep learning approaches have made considerable advancements in image restoration, while there are still challenges due to its ill-posed nature (Wang, Chen, and Hoi 2021b; Lu et al. 2023). The success of the self-supervised learning paradigm for high-level tasks, especially those using contrastive learning methods, has drawn great attention (Gui et al. 2023a). This inspires many researchers to make strides in improving the end-to-end learning paradigm for image restoration tasks, incorporating the concept of sample contrastive learning and bridging the gap between high-level and low-level tasks (Wu et al. 2021; Chen et al. 2022b; Ye et al. 2022; Wu, Jiang, and Liu 2023; Zheng et al. 2023). Image restoration tasks usually contain high-quality ground truth as the learning target (positive sample), and more attention is paid to obtaining appropriate negative examples. For example, Wu *et al.* (Wu et al. 2021) directly utilized low-quality input as the negative sample and introduced the contrastive paradigm for the image dehazing task. Wu *et al.* proposed a hard negative construction for image super-resolution tasks. Compared to the super-resolved results, hard negatives with similar image quality to the anchor sample can push it ahead effectively (Wu, Jiang, and Liu 2023). Most recently, Zheng *et al.* utilized multiple pre-trained models to provide consensual negatives and proposed a progressively improved negative lower bound for image dehazing (Zheng et al. 2023).

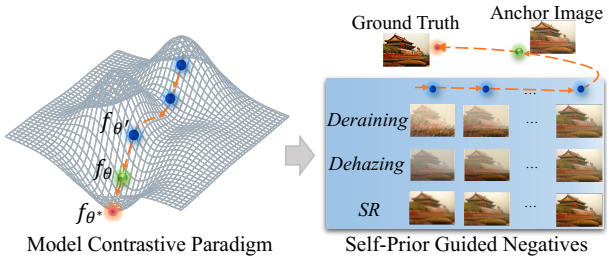


Figure 1: Illustration of the proposed model contrastive paradigm. We provide a common optimization space for it. To the target model f_θ, the proposed model contrastive paradigm exploits negative samples from the latency model $f_{\theta'}$ smoothly. Compared to task-oriented negatives in previous work, our model contrastive paradigm is task-agonist and general to various image restoration tasks. This provides a compact optimization space adaptively (pushing target model f_θ closer to assumed optimal f_{θ^*}).

While existing contrastive learning methods for low-level tasks, leveraging potent approaches such as hard negative mining (Wu, Jiang, and Liu 2023) and curriculum learning strategies (Zheng et al. 2023), have shown impressive performance, certain intrinsic limitations persist. Chiefly, there exists an *over-reliance on task-oriented prior*, leading to lim-

*Corresponding Author

ited generalization capability across multiple image restoration tasks. Many existing methods exhibit a pronounced task-oriented bias, where negative sample generation is often influenced by prior knowledge and empirical evaluation centered on a singular target task (Wu et al. 2021; Wu, Jiang, and Liu 2023; Zheng et al. 2023). Such approaches, while effective in isolation, often impede to a diverse set of image restoration tasks and model architectures. In light of these challenges, it compels us to ask: *Is there a task-agnostic and general method for negative sampling that could potentially enhance the performance of a diverse range of image restoration tasks?*

Taking this into consideration, we turn our focus to the target model itself, rather than sample selection. In recent literature, much of the spotlight has been directed towards negative sample collection, often overlooking a latent gem, the latency model, during the learning process. Specifically, the latency model, when operating within a close optimization step, shares strikingly similar parameters with the current model. This intrinsic similarity paves the way for the construction of proper negatives pertinent to the current anchor sample. To illustrate this concept, we introduce a toy example in Fig. 1. It becomes evident that, throughout the learning journey, the output of the latency model exhibits a suboptimal but congruent distribution relative to the current anchor sample. This alignment offers the potential to derive 'hard' negatives that are well suited to the task at hand. Furthermore, as the entire learning process is incrementally refined, the negatives in our model contrastive paradigm adopt a curriculum way naturally. Motivated by these insights, we put forth an innovative ***model contrastive paradigm*** for image restoration tasks. In contrast to previous approaches with complex negative mining strategies, the core of our proposed model contrastive paradigm lies in the self-promoted negatives between latency and current models. Notably, it is a task-agonist and general to diverse image restoration tasks.

In detail, we propose a model contrastive paradigm for various image restoration tasks (MCLIR) with a Self-Prior guided Negative loss (SPN). Importantly, the proposed MCLIR is compatible with existing approaches across different architectures and tasks. We retrain existing models by the proposed model contrastive and some results are presented in Fig. 2. Retrained models can be found to achieve promising improvement across multiple tasks and architectures. More specifically, based on lightweight EDSR (Lim et al. 2017) and SwinIR (Liang et al. 2021), our retrained models show superior performance on the Manga109 test dataset for $\times 4$ scale image super resolution, boasting improvements of **0.16 dB** and **0.12 dB** in terms of PSNR. Retrained IDT (Xiao et al. 2022) achieves **0.38 dB** and **0.7 dB** gains on Rain200L and SPA datasets for image deraining. Furthermore, for image dehazing, we gain a notable improvement of **3.41 dB** and **0.57 dB** compared to original FFANet (Qin et al. 2020a) and DehazeFormer (Song et al. 2023), respectively.

