Unfolding Taylor's Approximations for Image Restoration

Man Zhou **Xueyang Fu*** **Zeyu Xiao** **Gang Yang** **Aiping Liu** **Zhiwei Xiong**
University of Science and Technology of China
manman@mail.ustc.edu.cn xyfu@ustc.edu.cn

Abstract

Deep learning provides a new avenue for image restoration, which demands a delicate balance between fine-grained details and high-level contextualized information during recovering the latent clear image. In practice, however, existing methods empirically construct encapsulated end-to-end mapping networks without deepening into the rationality, and neglect the intrinsic prior knowledge of restoration task. To solve the above problems, inspired by *Taylor's Approximations*, we unfold Taylor's Formula to construct a novel framework for image restoration. We find the main part and the derivative part of Taylor's Approximations take the same effect as the two competing goals of high-level contextualized information and spatial details of image restoration respectively. Specifically, our framework consists of two steps, which are correspondingly responsible for the mapping and derivative functions. The former first learns the high-level contextualized information and the later combines it with the degraded input to progressively recover local high-order spatial details. Our proposed framework is orthogonal to existing methods and thus can be easily integrated with them for further improvement, and extensive experiments demonstrate the effectiveness and scalability of our proposed framework.

1 Introduction

Image restoration has long been an important task in computer vision, which demands a delicate balance between spatial fine-grained details and high-level contextualized information while recovering a latent clear image from a given degraded observation. It is a highly ill-posed issue as there exists infinite feasible results for single degraded image. Representative image restoration tasks include image deraining, image deblurring and multi-spectral image super-resolution.

Much research efforts have been devoted to solve the single image restoration problem, which can be categorized into two groups: traditional optimization methods (15; 38) and deep learning based methods. In detail, various natural image priors have been developed in traditional image restoration methods to regularize the solution space of the latent clear image, *e.g.*, low-rank prior (38; 44), dark channel prior (33; 34; 56), graph-based prior (28; 2), total variation regularization (5; 7; 1) and sparse image priors (29; 59; 30; 21). However, these priors need to be carefully designed and some methods are difficult to optimize, limiting their practical usage.

Convolutional neural networks (CNNs) have been applied in numerous high-level computer vision problems and obtained promising improvement in image restoration tasks over traditional methods (54; 65; 14; 26; 47; 25; 68; 69; 9; 39). However, most existing CNN-based image restoration methods empirically construct encapsulated end-to-end mapping networks without deepening into the rationality, and neglect the intrinsic properties of the degradation process. Specifically, it makes these methods unable to balance effectively the two competing goals of spatial fine-grained details

*Corresponding Author.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

and high-level contextualized information while recovering the latent clear images from the degraded inputs, limiting the model performance. Moreover, existing forward-pass mapping methods lack of sufficient interpretability, making it hard to analyze the effect of each module and preventing their further improvement.

To solve the aforementioned problems, in this paper, we first revisit the connection between Taylor's Approximation and image restoration task, and then propose to unfold Taylor's Formula as blueprints to construct a novel framework that enforces each process part to coincide with the intrinsic properties of image restoration task. We figure out that the main part and derivative part of Taylor's Approximation take the same effect as the above two competing goals of image restoration respectively. In this way, our approach deviates from existing methods that optimally balance the above goals in the overall recovery process and are independent from the degradation process. Specifically, we break down the overall recovery process into two manageable steps, corresponding to two specific operation steps that are responsible for the mapping and derivative functions respectively. The former first learns the high-level contextualized information and the later combines it with the degraded input to progressively recover local high-order spatial details. Moreover, our proposed framework is orthogonal to existing methods and thus can be easily integrated with them for further performance gain. The conducted extensive experiments demonstrate the effectiveness and scalability of our proposed framework over three popular image restoration tasks, image deraining, image deblurring and multi-spectral image super-resolution.

The contributions of this paper can be summarized as follows:

1) We introduce a new perspective for designing image restoration framework by unfolding the Taylor's Formula. To the best of our knowledge, this is the first effort to solve image restoration task inspired by Taylor's Approximations.
2) In contrast to existing image restoration methods that sub-optimally balance the competing goals of high-level contextualized information and local high-order spatial details of image restoration respectively in the overall recovery process, we decompose the recovery process into two manageable steps, corresponding to two specific steps that are responsible for the mapping and derivative functions respectively.
3) Our proposed framework (i.e., Deep Taylor approximation framework) is orthogonal to existing CNN methods. These methods can be easily and directly integrated with our framework for further improvement. Extensive experiments conducted on standard benchmarks demonstrate the effectiveness of the framework.

