Improving De-raining Generalization via Neural Reorganization

Jie Xiao†, Man Zhou†, Xueyang Fu*, Aiping Liu, Zheng-Jun Zha
University of Science and Technology of China, China
{ustchbxj,manman}@mail.ustc.edu.cn, {xyfu,aipingl,zhazj}@ustc.edu.cn

Abstract

Most existing image de-raining networks could only learn fixed mapping rules between paired rainy/clean images on single synthetic dataset and then stay static for lifetime. However, since single synthetic dataset merely provides a partial view for the distribution of rain streaks, deep models well trained on an individual synthetic dataset tend to overfit on this biased distribution. This leads to the inability of these methods to well generalize to complex and changeable real-world rainy scenes, thus limiting their practical applications. In this paper, we try for the first time to accumulate the de-raining knowledge from multiple synthetic datasets on a single network parameter set to improve the de-raining generalization of deep networks. To achieve this goal, we explore Neural Reorganization (NR) to allow the de-raining network to keep a subtle stability-plasticity trade-off rather than naive stabilization after training phase. Specifically, we design our NR algorithm by borrowing the synaptic consolidation mechanism in the biological brain and knowledge distillation. Equipped with our NR algorithm, the deep model can be trained on a list of synthetic rainy datasets by overcoming catastrophic forgetting, making it a general-version de-raining network. Extensive experimental validation shows that due to the successful accumulation of de-raining knowledge, our proposed method can not only process multiple synthetic datasets consistently, but also achieve state-of-the-art results when dealing with real-world rainy images.

1. Introduction

In recent years, deep learning models have achieved significant progress on single image de-raining task [32, 6, 4, 34, 43, 13, 50, 41, 28, 5, 42, 39, 33, 18]. This task aims to recover the clean image from its rain-polluted version, benefiting subsequent downstream computer vision tasks [16, 2, 19, 15, 31, 46, 26, 49, 25], *e.g.*, object detection, image classification, person identification, *etc*.

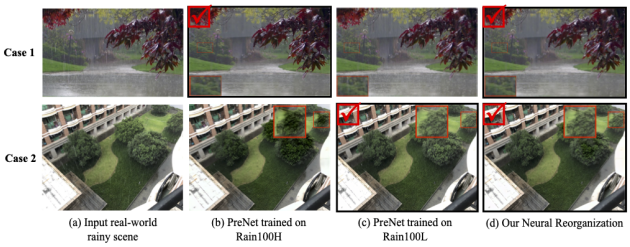

Figure 1: Visual comparison between the state-of-the-art PreNet and our Neural Reorganization on real-world rainy scene with heavy and light rain. The **red check mark** indicates that the network successfully works on it. Specifically, it is clear that our proposed Neural Reorganization scheme is capable of dealing with various real-world rainy scenes (*i.e.*, heavy and light rain). While the original PreNet could only take effect on specific type of rain streak and fail on untouched type of training stage.

While existing methods obtain promising results, most of them only focus on learning specific mapping rules on the given individual synthetic rainy dataset. They may suffer from severe performance degradation when applying to real-world scenes, as the learned specific mapping rule on given individual synthetic rainy dataset is incapable of describing them. As shown in Figure 1, taking the state-of-the-art de-raining network PreNet [29] for example, we first train PreNet on the given light rainy dataset Rain100L [38] and then test it on the collected real-world rainy samples from the Real-Internet dataset [35]. It is clear that the trained PreNet only works well on light rainy scene and fails to remove rain-streaks on heavy rainy samples (Case 1). Similarly, when trained on the given heavy rainy dataset Rain100H [38], the trained PreNet suffers from severe artifacts (Case 2) as shown in red box. Therefore, how to obtain a general-version image de-raining network (GID), which can tackle various types of rain streaks with a single set of parameters and generalizing well to complex real-world rainy scenes, has attracted more and more attention.

Inspired by high-level computer vision tasks, *e.g.*, im-

*Corresponding author. †Co-first authors contributed equally

age classification, a straightforward strategy to solve the above issue is to train a deep network by mixing multiple types of synthetic rainy datasets. However, this mix-up strategy is exposed to the following weaknesses: 1. The gradient from each task pulls the solution towards its optimum and the result is an equilibrium between the gradients of different tasks [10]. Since the contribution of each individual dataset to the total loss varies greatly, the optimization algorithm inclines to overlook the disadvantaged member of the mixed datasets. The biased allocation leads to undesirable performance when facing complex real-world scenes. 2. This mixing strategy leads to low computational efficiency in practice. Specifically, when a new dataset is introduced, the approach requires to remix all these datasets and retrain network from scratch. On the other hand, training the network on a chronological sequence of synthetic datasets rather than mixed datasets also acts as a possible approach to achieve GID. The main issue of the de-raining model regarding sequentially learning is that it is prone to catastrophic forgetting or catastrophic interference [27], *i.e*., training a model with new information interferes with previously learned knowledge. Alternatively, some researchers [21, 24, 20, 36, 43] focus on domain adaption between synthetic and real-world scenes. However, these methods could only learn specific one-to-one cross-domain mapping rules between individual synthetic dataset and real-world rainy images. When applying to multiple synthetic rainy datasets, these methods are incapable of learning the many-to-many mapping rules between all the synthetic domains and real-world one in a single network. Besides, all of them require the collection of a large number of real-world samples during the training phase.