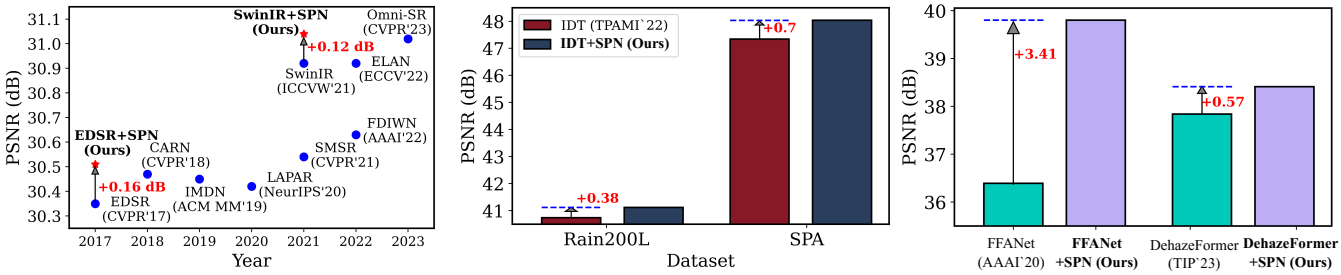
The main contributions of this work are:

- The paper proposes a novel approach MCLIR through a task-agnostic model contrastive paradigm, which provides the adaptive generation of negative samples directly from the target model itself. Unlike conventional methods that manually apply negative samples to a specific target task, the proposed model contrastive paradigm exhibits versatility across multiple tasks and models.
- The paper introduces the self-prior guided negative loss (SPN) for image restoration, which is seamlessly compatible with existing methods. SPN provides a simple to enhance existing image restoration models by integrating self-supervision principles within our model contrastive paradigm.
- The paper demonstrates the effectiveness of the proposed approach by retraining existing models with the MCLIR, which significantly improves image restoration across various tasks and architectures.

Related Work

Image Restoration Image restoration, crucial in image processing and computer vision, aims to restore high-quality images from degraded versions (Su, Xu, and Yin 2022). Deep learning, particularly CNNs, has revolutionized this domain (Wang, Chen, and Hoi 2021a; Jiang et al. 2023; Wang et al. 2021a; Gui et al. 2023b), with advancements in various CNN-based methods (Dong et al. 2016; Zhang et al. 2017; Zhang, Zuo, and Zhang 2018; Yang et al. 2017a), and the incorporation of novel architectures such as residual networks (Lim et al. 2017), attention mechanisms (Zhang et al. 2018; Liu et al. 2019), and UNet structures (Ronneberger, Fischer, and Brox 2015; Jiang et al. 2020). As well as some lightweight architectures were proposed for practical application (Jiang et al. 2021; Lamba and Mitra 2021; Hui et al. 2019; Wu et al. 2023b; Lu et al. 2022; Wu et al. 2023a). Recent trends include Transformer-based architectures (Liu et al. 2021; Dosovitskiy et al. 2021), delivering breakthroughs in many image restoration tasks (Liang et al. 2021; Song et al. 2023; Zamir et al. 2022; Xiao et al. 2022; Wang et al. 2022, 2023). In this work, we turn our attention to the optimizing strategy and exploit an effective model contrastive learning method to refresh the performance existing models.

Contrastive Learning Self-supervised learning, particularly contrastive methods, has seen remarkable success in high-level vision tasks (Gui et al. 2023a). In the field of low-level image restoration, researchers are exploring the integration of self-supervised regularization to bridge the gap between low- and high-level tasks (Wang et al. 2021b; Wu et al. 2021; Chen et al. 2022b; Ye et al. 2022; Zheng et al. 2023; Wu, Jiang, and Liu 2023). Typically, low-level tasks often have high-quality ground truth, making the identification of informative negative samples vital. Wu *et al.* employed low-quality inputs as negative samples and introduced a perceptual-based contrastive loss for more effective convergence (Wu et al. 2021). Wu *et al.* proposed a practical contrastive learning framework for single image super-resolution tasks, enhancing performance through hard negative construction and negative information interpolation (Wu, Jiang, and Liu 2023). Nonetheless, the adaptability



(a) Lightweight Image Super-Resolution (b) Image Deraining (c) Image Dehazing

Figure 2: Comparisons between models retrained by our proposed model contrastive paradigm and the originals. Retrained models can achieve remarkable improvements on various image restoration tasks.

and generalization of these methods across diverse image restoration models and tasks remain a challenge.

