HIRL: HYBRID IMAGE RESTORATION BASED ON HIERARCHICAL DEEP REINFORCEMENT LEARNING VIA TWO-STEP ANALYSIS

Xiaoyu Zhang[1,2], *Wei Gao*[1,2*]

[1]School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[2]Peng Cheng Laboratory, Shenzhen, China

ABSTRACT

The restoration of hybrid distorted images in real-world scenarios is still a difficult problem due to the fact that the degrading types and degrees are always unknown. Previous studies typically utilize multiple recovery tools to restore images. However, each tool adopted inevitably introduces additional noise and will affect the subsequent recovery results. To address this issue, in this paper, we propose a hierarchical deep reinforcement learning framework (HIRL), which balance both benefits and noises brought by each tool and select the appropriate type and degree tools. The proposed method endeavors to find a long-term optimal tool sequence, which is better than the greedy strategy that employs the tools with the largest short-term returns. Meanwhile, it benefits from a hierarchical design to reduce time consumption and complexity compared to a brute force strategy. Experiments demonstrate the superiority of the proposed method over other state-of-the-art methods. Furthermore, our framework is highly scalable and can be easily extended to the other recovery pipelines.

Index Terms— Hybrid Image Restoration, Hierarchical Reinforcement Learning, Deep Learning

1. INTRODUCTION

Image restoration is a general term for a series of low-level computer vision tasks, including denoising [1, 2], compression artifact removal [3, 4] and deblurring [5], etc. One of the important tasks is hybrid distortion recovery, which focuses on recovering images with a mix of multiple distortions.

Recently, it has been proposed to deal with the hybrid distortion problem by convolutional neural networks (CNNs). [1, 2, 4, 3, 5] use a single convolutional neural network for end-to-end recovery of hybrid distorted images. [6] and [7] design an attention model and a multi-branch network to deal with different parts of distortion, respectively. Although these methods can partly improve the performance of recovery by using elaborate and complex networks, such equal processing of various distorted images (e.g., mild and severe degrees) manner would introduce unnecessary time complexity. To address this issue, iterative image recovery by a series of meta-operators is proposed [8, 9, 10]. They operate with different intensities for different degrees of distortion, resulting in problem-adapted complexity. Since meta-operators inevitably generate additional noise during use and thus affect the subsequent recovery process, it is necessary to take into account both the quality gain and introduced noise when adopting the operator. Brute force search finds the best recovery path by traversing the combination of all tools. However, its time complexity grows exponentially with the number of distortion types and operations grows. In contrast, the greedy strategy reduces the complexity, but it focuses on the tools that bring the most incentive in the moment, which can easily fall into local optimization. It is difficult to obtain high quality images with a greedy strategy in the long run. As a typical paradigm of machine learning, reinforcement learning (RL) [11] is able to search for the global optimal solution with linear complexity. In recent year, RL-based image restoration approaches [8, 9, 10] train RL agents to select tools for incremental recovery. However, the correlation between the recovery tool and hybrid distortion types has not been fully investigated. Moreover, there is no distinction between different types and degrees of tools for hybrid distortion problems.

To this end, we study the methodological noise when applying tools in different orders, and rethink the connection between reinforcement learning and hybrid image recovery pipelines. In this paper, we propose a hierarchical search architecture (HIRL) based on hierarchical reinforcement learning [12]. Specifically, we separate the distortion case into different hierarchies according to different types and degrees, which can be easily extended to other image restoration tasks by introducing the corresponding tools in the toolset. In addition, we improve the reward scheme by introducing a perception-driven image quality assessment [13], which yields more visually friendly restorations.

To summarize, our contributions are as follows. (1) We investigate the correlation between restoration path and degra-

*Corresponding author: gaowei262@pku.edu.cn

This work was supported by Natural Science Foundation of China (61801303, 62031013), Guangdong Basic and Applied Basic Research Foundation (2019A1515012031), Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806163656003), Shenzhen Science and Technology Plan Basic Research Project (JCYJ20190808161805519), and CCF-Tencent Open Fund (RAGR20200114).