Although individual synthetic dataset only provides a partial view, combining multiple local views is conducive to inferring the panorama. If the de-raining model can make reasonable use of these multi-view datasets, it may gain remarkable promotion on generalization ability. However, existing methods [45, 40, 34, 29, 22, 17, 47, 38, 8] mainly fucus on designing more effective architecture of network to achieve better performance on single synthetic dataset. The de-raining networks obtained by the above methods resembles a static entity of de-raining knowledge. If the de-raining knowledge is directly extended without targeting the original dataset, it will lead to the catastrophic forgetting problem [3]. To address the above issues, in this paper, we propose a novel Neural Reorganization (NR), a brain-inspired scheme to endow the de-raining network with a dynamic stability-plasticity trade-off rather than naively pure stability after training phase. Specifically, we borrow the synaptic consolidation mechanism in the biological brain and knowledge distillation to design our NR algorithm. In this way, our NR algorithm enables the deep network to be trained on a list of synthetic rainy datasets by overcoming catastrophic forgetting. Therefore, the final trained network could aggregate de-raining knowledge from multiple datasets. In addition to improving the generalization ability on real-world rainy images, the trained network can also maintain considerable performance on all the synthetic datasets. Our main contributions are as follows:

1) To our best knowledge, it is the first attempt to introduce brain-inspired mechanism to improve the image de-raining generalization issue. Compared with mainstream transfer learning methods, our proposed method only requires synthetic rainy datasets that are easily accessible.
2) Inspired by learning and memory mechanism in the biological brain, we first propose Neural Reorganization. Our proposed algorithm facilitates a de-raining network to effectively accumulate de-raining knowledge from a continuous stream of correlated data by overcoming catastrophic forgetting.
3) Since our proposed NR algorithm is orthogonal to the mainstream de-raining methods focusing on the network architectures, it can be directly applied to these methods to improve the generalization ability.
4) Extensive experiments demonstrate that a single network trained with our proposed scheme can process multiple synthetic datasets consistently and achieve state-of-the-art results over real-world rainy scenes. In addition, we believe that our NR can provide a new perspective for other related low-level vision tasks in terms of improving generalization ability.

2. Related Work

Image de-raining network. Image de-raining aims to recover the clear images from its rainy counterparts. Recent remarkable progress has been achieved over the image de-raining field due to the exploitation of deep learning technology. Fu *et al*. [8, 7] propose a customized neural network for image rain removal, which outperforms handcrafted image de-raining methods by a large margin. The rainy images are decomposed into high- and low-frequency parts, and then fed them into the network for rain removal and enhancement, respectively. Yang *et al*. [38] introduce a novel de-raining pipeline by recursively taking full use of stage-wise de-rained results for removing different levels of rain streaks. To further improve the de-raining performance, researchers design an enormous number of advanced deep network architectures, including non-local module equipped encoder-decoder network [17], multi-stage de-raining neural network [47], conditional generative adversarial network [48], recurrent neural network [22, 29] and so on. However, the above deep learning-based methods lack of sufficient explanation due to roughly stacking black-box neural layers in an end-to-end