Our work distinguishes itself from existing methods (Wu et al. 2021; Wu, Jiang, and Liu 2023; Zheng et al. 2023) by introducing a task-agnostic model contrastive paradigm for image restoration. This novel approach shifts the focus from negative samples to negative models, offering a straightforward yet effective framework for a variety of image restoration tasks.

Method

Overall Framework

Deep learning-based image restoration methodologies have recently made considerable breakthroughs across multiple restoration tasks, including image super-resolution, image dehazing, image deblurring, image deraining, and so on. In this paper, we focus our attention on the vanilla end-to-end framework with deep learning methods.

Given a low-quality input image I^{LQ}, the target model f_θ (where θ denotes the model parameters) processes this input to produce a reconstructed image I^{Rec}. The high-quality counterpart of the image is represented as I^{HQ}. The optimization of f_θ is guided by minimizing the reconstruction loss $\mathcal{L}_{rec}$, defined as:

$$\min_{\theta} \mathcal{L}_{rec}(f_\theta(I^{LQ}), I^{HQ}), \tag{1}$$

where $\mathcal{L}_{rec}$ typically depends on metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Delving further, existing contrastive learning methods for image restoration (Wu et al. 2021; Zheng et al. 2023; Wu, Jiang, and Liu 2023) utilize negative samples I^{Neg}, extending the standard end-to-end learning with a negative loss $\mathcal{L}_{neg}$:

$$\min_{\theta} \mathcal{L}_{rec}(f_\theta(I^{LQ}), I^{HQ}) + \mathcal{L}_{neg}(f_\theta(I^{LQ}), I^{Neg}), \tag{2}$$

where the negative samples are predefined and generated based on task-specific priors, which limit their generalization capability across diverse image restoration tasks.

In this work, we introduce a novel model contrastive paradigm for image restoration, shifting from task-specific negative samples to negatives derived from the target model itself. We designate the target model as f_θ, representing its state in a given training iteration t. A key innovation is the development of a latency model, $f_{\theta'}$ (the negative model), which is utilized to produce adaptive negative samples for training f_θ. This approach simplifies the overall process in existing methods and provides a general framework for image restoration. The specifics of implementing negative models and the corresponding loss functions will be elaborated in subsequent sections.

Learning from History

We propose a practical implementation for leveraging latency models to generate informative negative samples during training, as introduced earlier. This method avoids the challenges of significant gaps between model checkpoints and the impracticality of frequent processing, especially with larger models. In detail, we introduce the efficient strategy of exponential moving averages (EMA) to achieve a smooth negative model. The update equation for the negative model is as follows:

$$\theta^{'} = w\theta^{'} + (1-w)\theta, \text{ s.t. } t\%s = 0. \tag{3}$$

In this schema, w denotes the update weight, s is the update step, and t captures the current iteration. To ensure the preservation of latency parameters, we adopt a long step s and selectively update the negative model f_θ at intervals of every s iteration.

Self-Prior Guided Negative Loss

At the heart of our model contrastive paradigm, a loss function mediates between the target reconstruction I^{Rec} and its negative counterpart, I^{Neg}. Here, we take the pre-trained VGG (Simonyan and Zisserman 2015) network as the embedding network to map the samples into a latent feature space, where $f^{Rec} = \text{VGG}(I^{Rec})$ and $f^{Neg} = \text{VGG}(I^{Neg})$. Then the proposed negative loss $\mathcal{L}_{neg}$ is formulated as

$$\mathcal{L}_{neg} = \|f^{Rec} - f^{Neg}\|_1. \tag{4}$$

Furthermore, our model-based contrastive paradigm can incorporate multiple negatives by adding more negative models. For more robustness, we take multiple distinct steps

to obtain several latency models. The combined negative loss, accounting for multiple negatives, is represented as

$$\mathcal{L}_{neg}^{N} = \frac{1}{N}\sum_{i=1}^{N} \|f^{Rec} - f_i^{Neg}\|_1, \tag{5}$$

where N denotes the total number of negative models and f_i^{Neg} corresponds to the latent feature of the i-th negative sample.

Compared to existing methods, our negative regularization is a self-prior guided loss function, where the negative samples stem from the target model itself. More important, it is general and transferable to existing image restoration models while retaining the original learning strategy. Typically, image restoration tasks utilize a reconstruction loss, $\mathcal{L}rec$, relying on metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). To validate the prowess of our model-based contrastive paradigm by $\mathcal{L}_{neg}$, we have retrained numerous exiting methods with it, testing across different image restoration tasks and architectures. The precise formulation of the reconstruction loss $\mathcal{L}_{rec}$ is dependent on the retraining method.

Generally, the total loss function within our model-based contrastive paradigm is defined as:

$$\mathcal{L} = \mathcal{L}_{rec} - \lambda\mathcal{L}_{neg}^{N}, \tag{6}$$

where $\mathcal{L}_{rec}$ represents the corresponding reconstruction loss adopted in existing method, and λ is the balancing coefficient. This simple formulation allows us to incorporate the proposed SPN with existing image restoration methods, enhancing their flexibility and adaptability to various tasks.