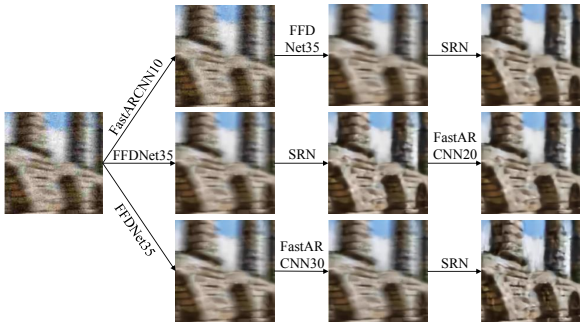

Fig. 1. Image restoration pipelines using different strategies. The three branches are the reverse order of the degrading process, greedy strategy, and brute force search, respectively.

dation process, and reveal the methodological noise brought by the tools. (2) We propose a hierarchical deep reinforcement learning based adaptive restoration pipeline (HIRL) and design a hierarchical set of tools based on distortion types and degrees. (3) We extend blind image quality assessment (BIQA) method in the reward scheme to guide RL agents for iterative recovery in scenarios without reference images. (4) Experimental results show that the proposed framework can effectively cope with hybrid distortion scenarios and outperforms other state-of-the-art methods.

2. PROPOSED METHOD

2.1. Hierarchical Design of Toolset

Taking into account the parameters and complexity of the model, we select a group of light-weight networks for distortion reductions. For denoising tasks, we use FFDNet [2] with different noise estimation maps distributed at 15, 25, 35, respectively. To solve compression artifacts problems, we reproduce FastARCNN [4] with training under different compression quality factors (QFs) of 30, 20, 10. Regarding deblurring, Sharpen [14], SRCNN [15] and SRN [5] are selected and sorted by their performance on Gaussian deblurring.

2.2. Analysis of Recovery Path

We perform hybrid distortion recovery on DIV2K [16] dataset using different tool combination strategies, and find that neither the degenerate reverse process nor greedy strategy can achieve optimal recovery. We take one of the recovery cases as shown in Fig. 1. The corresponding degradation processes are motion blur of 25, additive Gaussian noise distributed at 15, and JPEG compression with a quality factor of 10. The distorted image is enhanced with three different strategies: the reverse order of the degradation process, greedy strategy and brute force search. The top branch is recovered by the reverse order of the degradation process, performing FastARCNN10, FFDNet35 and SRN, respectively, with a PSNR gain of 1.157dB. The middle branch is the result of greedy strategy with the corresponding combination of tools FFDNet35, SRN and FastARCNN20, respectively, and the PSNR is improved by 1.323dB. The bottom branch represents the brute force search strategy, obtained with FFDNet35, FastARCNN30 and SRN, and the PSNR has increased by 1.426dB. The poor performance of the first strategy implies a non-correspondence between the optimal recovery path and the degradation process, and the FastARCNN10 used in the first step actually introduces more information loss. The greedy strategy also falls into a local optimum due to the additional noise introduced by the SRN in the second step, which affects the overall performance. The brute force search strategy finds the best recovery path by traversing all possible combinations, and thus considers both the quality improvement and the noise brought by the tool. However, the time complexity of brute force search grows exponentially as the length of repair paths and tool sets increase. Therefore, we propose a framework based on the hierarchical reinforcement learning. On the one hand, the RL-based training strategy enables the framework to search for globally optimal solutions, and on the other hand the hierarchical structure reduces the time complexity to a linear level.

2.3. Proposed Hierarchical Framework

Inspired by HIRO [17], the proposed hierarchical structure is illustrated as Fig. 2, we divide the restorations as two steps: type selection and degree selection. Specifically, when the RL agent selects a restoration tool, it first judges the tool type to be executed through the type strategy (TS) module and then chooses the restoration degree of the tool through degree strategy (DS) module. In each of these two modules, there are two networks for tool selection, i.e., the evaluation network and the target network, which are implemented by ResNet18 [18]. The output of the evaluation network is defined as the long-term value V of each tool, which is the maximum reward achieved after considering all possible recovery paths after taking the tool. Following the study of temporal difference method (TD) [19, 20], we replace all subsequent operations by the value of the next state:

$$V_s = \max_a \left(r^a_{s,s'} + \gamma V_{s'} \right), \tag{1}$$

where s, s' represent the images before and after each enhancement step, a indicates the action of using a recovery tool, and γ is an attenuation factor. $r^a_{s,s'}$ represents the reward from applying the current recovery tool, which is characterized by peak signal-to-noise ratio (PSNR). The target network inherits the parameters of the evaluation network with regular intervals, which is used to make the training more stable.