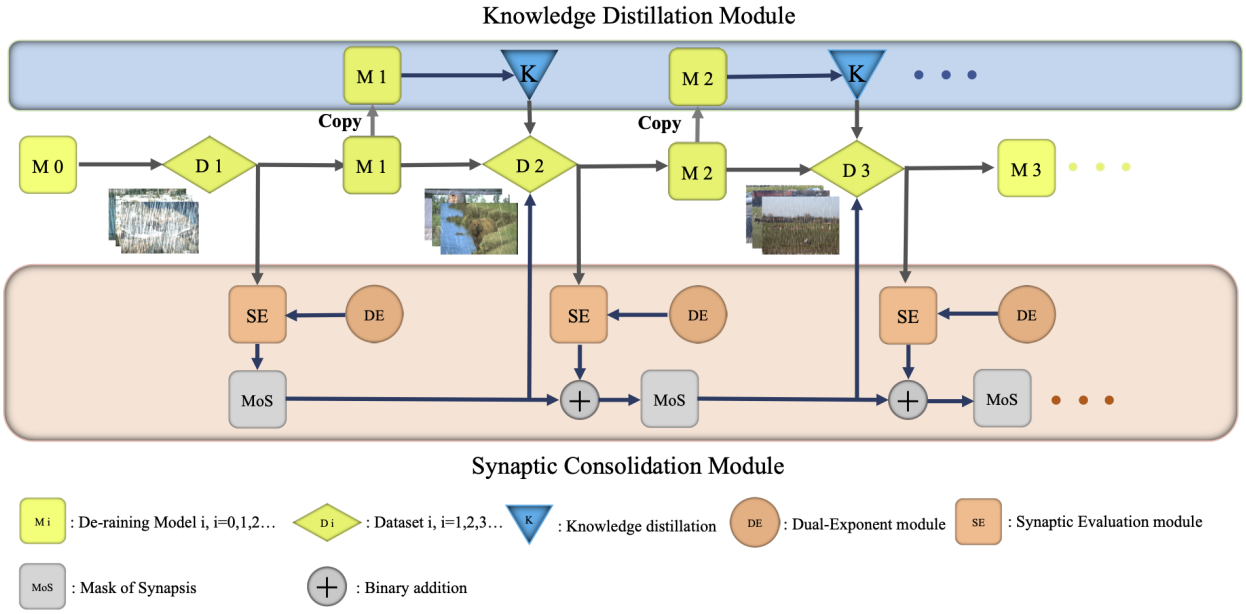

Figure 2: The detailed flowchart of our proposed Neural Reorganization scheme. It allows network to incrementally accumulate various de-raining knowledge from a sequence of synthetic datasets by overcoming catastrophic forgetting. Specifically, Neural Reorganization mainly consists of Synaptic Consolidation Module and Knowledge Distillation Module. Synaptic Consolidation Module aims to maintain a mask (MoS in figure), which is utilized to shield any changes to parameters important to previous datasets. Knowledge Distillation Module is responsible for instructing the model to learn new knowledge on the new dataset by knowledge distillation ONLY on the new dataset while maintaining the memories of previous tasks.

fashion, which hinders their further improvement. To this end, some researchers turn their attention on designing interpretable de-raining networks. In the representative work [34], Wang *et al*. first analyze the prior terms of both image and rain streaks, and then integrate them into a convolutional sparse coding inspired networks to implement image de-raining. The signal flow of the de-raining network keeps consistent with the optimization procedure. While existing deep learning-based methods [32, 6, 4, 34, 43, 13, 5] have obtained promising performance, most of them could only learn fixed mapping rules between paired rainy/clean images on a single type of rainy dataset. When dealing with multiple datasets, these methods cannot directly use a single parameter set to cover all the datasets due to catastrophic forgetting. To solve this problem, we explore a brain-inspired neural reorganization to endow the network with stability-plasticity trade-off, flexibly adapting to various rainy scenarios.

Image de-raining generalization. The other topic relevant to our work is how to improve the de-raining generalization ability. Jin *et al*. [14] employ an unsupervised de-raining generative adversarial network to tackle the generalization problem by introducing self-supervised constraints from the intrinsic statistics of unpaired rainy and clean images. Wei *et al*. [36] develop a semi-supervised transfer learning method, which also extract and exploit statistical prior to align the synthetic rainy domains and the real-world domain. Rajeev *et al*. [43] propose a Gaussian Process-based semi-supervised learning framework, which enables the network in learning to de-rain using synthetic datasets while generalizing better using unlabeled real-world images. Lin *et al*. [24] propose a two-stage weakly-supervised data distillation approach, which aligns unpaired rainy and clean images to generate supervision in a coarse-to-fine manner. Different from the above methods, our approach possesses two distinguishing characteristics: 1) our method acts as an effective training strategy and can be easily extended to existing de-raining networks; 2) our method only requires synthetic datasets during the training stage, which is more friendly for practical application. The reason is that the synthetic datasets are easier to obtain while collecting high quality real-world samples is difficult and laborious.

3. Neural Reorganization

Recent years have witnessed promising advance in single image de-raining task. However, due to various challenges in obtaining real-world labeled image de-raining datasets, existing methods are trained only on fixed synthetically generated dataset, which only provides a limited perspective for the distribution of rain streaks. This results in a huge

obstacle for de-raining network trained on single synthetic dataset to generalize to real-world rain scenes. Inspired by learning and memory mechanisms in the biological brain [37, 9], we propose Neural Reorganization to endow the de-raining network with dynamic stability-plasticity trade-off after traditional training phase so that de-raining model with a single parameter set could be competent across various types of rain streaks provided by multiple synthetic datasets, which points out a novel way to improve generalization ONLY with synthetic datasets.

3.1. Synaptic Consolidation

One prominent characteristic of the mammalian brain is its capacity to integrate new information throughout life while stably maintaining memories. These two seemingly mutually exclusive properties of the brain are coincident with plasticity-stability trade-off of synaptic connections [37]. Not only are synaptic connections capable of undergoing rapid changes in response to new experience but also can serve as substrates for long-term information storage. Researches [37, 9] have revealed that novel experience could lead to the formation of synaptic connections, which provide a structural basis for learning, but most of these connections are eliminated by a protracted process. Then, these surviving connections, together with most connections formed early during development and surviving experience-dependent elimination, are preserved and provide a structural basis for memory retention throughout the entire life of an animal. Inspired by these findings, we design Synaptic Consolidation mechanism to shield changes to important synapses when learning on the new task.

Clearly, the proposed Synaptic Consolidation should be composed of two parts, with one concentrating on figuring out importance of synapses, and the other aiming to determine the ratio of synapses that should be consolidated. Since what really matters in this work is the ranking of synaptic importance, which reminds us about the most influential portion of synapses for the current task and facilitates subsequent consolidation of these synapses, we ONLY need a reasonable way to calculate the importance of synapses. In this paper, we accumulate the gradients over the given data points to obtain importance weight Ω:

$$\Omega = \mathop{\mathbb{E}}_{x\sim\mathbb{P}^n}\left[\left|\frac{\partial\mathcal{L}(f(x;\theta),y)}{\partial\theta}\right|\right], \tag{1}$$

where $f(\cdot)$ and $\mathcal{L}(\cdot,\cdot)$ refer to the de-raining model and the objective loss function respectively, $|\cdot|$ denotes element-wise absolute operator and $\mathbb{P}$ denotes distribution of rainy images. The detailed description of Eq. (1) can be found in supplemental file.