Remark

In this study, we propose the model contrastive paradigm for image restoration tasks. This approach significantly simplifies the construction of negatives while simultaneously providing adaptive and effective ones. Contrary to existing approaches (Wu et al. 2021; Wu, Jiang, and Liu 2023; Zheng et al. 2023), the proposed model contrastive paradigm does not rely on the restrictions of task-oriented priors, making it versatile and universal to various image restoration tasks. In essence, our contribution lies in providing a task-agonist and general approach to the construction of negative samples, expanding the scope of contrastive learning in the field of image restoration.

Experiment

Experimental Settings

In this section, we apply the following four image restoration tasks to test and validate the effectiveness of the proposed method.

Image Super-Resolution We retrain the CNN-based EDSR (Lim et al. 2017) and Transformer-based SwinIR (Liang et al. 2021). We take 800 images from DIV2K (Agustsson and Timofte 2017) for training and test them on five benchmark datasets.

Image Dehazing We employ the CNN-based FFANet (Qin et al. 2020a) and Transformer-based DehazeFormer (Song et al. 2023) as our baselines and retrain them for image dehazing tasks. Following (Song et al. 2023), we utilize the indoor training dataset (ITS) and RESIDE-6K to separately train the models and evaluate on the synthetic objective testing set (SOTS).

Image Deblurring For image deblurring, we retrain and evaluate NAFNet (Chen et al. 2022a) on GoPro dataset (Nah, Hyun Kim, and Mu Lee 2017).

Image Deraining For image deraining, we conduct experiments on several publicly available benchmarks, namely Rain200L/H (Yang et al. 2017b), DID-Data (Li et al. 2018), DDN-Data (Fu et al. 2017), and SPA-Data (Wang et al. 2019). We adopt the Transformer-based IDT model (Xiao et al. 2022) as our baseline and subsequently retrain it using the proposed model contrastive paradigm.

Given the diverse training settings across various image restoration tasks, we ensure a fair comparison by strictly adhering to the original training configurations of each retrained model, as specified in their literature.

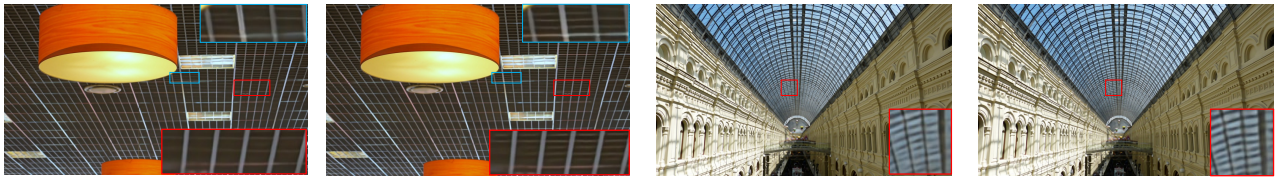
Comparison Results

Image Super-Resolution We first investigate the image super-resolution task, utilizing both CNN-based EDSR and Transformer-based SwinIR as baselines to evaluate our proposed self-prior guided negative regularization. Results presented in Tab. 1 across various architectures and model capacities. Our proposed model contrastive paradigm can achieve considerable improvements when retraining lightweight models such as EDSR and SwinIR, achieving average gains of 0.07 dB and 0.09 dB on scale $\times 4$, respectively. Moreover, the retrained EDSR model demonstrates a noteworthy average improvement of 1.3 dB for the scale $\times 2$ task. In particular, there are substantial improvements from our re-training on the Urban100 and Manga109 datasets. This implies that our model contrastive paradigm facilitates a more accurate and compact convergence in the restoration process. Moreover, comparisons with current state-of-the-art methods are presented in Fig. 2 (a). Remarkably, the retrained EDSR and SwinIR can eclipse the performance of several contemporary methods. These underscore the promise of our model contrastive paradigm.

In addition to the lightweight models, we also retrain the large SwinIR model. Our proposed approach exhibits its general applicability to larger models, providing an average improvement of 0.05 dB. It is worth noting that the improvements seen in the larger models were relatively smaller compared to the lightweight models. This observation is consistent with the understanding that lightweight models, owing to their smaller capacity, may struggle to find an optimal solution. Therefore, our proposed model contrastive paradigm provides significant assistance in these cases. Conversely, larger models have the capacity to fit or even overfit the dataset, hence the model contrastive paradigm improves them within a lesser margin. Visual results are showcased in Fig. 3. It is also evident that our model contrastive paradigm