Training Process. First of all, we input the distorted image s to TS evaluation network and output value $T(s)$ of each tool type. The type t with the greatest value is selected in a probability of 1-ε, or the type is randomly selected in probability of ε. The DS evaluation network works similarly with the TS

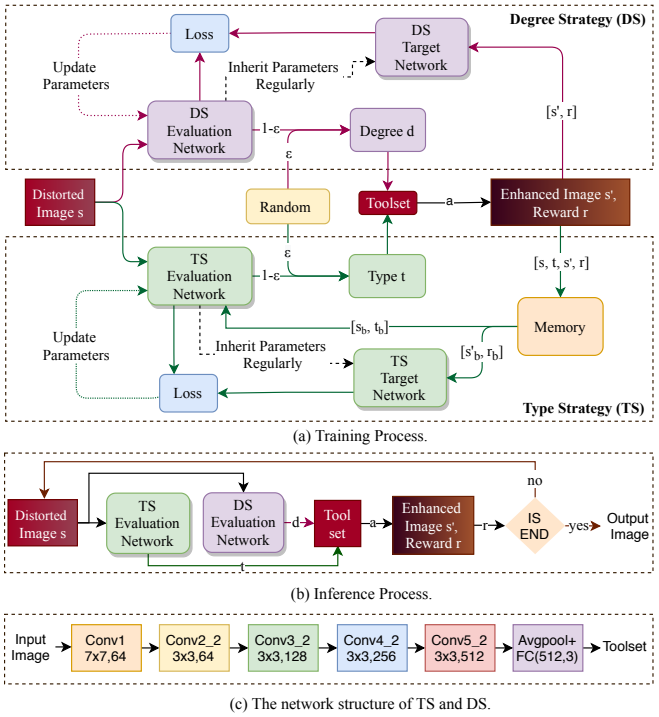

Fig. 2. The proposed framework for efficient image restoration based on hierarchical reinforcement learning. (a) Training process with the details of degree strategy module and type strategy module. (b) Inference process. (c) The network structure of TS and DS, which is implemented by ResNet18.

evaluation network and outputs the selected degree d. After that, we apply the tool a corresponding to t, d to the current image s to obtain the enhanced image s' and reward r. Then the DS evaluation network is trained by minimizing:

$$Loss_{DS} = \|r + \gamma D'_{max}(s', :) - D(s, d)\|_2^2, \qquad (2)$$

where $D(s, d)$ represents the value of degree d for image state s from DS evaluation network. $D'_{max}(s', :)$ is the largest value obtained after processing s' with DS target network. Hereafter, we store the distorted image s, tool type t, enhanced image s' and reward r as a transition $[s, t, s', r]$. After storing a certain number of transitions, we sample a batch of transitions from memory to update the TS evaluation network:

$$Loss_{TS} = \|r_b + \gamma T'_{max}(s'_b, :) - T(s_b, t_b)\|_2^2, \qquad (3)$$

where $T(s_b, t_b)$ is a batch values of types t_b for image states s_b from TS evaluation network. $T'_{max}(s'_b, :)$ is the maximum values for image states s'_b from TS target network. After this training, we establish a hierarchical framework for further tool selection.

Inference Process. For inference, the type and degree of the tools are selected sequentially by the TS and DS evaluation networks. The images are then recovered step by step through the tools until there is no reward, where the reward is measured by PSNR and PaQ-2-PiQ [13]. PaQ-2-PiQ is introduced to guide recovery in scenes without reference images, while it is more consistent with the quality of the human visual system due to its training on a subjective dataset.

3. EXPERIMENTAL RESULTS

3.1. Experimental Setup

Datasets. We perform experiments on 800 images from DIV2K [16] for training. The datasets of CSIQ [21] and 100 images from DIV2K [16] are used for testing. There are 30 reference images in CSIQ, and each distortion has 5 distortion levels. We choose additive white Gaussian noise, JPEG, and blur datasets to test the performance of the model on single distortion. The hybrid distortions are degraded on DIV2K with three levels: mild, moderate and severe. Specifically, the mild distortion is Gaussian noise uniformly distributed at 15, JPEG with quality factor of 30 and Gaussian blur with kernel of 5. The moderate dataset is degraded by Gaussian noise 25, JPEG 20 and Gaussian blur 10. And the severe dataset is degraded by Gaussian noise 35, JPEG 10 and Gaussian blur 15. Moreover, we experiment on CLIVE dataset [22] to verify the performance of the model in the real scenarios.
Training Details. The single-network approaches are trained under Gaussian blur with a kernel of {5,10,15}, JPEG compression with a quality factor of {10, 20, 30}, Gaussian noise whose variance is distributed at {15, 25, 35}, and hybrid distortions, respectively, each of which contains 2400 images. The RL agents of PixelRL, RL-Restore, and HIRL are trained on the hybrid distortion datasets, respectively.
Network Configuration. We initialize ε to 0.5 and then decrease it by 0.1 every 100 epochs until 0.1. γ is set to 0.9. The memory size of transitions is 100 and the sampling numbers is 30. Moreover, the target network inherit parameters from the evaluation network after every 30 times of training.