Inspired by decreasing tendency of synaptic plasticity according to age in biological brain and the exponential model proposed in [37, 9] to describe changes in total spine [1] number as a function of age, we design a Dual-Exponent module to decide the proportion of synapses which are remarkably contributive to the performance of current task and need to be consolidated to overcome catastrophic forgetting during follow-up learning. The Dual-Exponent module is of the mathematical form:

$$P(n) = \alpha_s e^{-\frac{n}{\tau_s}} + \alpha_l e^{-\frac{n}{\tau_l}}, \tau_l \gg \tau_s > 0, n = 0, 1, 2\ldots, \tag{2}$$

where P denotes the calculated proportion of synapses to be consolidated, n denotes dataset id, τ_l and τ_s refer to time constants used to depict the rate of decline of synaptic plasticity. The first exponential component with relatively smaller time constant τ_s and larger coefficient α_s indicates a quite proportion of parameters of network are free to update, imitating the great plasticity of the biological brain in the early stages of development. The second component of Eq. (2) corresponds to slower and relatively minute decline of plasticity, which is inspired by stability of synaptic connection in adulthood. Besides, due to de-raining network with fixed capacity, $P(n)$ should satisfy the normalization condition:

$$\sum_{n=0}^{\infty} P(n) = 1. \tag{3}$$

From mathematical form of $P(n)$ expressed in Eq. (2) and sum law of limits, Eq. (3) is converted to:

$$\sum_{n=0}^{\infty} \alpha_s e^{-\frac{n}{\tau_s}} + \sum_{n=0}^{\infty} \alpha_l e^{-\frac{n}{\tau_l}} = 1, \tau_l \gg \tau_s > 0, \tag{4}$$

which is composed of two terms of infinite geometric series. We take the closed form of the geometric series, so Eq. (4) is transformed:

$$\frac{\alpha_s}{1-e^{-\frac{1}{\tau_s}}} + \frac{\alpha_l}{1-e^{-\frac{1}{\tau_l}}} = 1. \tag{5}$$

From Eq. (5), $\alpha_s, \alpha_l, \tau_s, \tau_l$ are not independent hyper-parameters and should meet the constraint:

$$\alpha_s = (1 - \frac{\alpha_l}{1-e^{-\frac{1}{\tau_l}}})(1-e^{-\frac{1}{\tau_s}}). \tag{6}$$

Hence, we can freely determine the values of α_l, τ_l, and τ_s, then α_s is automatically determined by the Eq. (6). Equipped with Eq. (1) and (2), it is qualified to decide whether individual synaptic connection should be consolidated for current task. Specifically, for each parameter tensor θ_{ij}, Synaptic Consolidation produces a boolean tensor mask with the same dimension, of which the element "True" means the corresponding parameter is important for current task and should be protected from updating. Figure 3b presents the details of Synaptic Consolidation.

[1] A spine (or dendritic spine) is a small membranous protrusion from a neuron's dendrite that typically receives input from a single axon at the synapse.

Neural Reorganization Algorithm:
Start with:
Dataset 0, 1 …, N-1
Initialize:
model ← Train on Dataset 0
old_model ← model
Mask ←SC(model, Dataset 0, 0) // Synaptic Consolidation
Train:
for i in 1, 2…, N-1
model← KD(Dataset i, old_model, Mask) // Knowledge Distillation
Mask ←Mask ∪ SC(model, Dataset i, i)
old_model ← model

(a) The main frame of Neural Reorganization

Synaptic Consolidation (SC) :
Input:
model, Dataset, task_id
Output:
$Mask$ // boolean tensor
Process:
$\Omega_{ij} \leftarrow \frac{1}{M}\sum_{k=1}^{M}|\frac{\partial l(x_k;\theta_{ij})}{\partial \theta_{ij}}|$
Importance ← serializing and descending sort Ω
p ←$\alpha_s e^{-\frac{id}{\tau_s}} + \alpha_l e^{-\frac{id}{\tau_l}}$
index ← integer (p× #parameters)
Flag ← Importance[index] //thretshold to quantify Ω
$Mask_{ij} \leftarrow \Omega_{ij} >$ Flag //binary classification of synapses for current task:
// True means corresponding parameter is protected from updating

(b) Procedure for Synaptic Consolidation

Knowledge Distillation (KD):
Input:
Dataset, old_model, Mask
Output:
model
Process:
model ← old_model //initialize the model using old parameters
Repeat until convergence
for x_k, y_k in Dataset:
$y \leftarrow$ model(x_k), $y_{old} \leftarrow$ old_model(x_k)
tot_loss ← $\mathcal{L}(x_k, y) + \lambda\mathcal{L}(y, y_{old})$ //knowledge distillation on current Dataset
gradients ← Backward on tot_loss
gradients ← Mask(gradients) //shield changes to important synapsis
model ← Update(model, gradients) //update model using masked gradients

(c) Procedure for Knowledge Distillation

Figure 3: Procedure of our Neural Reorganization.