Methods	Architecture	Scale	Avg. PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
EDSR-light	CNN	×2	34.06/0.9303	33.57/0.9175	32.16/0.8994	31.98/0.9272	38.54/0.9769
+SPN (Ours)			**34.19/0.9313**	**33.67/0.9182**	**32.21/0.9001**	**32.23/0.9297**	**38.64/0.9772**
EDSR-light		×4	28.14/0.8021	28.58/0.7813	27.57/0.7357	26.04/0.7849	30.35/0.9067
+SPN (Ours)			**28.21/0.8040**	**28.63/0.7829**	**27.59/0.7369**	**26.12/0.7878**	**30.51/0.9085**
SwinIR-light	Transformer	×4	28.46/0.8099	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151
+SPN (Ours)			**28.55/0.8114**	**28.85/0.7874**	**27.72/0.7414**	**26.57/0.8010**	**31.04/0.9158**
SwinIR		×4	28.88/0.8190	28.94/0.7914	27.83/0.7459	27.07/0.8164	31.67/0.9226
+SPN (Ours)			**28.93/0.8198**	**29.01/0.7923**	**27.85/0.7465**	**27.14/0.8176**	**31.75/0.9229**

Table 1: Comparison results of image super-resolution. We take CNN-based EDSR (Lim et al. 2017) and Transformer-based SwinIR (Liang et al. 2021) as our baselines. Results of retrained models by the proposed model contrastive paradigm are in bold. Avg. presents the mean value of four test datasets.



SwinIR-light | **SwinIR-light +SPN (ours)** | SwinIR | **SwinIR +SPN (ours)**

Figure 3: Visual comparison for image super-resolution tasks. Displayed are results from both the original SwinIR models and those retrained by our model contrastive paradigm. The enhancements brought about by our approach are clearly evident.

Input PSNR/SSIM | IDT 37.35/0.9891 | **IDT+SPN (Ours) 37.72/0.9897** | Divergence

Figure 4: Visual comparisons between IDT and our retrained one. A divergence map delineates the differences between the IDT output and ours, highlighting the improvement, particularly in degraded regions.

can improve existing methods, particularly in bringing more precise textures.

Image Deraining We take the advanced IDT (Xiao et al. 2022) as a baseline and retrain it by the proposed model contrastive paradigm on five datasets, and Tab. 2 provides a comprehensive comparison. The retrained IDT achieves an average PSNR improvement of 0.26 dB across all test datasets. For instance, it showcases a gain of 0.7 dB in comparison to its original in the SPA dataset. Furthermore, visual results are illustrated in Figure 4, accompanied by a divergence map. It intuitively highlights enhancements in the degraded regions. These insights underscore the ability of our model contrastive paradigm to refine target degradation.

Image Dehazing Following the recent work (Song et al. 2023), we utilize the SOTS-indoor and SOTS-mix datasets for testing. In detail, we retrain FFANet (Qin et al. 2020b) and DehazeFormer (Song et al. 2023), with the results detailed in Fig. 3. From the table, a remarkable enhancement is achieved over FFANet, exhibiting gains of 3.41 dB and 0.69 dB on the indoor and mixed test datasets, respectively. Concurrently, the Transformer-based model DehazeFormer also records advancements, evident across various model scales. Specifically, the retrained DehazeFormer-B achieves the most pronounced improvement of 0.57 dB on the indoor test dataset. Intriguingly, the retrained FFANet even outperforms the DehazeFormer. It is reasonable that FFANet is a CNN-based model, which has a large model capacity with heavy parameter counts, and our model contrastive paradigm appears to steer FFANet towards more optimized results. Some visual results are presented in Fig. 5. One can find that the retrained FFANet is clearer with fewer artifacts.

Image Deblurring Results of the retrained NAFNet (Chen et al. 2022a) for image deblurring are in Tab. 4. One can find that the retrained NAFNet showcases notable en-

Methods	Avg. PSNR/SSIM	Rain200L PSNR/SSIM	Rain200H PSNR/SSIM	DID PSNR/SSIM	DDN PSNR/SSIM	SPA PSNR/SSIM
(CVPR21) MPRNet	36.17/0.9543	39.47/0.9825	30.67/0.9110	33.99/0.9590	33.10/0.9347	43.64/0.9844
(AAAI'21) DualGCN	36.69/0.9604	40.73/0.9886	31.15/0.9125	34.37/0.9620	33.01/0.9489	44.18/0.9902
(ICCV'21) SPDNet	36.54/0.9594	40.50/0.9875	31.28/0.9207	34.57/0.9560	33.15/0.9457	43.20/0.9871
(CVPR'22) Uformer-S	36.95/0.9505	40.20/0.9860	30.80/0.9105	34.46/0.9333	33.14/0.9312	46.13/0.9913
(CVPR'22) Restormer	37.49/0.9530	40.58/0.9872	31.39/0.9164	35.20/0.9363	34.04/0.9340	46.25/0.9911
(CVPR'23) DRSformer	38.33/0.9676	41.23/0.9894	32.17/0.9326	35.35/0.9646	34.35/0.9588	48.54/0.9924
(TPAMI'22) IDT	37.77/0.9593	40.74/0.9884	32.10/0.9343	34.85/0.9401	33.80/0.9407	47.34/0.9929
(Ours) IDT+SPN	**38.03/0.9610**	**41.12/0.9893**	**32.17/0.9352**	**34.94/0.9424**	**33.90/0.9442**	**48.04/0.9938**