3.2. Performance Comparisons

We present quantitative results on test datasets in Table 1. By experimenting with specific distortion cases on CSIQ dataset, we find that the adaptive use of multiple tools have better quality compared with single tools in the toolset. For experiments on multiple distortion cases of DIV2K, the proposed HIRL outperforms other approaches due to the comprehensive hierarchical toolset and the ability of the hierarchical framework to find a better combination of tools. The visualization results are shown in Fig. 3, which shows that proposed HIRL can effectively suppress noise and blocking effects, and reproduce fine texture details. In addition, the experiments on CLIVE dataset are shown in Fig. 4. The introduction of blind quality assessment metric helps to improve the recovery quality by guiding RL agents in the absence of a reference image.

Table 1. Comparisons of PSNR (dB) and SSIM for different distortion types on CSIQ and DIV2K datasets.

Distortion	SRCNN [15]	FastARCNN [4]	SRN [5]	FFDNet [2]	PixelRL [9]	RL-Restore [8]	HIRL
AWGN	28.26/0.6785	31.95 /0.8515	29.41/0.7298	32.59/0.8435	31.21/0.8121	28.39/0.6974	**33.54/0.8894**
JPEG	28.45/0.7992	28.92/0.8169	28.46/0.7992	28.91/0.8160	27.95/0.7901	26.40/0.7267	**29.23/0.8247**
BLUR	28.76/0.8164	28.28/0.8057	29.04/0.8123	28.28/0.8058	27.43/0.7897	25.64/0.7354	**29.80/0.8527**
Mild	22.91/0.5129	25.04/0.6148	22.93/0.5130	25.17/0.6661	25.25/0.6251	25.20/0.6720	**25.82/0.6861**
Moderate	21.32/0.3254	24.28/0.6145	22.48/0.4114	25.19/0.6136	23.09/0.5280	25.52/0.6563	**26.18/0.6626**
Severe	14.32/0.0886	21.00/0.3524	16.81/0.1404	21.31/0.3997	14.92/0.0985	23.33/0.5473	**23.81/0.5665**
Average	24.00/0.5368	26.58/0.6759	24.86/0.5677	26.91/0.7908	24.97/0.6073	25.74/0.6725	**28.06/0.7470**

Table 2. Comparisons of PSNR gains (dB) for brute force and the proposed method with the same trials.

Schemes	9 Trials	90 Trials	819 Trials	7380 Trials
Brute Force	1.338	1.586	1.721	1.847
HIRL	**1.618**	**1.866**	**1.958**	**1.958**

Table 3. Ablation study on enhancement path lengths by using PSNR (dB) and SSIM.

Distortion	path=3	path=5	path=7
mild	25.41/0.6763	25.72/0.6785	**25.82/0.6861**
moderate	26.09/0.6626	26.16/0.6646	**26.18/0.6677**
severe	23.70/0.5639	23.71/0.5664	**23.81/0.5665**

3.3. Ablation Study

We compare HIRL with brute force search on the DIV2K hybrid distortion dataset, as shown in Table 2. We compare the PSNR gain of both for equal number of search steps. For brute force search strategy, the number of search steps corresponding to the effective path length n is $\sum_{i\in[1,n]} m^i$, where m is the size of the toolset. We enumerate the cases where n is $\{1, 2, 3, 4\}$ and the number of search steps are $9, 90, 819,$ and 7380, respectively. Experiments results show that the proposed method can obtain more gains with the same time complexity. Moreover, we perform experiments with limited path lengths of 3, 5, 7 on DIV2K distorted images. As seen in Table 3, the increased number of paths contributes to more sufficient explorations and produces higher quality results.