2 Related Work

Image restoration aims to recover the latent clean images from its degraded observation, which is beneficial for several downstream high-level computer vision tasks, *e.g.*, object detection, image classification and image segmentation. It is a highly ill-posed problem as there exists infinite feasible results for a single degraded image. To this end, it attracts more attention over the whole community. Representative tasks of different degradation include image deraining, image deblurring.

Image deraining aims to remove the rain streaks while maintaining its image textures, and has gained much advances owing to the breakthrough of deep learning (11; 62; 63; 12; 50; 27; 28; 37; 49; 8; 70). Fu *et al.* (10) early attempt to design a CNN-based architecture for image deraining and obtain favorable performance with a larger margin than classic promising methods such as guided filter, nonlocal means filtering and more (52; 3; 22; 36). Yang *et al.* (58) introduce a multi-task network with a series of contextualized dilated convolution and recurrent framework to achieve joint detection and removal of rain streaks. Ren *et al.* (37) present the PreNet as a simple baseline for image deraining that combines Resblocks and recurrent layers in a progressive and recursive manner. Overall, the state-of-the-art deep learning methods rely on designing complex network structure. However, these complicated network architectures make them lack of evident interpretability and still have room for further improvement.

Image deblurring is a typical ill-posed problem which aims at generating a sharp latent image from a blurry observation. Early Bayesian-based iterative deblurring methods include the Wiener filter (51) and the Richardson-Lucy algorithm (40). Later works commonly rely on developing effective image priors (24; 42; 53; 71) or sophisticated data terms (6). Recently, several CNN-based methods have been proposed for image deblurring (16; 25; 32; 65; 26; 47; 43; 14; 66; 35; 64; 45; 39). For example,

Sun *et al.* (46) propose a CNN-based model to estimate a kernel and remove non-uniform motion blur. Chakrabarti (4) uses a network to compute estimations of sharp images that are blurred by an unknown motion kernel. Nah *et al.* (32) propose a multi-scale loss function to apply a coarse-to-fine strategy. Kupyn *et al.* propose DeblurGAN (25) and DeblurGAN-v2 (26) to remove blur based on adversarial learning. Despite of the encouraging performance achieved by CNN-based methods for image deblurring, they fail to reconstruct sharp results with good interpretability.

The Taylor's Approximations is one of the most popular methods in numerical approximation. In previous researches, Tukey (48) is the first to exploit the Taylor's series expansion to study error propagation. Afterwards, a number of studies (17; 19; 20; 23; 18) have been made and based on the first-order Taylor's expansion. It is, after all, a liner approximation that results in the inaccurate values. Xue *et al.* (55) develop above methods by high-order Taylor series expansion, which further improve the performance . However, it has never been explored for image restoration. In this paper, we revisit the connection between the Taylor's Approximations and image restoration process for the first time, and propose to unfold the Taylor's Formula to construct a novel framework, named Deep Taylor's Approximations Framework.

3 Proposed Method

3.1 Revisiting Taylor's Approximations and Image Restoration

Problem Formulation Image degradation is the most common and inevitable phenomena in imaging systems. It is commonly formulated as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{y}$, $\boldsymbol{x}$, $\boldsymbol{A}$ and $\boldsymbol{N}$ denote a degraded observation, a latent clear image, a degraded matrix and noise respectively. Specifically, when $\boldsymbol{A}$ is the identity matrix and $\boldsymbol{N}$ is the rain streak, it transforms as image deraining task. Moreover, when $\boldsymbol{A}$ is the blurry matrix and $\boldsymbol{N}$ is the additional noise, it turns to image deblurring task. In terms of multi-spectral image super-resolution, the A indicates the down-sampling and blurring operation while N is the additional noise. Observing Equation (1), for the inverse process, it can be reformulated as

$$\boldsymbol{y}_0 = \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{y} - \boldsymbol{N}), \tag{2}$$

where we denote the term $\boldsymbol{A}\boldsymbol{x}$ as $\boldsymbol{y}_0$, representing the latent image without additional noise. In detail, $\boldsymbol{N}$ is rain streak R over image deraining task while acting as the additional noise in image deblurring and multi-spectral image super-resolution.

Existing Methods Most of the existing CNN-based image restoration methods empirically construct encapsulated end-to-end mapping networks which neglect the intrinsic prior knowledge of image restoration task. Specifically, it acts as

$$\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{y}), \tag{3}$$

where $\boldsymbol{F}$ is the existing-designed mapping network, which empirically learns the mapping between the degraded and clean pairs $\boldsymbol{x}$ and $\boldsymbol{y}$ in the overall recovery process.