3.2. Knowledge Distillation

Except for synaptic consolidation, the biological brain also has profound mechanisms to remodel synapses when learning new knowledge so as to reconcile learning and memory. Inspired by [23], we exploit knowledge distillation [11] to imitate the "cautious" learning style in biological brains. Using only examples for the new dataset, we optimize both for performance on the new dataset and for preservation of responses on the previous tasks. Specifically, when model is trained on dataset n ($n > 0$), the total loss $\mathcal{L}_t$ should be of the form:

$$\mathcal{L}_t(x^n, y^n; \theta, \theta^{n-1}) = \mathcal{L}(f(x^n; \theta), y^n) + \lambda\mathcal{L}(f(x^n; \theta), f(x^n; \theta^{n-1})), \quad (7)$$

where $\mathcal{L}$ denotes conventional loss to train the network, θ^{n-1} denotes the optimized parameter set after training on dataset $n-1$, x^n and y^n refer to rainy input and corresponding label from dataset n, θ represents the parameter set to be optimized. It is noteworthy that knowledge distillation works in online manner as only data from dataset n is required in Eq. (7). Figure 3c shows the detailed procedure of Knowledge Distillation and the intact illustration of our proposed Neural Reorganization is in Figure 3.

4. Experiment

In this section, to validate the effectiveness of our proposed Neural Organization scheme, we first conduct extensive experiments by integrating it with the state-of-the-art de-raining network PreNet [29] for performance evaluation and model visualization. Then, we conduct the ablation to verify the function of proposed modules and support our idea that multi-view synthetic data is beneficial to model generalization. Finally, we present performance comparison with the representative transfer learning method: Syn2Real [43].

4.1. Dataset and Performance Metric

Synthetic data. We conduct model training ONLY on four widely-used synthetic datasets, including Rain100H [38], Rain100L [38], Rain800 [48] and Rain14000 [8] in this work. Both Rain100L and Rain100H are first proposed in [38] and composed of the rainy images with only one type and five types of rain streaks, respectively. Rain800 [48] consists of 700 training images and 100 testing images, which are chosen from UCID dataset [30] and BSD-500 dataset [1]. Rain14000 [8] includes 12,600 rainy images for training and 1,400 rainy images for testing, which are synthesized from 1,000 clean images with 14 kinds of different rain-streak orientations and magnitudes.

Real-world data. To measure the generalization ability of de-raining model in real-world scenarios, we conduct model evaluation on the recent public real-world rainy data set SPA-Data [35], which contains nearly 0.64 million rainy/clean image pairs for training and 1000 pairs for testing. In this work, we utilize the testing set, which includes 1000 real-world rainy images with their labeled clean images, of SPA-Data [35] to test the generalization ability of de-raining model. **It should be noted that we never train de-raining network on the training set of SPA-Data [35].** Besides, we also perform qualitative comparisons between our Neural Reorganation equipped network and their baseline counterparts on Real-Internet [35], which includes 146 real-world rainy images collected from Internet without ground truth.

Performance metric. In this paper, we employ the most common peak-signal-to-noise ratio (PSNR) [12] and structure similarity (SSIM) [44] as quantitative metrics of the

Training Strategy	Rain100H		Rain800		Rain14000		Rain100L		SPA-Data	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SI	18.01	0.611	23.70	0.760	28.48	0.865	37.44	0.978	34.91	0.951
Mix	26.74	0.861	23.74	0.843	30.98	0.927	31.53	0.948	34.85	0.955
our NR	25.93	0.851	24.53	0.816	30.42	0.907	36.34	0.973	**35.60**	**0.959**

Table 1: Comparison of quantitative results in terms of PSNR and SSIM. All the models are only trained on synthetic dataset Rain100H, Rain100L, Rain1400 and Rain800 without real-world SPA-Data. The corresponding experiment setting can be referred as section 4.3.

Test set / Training set	Rain100H		Rain100L		Rain14000		Rain800		SPA-Data	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain100H	**29.37**	**0.897**	34.71	0.969	28.39	0.878	22.45	0.810	34.50	0.952
Rain100L	17.83	0.611	**37.16**	**0.977**	27.60	0.851	23.58	0.755	34.99	0.953
Rain1400	15.16	0.430	28.90	0.897	**31.78**	**0.924**	23.03	0.807	33.70	0.952
Rain800	15.78	0.475	29.33	0.905	36.34	0.973	**25.86**	**0.871**	33.71	0.957
our NR	25.93	0.851	36.34	0.973	30.42	0.907	24.53	0.816	**35.60**	**0.959**

Table 2: Comparison of quantitative results in terms of PSNR and SSIM. In the middle four lines, the model is trained on individual synthetic dataset Rain100H, Rain100L, Rain1400 and Rain800 respectively and tested on all the aforementioned datasets as well as real-world rainy dataset SPA-Data. Our NR performs the best generalization over real-world scene.

model performance. Furthermore, we also provide qualitative comparisons about visual effect of image de-raining results between our Neural Reorganization and baselines.

4.2. Training Details

For fair comparison, all the parameters setting and training techniques keep consistent as reported in original paper. We train the model on the dataset sequence: Rain100H→Rain800→Rain14000→Rain100L. Inspired by biological advances [37], we set the value of the hyper-parameter α_l to 0.8%, which is the approximate proportion of surviving spines in the brain of adult mice when learning new skills. Besides, τ_s, τ_l is set to 1 and 10 respectively, satisfying the condition $\tau_l \gg \tau_s$. By Equation 6, the value of α_s is 0.5789. The coefficient λ of balancing the knowledge distillation loss and original training loss is set to 0.8. Further, all the experiments are implemented on NVIDIA GTX 1080Ti GPUs.