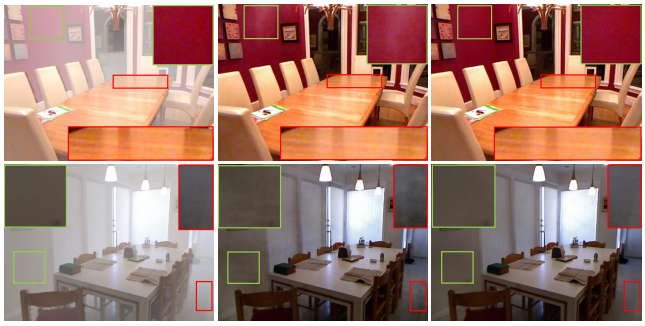
Table 2: Comparison results of image deraining. We take IDT (Xiao et al. 2022) as the benchmark and retrain it with the proposed model contrastive paradigm. We evaluate the performance on several image deraining datasets, and our results are in bold (Average performance of the five datasets is calculated in the Avg. column).

Methods	ITS SOTS-indoor		RESIDE-6K SOTS-mix	
	PSNR	SSIM	PSNR	SSIM
(CVPR'21) AECR-Net	37.17	0.990	-	-
(ICLR'23) SFNet	41.24	0.996	-	-
(CVPR'23) C^2PNet	42.56	0.995	-	-
(AAAI'20) FFANet	36.39	0.989	29.96	0.973
(Ours) FFANet+SPN	**39.80**	**0.995**	**30.65**	**0.976**
(TIP'23) DehazeFormer-T	35.15	0.989	30.36	0.973
(Ours) DehazeFormer-T+SPN	**35.51**	**0.990**	**30.44**	**0.974**
(TIP'23) DehazeFormer-S	36.82	0.992	30.62	0.976
(Ours) DehazeFormer-S+SPN	**37.24**	**0.993**	**30.77**	**0.978**
(TIP'23) DehazeFormer-B	37.84	0.994	31.45	0.980
(Ours) DehazeFormer-B+SPN	**38.41**	**0.994**	**31.57**	**0.981**

Table 3: Comparison results of image dehazing. The results of our retrained models are in bold

.

hancements compared to the original and outperforms the Transformer-based Restormer (Zamir et al. 2022).



Input and GT patches FFANet **FFANet+SPN (Ours)**

Figure 5: Visual results of FFANet and our retrained one for image dehazing.

Comparison to Existing Sample Contrastive Paradigm In contrast to existing methods that employ task-oriented negatives, our model contrastive paradigm provides straightforward enhancement to image restoration tasks. We furnish a comprehensive comparison between exiting contrastive approaches and our innovative model contrastive paradigm. For the image dehazing task, we examine previous methods based on FFANet, including contrastive regularization (CR) (Wu et al. 2021), curricular contrastive regularization (CCR) (Zheng et al. 2023), and our self-prior guided negative loss (SPN). The comparative results can be found in Tab. 5. Instead of merely employing low-quality input images as negative samples, CCR achieves significant advancements by utilizing consensus negatives from pre-trained models and the curriculum learning approach. Intriguingly, our SPN outperforms other methods, achieving the best improvement without additional prior knowledge. In addition, for image super-resolution (Wu, Jiang, and Liu 2023), our model contrastive paradigm again registers the highest gains, as presented in Tab. 5. This consistent performance underscores the fact that our proposed model contrastive paradigm is task-agnostic and highly effective for diverse image restoration tasks.

Ablation Studies

Impact of Negative Model This ablation study aims to elucidate the role of our negative model by investigating various configurations, including a randomly initialized and fixed SwinIR model, a pre-trained and fixed SwinIR model, and our standard model. The results, detailed in Tab. 6, highlight the effectiveness of our approach. The baseline model is denoted by '-,' indicating the origin SwinIR. We found that a negative model with random and fixed parameters hinders the target SR model, failing to provide effective negatives. Conversely, using a pre-trained and fixed negative model shows improvement, but not as much as our model contrastive. This suggests the importance of adaptive and curriculum-based negative generation in our model contrastive paradigm, aligned with the findings of existing methods (Zheng et al. 2023).

Impact of Negative Step We conducted an ablation study, in which FFANet was retrained with various negative models, each differentiated by its update iteration, s. The results are tabulated in Tab. 7. It becomes evident that the choice of step s distinctly influences both the negative model and per-

Methods	MIMO-UNet	HINet	MAXIM	Restormer	UFormer	NAFNet	**NAFNet+SPN (Ours)**
PSNR	32.68	32.71	32.86	32.92	32.97	32.87	**32.93**
SSIM	0.959	0.959	0.961	0.961	0.967	0.9606	**0.9619**

Table 4: Comparison results of image deblurring. We take NAFNet (Chen et al. 2022a) as the benchmark and retrain it with the proposed model contrastive paradigm on GoPro dataset.