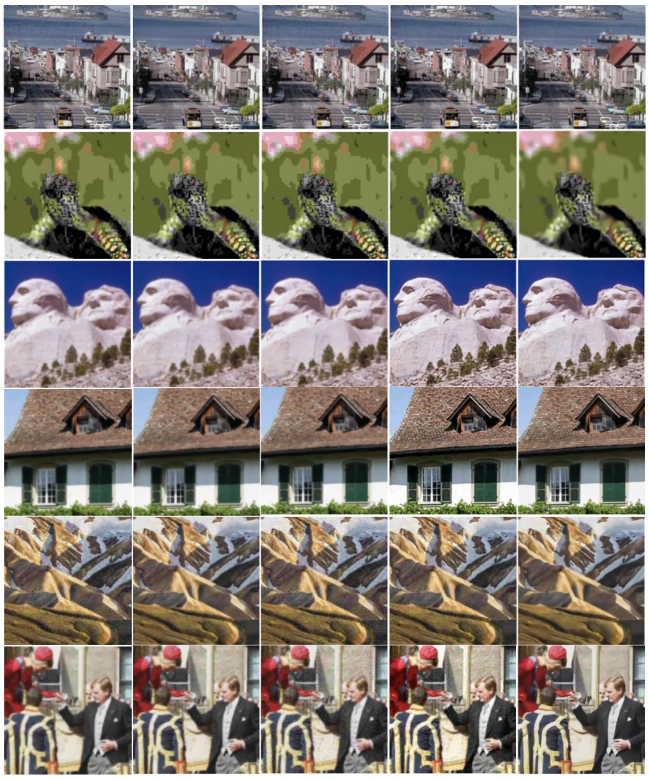

Fig. 3. Comparison of visual image quality on CSIQ and DIV2K datasets. The results are presented from top to bottom for AWGN, JPEG, BLUR with single distortion and mild, moderate, and severe with hybrid distortions.

4. CONCLUSION

We propose a novel image restoration method based on hierarchical reinforcement learning (HIRL), which can effectively remove mixed distortion artifacts. Firstly, we illustrate that the alternate use of tools would introduce method noise, and thus take advantage of the delay satisfaction characteristics of reinforcement learning to obtain the global optimal recovery path. Secondly, we propose to explore the hierarchical structure of complex problems, which can better model the restoration pipeline and is easy to be extended for more distortion cases. Finally, we present a blind image quality assessment method that guides the framework to recover in a visually friendly direction in the absence of a reference image. Extensive experimental results demonstrate the superiority of our proposed method over the other state-of-the-art methods. Furthermore, more decoupled tools and more comprehensive evaluation metrics for RL guidance in the absence of reference images should be explored in the future.

5. REFERENCES

[1] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual

Fig. 4. Comparison of image restoration performances for different methods on the wild dataset CLIVE, including RL-Restore, and the proposed method without and with blind image quality assessment (BIQA) metric of PaQ-2-PiQ.

learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[2] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[3] Xiaoshuai Zhang, Wenhan Yang, Yueyu Hu, and Jiaying Liu, "Dmcnn: Dual-domain multi-scale convolutional neural network for compression artifacts removal," in *ICIP*, 2018, pp. 390–394.

[4] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang, "Compression artifacts reduction by a deep convolutional network," in *ICCV*, 2015, pp. 576–584.

[5] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia, "Scale-recurrent network for deep image deblurring," in *CVPR*, 2018, pp. 8174–8182.

[6] Masanori Suganuma, Xing Liu, and Takayuki Okatani, "Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions," in *CVPR*, 2019, pp. 9039–9048.

[7] Guocheng Qian, Jinjin Gu, Jimmy S Ren, Chao Dong, Furong Zhao, and Juan Lin, "Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution," *arXiv preprint arXiv:1905.02538*, 2019.

[8] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy, "Crafting a toolchain for image restoration by deep reinforcement learning," in *CVPR*, 2018, pp. 2443–2452.

[9] Ryosuke Furuta, Naoto Inoue, and Toshihiko Yamasaki, "Pixelrl: Fully convolutional network with reinforcement learning for image processing," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1704–1719, 2020.

[10] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy, "Path-restore: Learning network path selection for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.

[11] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[12] Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum, "Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation," in *NeurIPS*, 2016, pp. 3675–3683.

[13] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik, "From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality," in *CVPR*, 2020, pp. 3572–3582.

[14] Youngbae Kim, Yeong Jun Koh, Chulwoo Lee, Sehoon Kim, and Chang Su Kim, "Dark image enhancement based onpairwise target contrast and multi-scale detail boosting," in *ICIP*, 2015.

[15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2015.

[16] Eirikur Agustsson and Radu Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *CVPRW*, July 2017.

[17] Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine, "Data-efficient hierarchical reinforcement learning," in *NeurIPS*, 2018, pp. 3307–3317.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[19] Gerald Tesauro et al., "Temporal difference learning and td-gammon," *Communications of the ACM*, vol. 38, no. 3, pp. 58–68, 1995.

[20] Dimitri Bertsekas, *Abstract dynamic programming*, Athena Scientific, 2018.

[21] Eric Cooper Larson and Damon Michael Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, pp. 011006, 2010.

[22] Deepti Ghadiyaram and Alan C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, 2015.