4.3. Results on Benchmark Datasets

To verify that our Neural Reorganization is able to promote generalization ability of de-raining network, we conduct both qualitative and quantitative experiments on the above datasets and performance metrics.

Baseline setup. One of the baselines is organized as sequentially and independently feeding multiple rainy datasets into the network for training (denoted by SI). In this setting, due to the catastrophic forgetting, the weights well-trained on the previous datasets are inevitably covered and updated by feeding the new rainy dataset, resulting in the rapid performance degradation on previous datasets. Besides, we also compare our method with the other baseline that the network is trained on mixed data of multiple datasets (Rain100H + Rain800 + Rain14000 + Rain100L). For convenience, we refer this strategy simply as Mix.

Quantitative comparison. Table 1 reports the comparisons between Neural Reorganization equipped networks and the corresponding baselines. Obviously, the SI baseline suffers from abrupt performance degradation on Rain100H, Rain800 and Rain14000 after completing training of the dataset sequence, which is attributed to catastrophic forgetting so that it produces worst de-raining results when countered with real-world rain images from SPA-Data. **It is worth noting that none of images from SPA-Data (training set or testing set) participates in the network training.** Due to the light rainy images contributing relatively smaller to total loss compared to medium and heavy rainy images, simply mixing the four datasets will cause the optimization algorithm to ignore light rain streaks (Rain100L) and put over-emphasis on medium and heavy rain streaks. Besides, since individual synthetic dataset provides an approximation about local distribution of real-world rain streaks, either such under-fitting or over-fitting on single synthetic dataset hinders generalization of de-raining model. The unbalanced allocation, which is determined by the capacity and rain streak properties of individual dataset, results in undesirable performance on real-world SPA-Data. As shown in Table 1, the de-raining model equipped our Neural Reorganization scheme achieves relatively homogeneous performance across multiple synthetic

datasets consistently and gain the state-of-the-art generalization ability over real-world dataset SPA-Data.

Qualitative comparison. Figure 4 presents visual comparisons of de-raining results on SPA-Data [35] over PreNet trained by 7 different strategies, which consist of three aforementioned ones and four individual dataset training approaches. Since Rain100L and Rain14000 do not contain heavy rain streaks, PreNet trained on individual Rain100L or Rain14000 fails to remove rain streaks clearly (the first and fourth rows in Figure 4). Similarly, PreNet trained on Rain100H or Rain800 is able to remove heavy rain-streaks while blurring the images (the second and third rows in Figure 4). Furthermore, due to the catastrophic forgetting, PreNet trained on Rain100H→Rain800→Rain14000→Rain100L sequence and Mix strategy can not remove heavy streaks (the first and fourth rows in Figure 4), or maintain detailed background information (the second and third rows in Figure 4). For PreNet trained with our NR (the last second row), it can be seen clearly that our NR obtain the most visually pleasing results. The same effect of our NR can also be found in Figure 5, which shows de-raining results on the real-world Real-Internet dataset [35].

4.4. Ablation Studies

De-raining on single synthetic dataset. Taking the promising PreNet, we perform extensive experiments to reveal the fatal shortcomings of existing image de-raining methods, which mainly focus on design sophisticated network architectures to obtain the-state-of-art image de-raining performance on single synthetic dataset. Table 2 gives a full description about the performance of PreNet trained on single synthetic dataset. With the fact that the highest value of PSNR and SSIM happened on the diagonal of the Table 2, we conclude that PreNet trained on single synthetic dataset tends to overfit on the biased distribution provided by this dataset, rather than acquire more general image de-raining knowledge. In contrast, PreNet equipped with our Neural Reorganization scheme can be competent across multiple synthetic datasets and then achieve state-of-the-art results on SPA-Data dataset [35].

Effect of each module. The Synaptic Consolidation (SC) module imitates the synaptic consolidation mechanism in the biological brains, and the synapses consolidated can preserve representative rain pattern. The Knowledge Distillation (KD) module emulates "cautious" learning style in biological brains, which enables synapses consolidated during different training phases to cooperate well. Therefore, both SC and KD can promote the generalization ability. Below we show the ablation experiments of PreNet with different modules to validate their functions. It is clear that promotion can be gained with either module and our NR achieves best performance.

Setting	Without Both	Without SC	Without KD	our NR
PSNR	34.91	35.20	35.14	35.60
SSIM	0.951	0.957	0.957	0.959

Table 3: Generalization of PreNet with different setups.

Promoting generalization progressively. We also conduct experiments to demonstrate that PreNet can promote its generalization progressively during the training on sequence Rain100H→Rain800→Rain14000→Rain100L. Specifically, after one training phase, we evaluate the model on SPA-Data and the result is reported in Table 4. Obviously, through sequentially training with NR, model generalization is promoted progressively.

Training phase	Rain100H	Rain800	Rain14000	Rain100L
PSNR	34.50	34.75	35.21	35.60
SSIM	0.952	0.957	0.957	0.959

Table 4: Generalization of different training phases.