Methods	Task & Dataset	PSNR	SSIM
FFANet	Image Dehazing (SOTS-indoor)	36.39	0.9886
+CR		36.74	0.9906
+CCR		39.24	0.9937
+SPN (Ours)		**39.80**	**0.9947**
EDSR	SISR (Urban100)	26.04	0.7849
+PCL		26.07	0.7863
+SPN (Ours)		**26.12**	**0.7878**

Table 5: A comparative analysis of the existing contrastive paradigms versus our proposed model contrastive approach. Typically, existing methods are task-oriented and are proposed for image dehazing (CR (Wu et al. 2021) and CCR (Zheng et al. 2023)) and image super-resolution (PCL (Wu, Jiang, and Liu 2023)), separately. Our model contrastive paradigm is task-agonist and outperform existing methods.

Negative Model	-	Random	Pre-trained	Default
Avg. PSNR	28.46	28.44	28.51	28.55

Table 6: Results of retrained lightweight SwinIR with different negative models.

Step s	100	500	1000	2000	All
FFANet	38.90	38.54	38.27	37.55	**39.80**

Table 7: Ablation studies of the negative step s. We retrain FFANet with different negative step s on indoor dataset.

formance. A smaller step provides more challenging negatives compared to the larger step, leading to superior performance. In our experiments, we employ multiple negative models with the four steps together, achieving the best results. The incorporation of multiple negative models ensures a consistent supply of robust negatives.

Impact of Balancing Coefficient We study the influence of the coefficient λ in Eq. (3), and results are presented in Tab. 8, where the value 0 means the original loss L_1 without our negative regularization. When λ is set to 1e-2 or 1e-3, there is collapse during training, especially in the early stage. This is reasonable because large negative loss can influence the model updating towards the optimal direction. When λ is 1e-4, our model contrastive paradigm achieves the best performance. Considering that the proposed model contrastive paradigm is general to various image restoration tasks, in this paper we take $\lambda = 1e-4$ as the default value across different tasks.

λ	0	1e-2	1e-3	5e-4	1e-4	1e-5
EDSR-light	28.58	-	-	28.56	28.63	28.60

Table 8: Ablation studies on coefficient in total loss function.

w in EMA	0	0.01	0.1	0.5	0.9	0.999
EDSR-light	-	-	**28.63**	28.60	26.60	28.61

Table 9: Ablation studies on the updating weight in EMA.

Impact of Updating Weight in EMA In reference to Eq. (3), we examine the influence of varying updating weights. The outcomes are detailed in Tab. 9. In general, a smaller value of w retains less latency information, bringing it in closer alignment with the target model. This results in the generation of the most challenging negatives. The symbol '-' indicates model collapse during training. As w increases, it becomes evident that the optimal performance is attained at $w = 0.1$. On the basis of our empirical observations, we consistently use $w = 0.1$ as the default setting in all subsequent experiments.

Discussion and Limitation

In this paper, we proposed a novel model contrastive paradigm for low-level image restoration tasks, improving existing models. Although our experiments cover a range of tasks, areas such as JPEG artifact removal, image denoising, and certain real-world scenarios remain unexplored. Our ablation studies focused on image super-resolution, as shown in Tab. 6 and 7 but we did not extensively evaluate our hyperparameters across all tasks. These unexamined aspects offer valuable opportunities for future research.

Conclusion

In this study, we propose an innovative model contrastive paradigm for various low-level tasks. Compared to the task-oriented negatives in existing methods, the proposed model contrastive paradigm, constructing negatives from the target model itself, is task-agonist and general to various image restoration tasks by a self-prior guided negative loss (SPN). Our proposed SPN is straightforward to implement. We have retrained several image restoration models, and they achieve significant improvements across various tasks and architectures. In the future, we believe it would be meaningful to evaluate our proposed paradigm in more dense prediction tasks, potentially offering fresh insights and advances for the community.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (U23B2009, 92270116).

References

Agustsson, E.; and Timofte, R. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022a. Simple Baselines for Image Restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 13667 of *Lecture Notes in Computer Science*, 17–33. Springer.

Chen, X.; Pan, J.; Jiang, K.; Li, Y.; Huang, Y.; Kong, C.; Dai, L.; and Fan, Z. 2022b. Unpaired deep image deraining using dual contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017–2026.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38: 295–307.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR)*.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. W. 2017. Removing Rain from Single Images via a Deep Detail Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1715–1723.