4.5. Comparison with Transfer Learning Method

In this section, we conduct experimental comparison between the representative transfer learning method Syn2Real [43] and our proposed Neural Reorganization. The difference between the them is whether real-world images are utilized in the training stage. In particular, it should be emphasized that one of key advantages of our proposed scheme is not requiring the expensively and laboriously collected real-world samples in the training stage. For fairness, we follow the same setting as reported in the original paper [43], except replacing the labeled synthetic data of Syn2Real with Rain100H+Rain100L+Rain14000+Rain800 mixed datasets. And, we train the network with the same data sequence: Rain100H-Rain100L-Rain14000-Rain800. The corresponding results are shown in Table 5 on evaluation over real-world rainy dataset SPA-Data. As can be seen clearly, our NR can obtain better generalization performance than Syn2Real. It further testifies the effectiveness of our raised scheme.

Method	SPA-Data	
	PSNR	SSIM
Syn2Real	32.87	0.954
our NR	32.97	0.956

Table 5: Comparison of quantitative results between our NR and Syn2Real [43] in terms of PSNR and SSIM.

5. Conclusion

We present a brain-inspired Neural Reorganization scheme to endow the network with a subtle stability-plasticity trade-off for improving de-raining generalization.

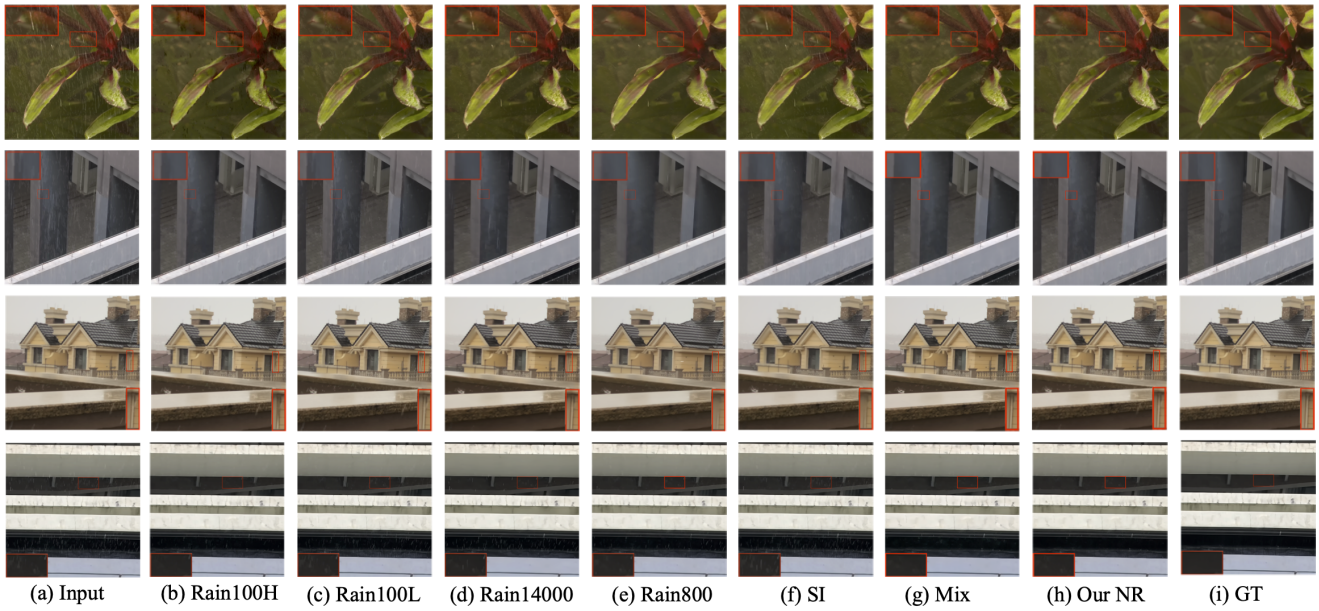

Figure 4: Visual quality comparisons of rain-streaks removal on rainy images from SPA-Data using de-raining model PreNet. (a) Input: rainy images from testing set of SPA-Data. (b-e): de-raining results based on PreNet trained on corresponding individual synthetic dataset. (f) SI: de-raining results using Prenet sequentially and independently trained on dataset sequence Rain100H-Rain100L-Rain14000-Rain800. (g) Mix: de-raining results using PreNet trained on Rain100H+Rain100L+Rain14000+Rain800 mixed dataset. (h) Our NR: de-raining results using PreNet trained on the same sequence as (f) with our Neural Reorganization. (i) GT: the clean image of (a).

Figure 5: Visual quality comparisons of rain-streaks removal on collected real-world rainy images from Internet for de-raining model PreNet equipped with different strategies.

We only use multiple synthetic datasets during training, while the performance on real-wolrd rainy images can be significantly improved. Since our method is orthogonal to existing de-raining methods, it can be easily inserted into them to improve their generalization abilities. Through extensive experiments, we demonstrate that the proposed method is able to process multiple synthetic datasets consistently and achieve state-of-the-art results over real-world scene.