Gui, J.; Chen, T.; Cao, Q.; Sun, Z.; Luo, H.; and Tao, D. 2023a. A Survey of Self-Supervised Learning from Multiple Perspectives: Algorithms, Theory, Applications and Future Trends. *ArXiv*, abs/2301.05712.

Gui, J.; Cong, X.; Cao, Y.; Ren, W.; Zhang, J.; Zhang, J.; Cao, J.; and Tao, D. 2023b. A Comprehensive Survey and Taxonomy on Single Image Dehazing Based on Deep Learning. *ACM Comput. Surv.*, 55(13s).

Hui, Z.; Gao, X.; Yang, Y.; and Wang, X. 2019. Lightweight Image Super-Resolution with Information Multi-distillation Network. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM)*, 2024–2032.

Jiang, J.; Wang, C.; Liu, X.; and Ma, J. 2023. Deep Learning-Based Face Super-Resolution: A Survey. In *ACM Comput. Surv.*, volume 55.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 8346–8355.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, Z.; Wang, X.; Jiang, J.; and Lin, C.-W. 2021. Rain-Free and Residue Hand-in-Hand: A Progressive Coupled Network for Real-Time Image Deraining. *IEEE Transactions on Image Processing*, 30: 7404–7418.

Lamba, M.; and Mitra, K. 2021. Restoring Extremely Dark Images in Real Time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3487–3497. Computer Vision Foundation / IEEE.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 11211, 262–277. Springer.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L. V.; and Timofte, R. 2021. SwinIR: Image Restoration Using Swin Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 1833–1844.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Lee, K. M. 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1132–1140.

Liu, X.; Ma, Y.; Shi, Z.; and Chen, J. 2019. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7313–7322.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 9992–10002.

Lu, Y.; Lin, Y.; Wu, H.; Luo, Y.; Zheng, X.; Xiong, H.; and Wang, L. 2023. Priors in Deep Image Restoration and Enhancement: A Survey. arXiv:2206.02070.

Lu, Z.; Li, J.; Liu, H.; Huang, C.; Zhang, L.; and Zeng, T. 2022. Transformer for Single Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 456–465. IEEE.

Nah, S.; Hyun Kim, T.; and Mu Lee, K. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3883–3891.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020a. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 11908–11915. AAAI Press.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020b. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 11908–11915.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Simonyan, K.; and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICLR)*.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.

Su, J.; Xu, B.; and Yin, H. 2022. A Survey of Deep Learning Approaches to Image Restoration. *Neurocomput.*, 487(C): 46–65.

Wang, H.; Wu, Y.; Li, M.; Zhao, Q.; and Meng, D. 2021a. Survey on rain removal from videos or a single image. In *Science China Information Sciences*, volume 65.

Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; and Guo, Y. 2021b. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10581–10590.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. H. 2019. Spatial Attentive Single-Image Deraining With a High Quality Real Rain Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12270–12279. Computer Vision Foundation / IEEE.

Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023. Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2654–2662. AAAI Press.

Wang, Z.; Chen, J.; and Hoi, S. C. H. 2021a. Deep Learning for Image Super-Resolution: A Survey. In *IEEE Trans. Pattern Anal. Mach. Intell.*, volume 43, 3365–3387. United States.

Wang, Z.; Chen, J.; and Hoi, S. C. H. 2021b. Deep Learning for Image Super-Resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(10): 3365–3387.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A General U-Shaped Transformer for Image Restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17662–17672.

Wu, G.; Jiang, J.; Bai, Y.; and Liu, X. 2023a. Incorporating Transformer Designs into Convolutions for Lightweight Image Super-Resolution. arXiv:2303.14324.

Wu, G.; Jiang, J.; Jiang, K.; and Liu, X. 2023b. Fully 1×1 Convolutional Network for Lightweight Image Super-Resolution. arXiv:2307.16140.

Wu, G.; Jiang, J.; and Liu, X. 2023. A Practical Contrastive Learning Framework for Single-Image Super-Resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10551–10560.

Xiao, J.; Fu, X.; Liu, A.; Wu, F.; and Zha, Z.-J. 2022. Image De-raining Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–18.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017a. Deep Joint Rain Detection and Removal from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1685–1694.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017b. Deep Joint Rain Detection and Removal from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1685–1694.

Yang, W.; Tan, R. T.; Wang, S.; Fang, Y.; and Liu, J. 2021. Single Image Deraining: From Model-Based to Data-Driven and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11): 4059–4077.

Ye, Y.; Yu, C.; Chang, Y.; Zhu, L.; Zhao, X.-l.; Yan, L.; and Tian, Y. 2022. Unsupervised Deraining: Where Contrastive Learning Meets Self-similarity. In *Proceedings of theIEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5811–5820.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M. 2022. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5718–5729.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.*, 26(7): 3142–3155.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3262–3271.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In *Proceedings of the European conference on computer vision (ECCV)*, 286–301.

Zheng, Y.; Zhan, J.; He, S.; Dong, J.; and Du, Y. 2023. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5785–5794.