Acknowledgement

This work was supported by the National Key R&D Program of China under Grand 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants U19B2038 and 61901433, the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-025, and the USTC Research Funds of the Double First-Class Initiative under Grant YD2100002003.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010. 5

[2] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003. 1

[3] M. Delange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2

[4] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. L. Wang, and M. Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 3

[5] Y. Du, J. Xu, Q. Qiu, X. Zhen, and L. Zhang. Variational image deraining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020. 1, 3

[6] Y. Du, J. Xu, X. Zhen, M. Cheng, and L. Shao. Conditional variational image deraining. *IEEE Transactions on Image Processing*, 29:6288–6301, 2020. 1, 3

[7] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 2

[8] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2, 5

[9] J. Grutzendler, N. Kasthuri, and WB Gan. Long-term dendritic spine stability in the adult cortex. *Nature*, 420(6917):812–816, 2002. 4

[10] R. Hadsell, D. Rao, AA. Rusu, and R. Pascanu. Embracing change: Continual learning in deep neural networks. *Trends in Cognitive Sciences*, 24(12):1028–1040, 2020. 2

[11] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 5

[12] Q. Huynh-Thu and M. Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008. 5

[13] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 3

[14] X. Jin, Z. Chen, J. Lin, Z. Chen, and W. Zhou. Unsupervised single image deraining with self-supervised constraints. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2761–2765, 2019. 3

[15] O. L. Junior, D. Delgado, V. Goncalves, and U. Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6, 2009. 1

[16] L. Kang, C. Lin, and Y. Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742–1755, 2012. 1

[17] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin. Non-locally enhanced encoder-decoder network for single image de-raining. *acm multimedia*, 2018. 2

[18] R. Li, L.-F. Cheong, and R. T. Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1

[19] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1

[20] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3833–3842, 2019. 2

[21] S. Li, W. Ren, F. Wang, I. B. Araujo, and X. Cao. A comprehensive benchmark analysis of single image deraining: Current challenges and future perspectives. *International Journal of Computer Vision*, (2–3), 2021. 2

[22] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 2

[23] Z. Li and D. Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. 5

[24] H. Lin, Y. Li, X. Fu, X. Ding, Y. Huang, and J. Paisley. Rain o'er me: Synthesizing real rain to derain with data distillation. *IEEE Transactions on Image Processing*, 29:7668–7680, 2020. 2, 3

[25] D. Liu, H. Zhang, F. Wu, and Z.J. Zha. Learning to assemble neural module tree networks for visual grounding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4673–4682, 2019. 1

[26] J. Liu, Z.J. Zha, X. Chen, Z. Wang, and Y. Zhang. Dense 3d-convolutional neural network for person re-identification in videos. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(1s):1–19, 2019. 1

[27] GI. Parisi, R. Kemker, JL. Part, C. Kanan, and S. Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. 2

[28] W. Ran, Y. Yang, and H. Lu. Single image rain removal boosting via directional gradient. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020. 1

[29] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng. Progressive image deraining networks: A better and simpler baseline. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3932–3941, 2019. 1, 2, 5

[30] G. Schaefer and M. Stich. Ucid: An uncompressed color image database. volume 5307, pages 472–480, 01 2004. 5

[31] M. S. Shehata, J. Cai, W. M. Badawy, T. W. Burr, M. S. Pervez, R. J. Johannesson, and A. Radmanesh. Video-based automatic incident detection for smart roads: The outdoor environmental challenges regarding false alarms. *IEEE Transactions on Intelligent Transportation Systems*, 9(2):349–360, 2008. 1

[32] C. Wang, X. Xing, Y. Wu, Z. Su, and J. Chen. Dcsfn: Deep cross-scale fusion network for single image rain removal. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, page 1643–1651, New York, NY, USA, 2020. Association for Computing Machinery. 1, 3

[33] G. Wang, C. Sun, and A. Sowmya. Erl-net: Entangled representation learning for single image de-raining. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5643–5651, 2019. 1

[34] H. Wang, Q. Xie, Q. Zhao, and D. Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2, 3

[35] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, 2019. 1, 5, 7

[36] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu. Semi-supervised transfer learning for image rain removal. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3872–3881, 2019. 2, 3

[37] G. Yang, F. Pan, and WB Gan. Stably maintained dendritic spines are associated with lifelong memories. *Nature*, 462:920–4, 11 2009. 4, 6

[38] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 2, 5

[39] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020. 1

[40] Y. Yang and H. Lu. Single image deraining via recurrent hierarchy enhancement network. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1814–1822, 2019. 2

[41] Y. Yang, W. Ran, and H. Lu. Rddan: A residual dense dilated aggregated network for single image deraining. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020. 1

[42] R. Yasarla and V. M. Patel. Confidence measure guided single image de-raining. *IEEE Transactions on Image Processing*, 29:4544–4555, 2020. 1

[43] R. Yasarla, V. A. Sindagi, and V. M. Patel. Syn2real transfer learning for image deraining using gaussian processes. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2, 3, 5, 7

[44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[45] SW. Zamir, A. Arora, S. Khan, M. Hayat, FS. Khan, M. Yang, and L. Shao. Multi-stage progressive image restoration. *arXiv preprint arXiv:2102.02808*, 2021. 2

[46] Z.J. Zha, D. Liu, H. Zhang, Y. Zhang, and F. Wu. Context-aware visual policy network for fine-grained image captioning. *IEEE transactions on pattern analysis and machine intelligence*, 2019. 1

[47] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2

[48] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3943–3956, 2020. 2, 5

[49] Z. Zhang, Y. Shi, C. Yuan, B. Li, P. Wang, W. Hu, and Z.J. Zha. Object relational graph with teacher-recommended learning for video captioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13278–13288, 2020. 1

[50] H. Zhu, C. Wang, Y. Zhang, Z. Su, and G. Zhao. Physical model guided deep image deraining. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020. 1