Our Motivation: Revisiting Image Restoration Task Different from most existing CNN-based methods, we try to associate Taylor's Approximations with image restoration process, which has not been considered in previous methods. Referring to Equation (2), we define the inverse process using the function $\boldsymbol{F}$, and then we can obtain the relationship as

$$\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{y}_0) = \boldsymbol{F}(\boldsymbol{y} - \boldsymbol{N}). \tag{4}$$

It can be figured out that the mapping function $\boldsymbol{F}$ in our framework corresponds the inverse operation of degradation matrix $\boldsymbol{A}$.

Different from most existing CNN-based methods, our framework considers the intrinsic prior knowledge of image restoration task and has better interpretability. The inverse operation of degradation matrix $\boldsymbol{A}$, indicating the generalized inverse mathematically, is denoted as

$$\boldsymbol{F} = \boldsymbol{A}^{-}. \tag{5}$$

Let $-\boldsymbol{N} = \boldsymbol{\epsilon} = \boldsymbol{y}_0 - \boldsymbol{y}$, we expand Equation (4) with an infinite-order Taylor's series expansion of the variable $\boldsymbol{y}$, written compactly as

$$\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{y}_0) = \boldsymbol{F}(\boldsymbol{y} + \boldsymbol{\epsilon}) \tag{6}$$

$$= \boldsymbol{F}(\boldsymbol{y}) + \frac{1}{1!}\frac{\partial \boldsymbol{F}(\boldsymbol{y})}{\partial y}\boldsymbol{\epsilon} + \frac{1}{2!}\frac{\partial^2 \boldsymbol{F}(\boldsymbol{y})}{\partial^2 y}\boldsymbol{\epsilon}^2 + \ldots + \frac{1}{k!}\frac{\partial^k \boldsymbol{F}(\boldsymbol{y})}{\partial^k y}\boldsymbol{\epsilon}^k + \ldots \tag{7}$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\partial^k \boldsymbol{F}(\boldsymbol{y})}{\partial^k y}\boldsymbol{\epsilon}^k, \tag{8}$$

When only regarding n order Taylor's Approximations, it can be simplified as

$$\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{y}) + \sum_{k=1}^{n} \frac{\partial^k \boldsymbol{F}(\boldsymbol{y})}{\partial^k y}\boldsymbol{\epsilon}^k. \tag{9}$$

It can be separated into two parts for consideration. In detail, the first term, defining as main part, $\boldsymbol{F}(\boldsymbol{y})$ represents the high-level contextualized information, which gives a constant approximation of clear image (ground truth) while the rest is the local high-order spatial details. When we revisit the connection between the goals of image restoration and Equation (9), we can clearly find that the main part and derivative part of Taylor's Approximations take the same effect as the two competing goals of high-level contextualized information and spatial details of image restoration respectively. In this way, our approach deviates from existing methods that optimally balance the above goals in the overall recovery process and are independent from the degradation process. Specifically, we break down the overall recovery process into two manageable steps. The former first learns the high-level contextualized structures and the later combines them with the degraded input to progressively recover local high-order spatial details. The details of our proposed framework are illustrated as bellow.

3.2 Deep Taylor's Approximations Framework

Based on above analysis, we reasonably associate Taylor's Approximations with image restoration task, which has not been considered in previous restoration methods. In detail, we propose to unfold Taylor's Formula as blueprints to construct a novel framework, named Deep Taylor's Approximations Framework. In this way, almost every process part one-to-one corresponds to each operation involved in Taylor's Formula, thereby breaking down the overall recovery process into two manageable steps. Specifically, our framework consists of two operation parts, which are correspondingly responsible for the mapping and derivative functions.

Mapping Function Part As aforementioned function $\boldsymbol{F}$, it takes responsibility for mapping the degraded input to the approximation of expected latent clear image. As shown in Figure 1, it can be implemented with existing CNN-Based or traditional methods. In the following, we choose several representative image restoration methods as $\boldsymbol{F}$ to validate the effectiveness and scalability.

Derivative Function Part Recalling Equation (9), for the k order derivative part, it can be written as

$$\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k = \frac{\partial^k \boldsymbol{F}(\boldsymbol{y})}{\partial^k y}\boldsymbol{\epsilon}^k, \tag{10}$$

differentiating above k order part $\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k$ for y as

$$\frac{\partial \boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k}{\partial \boldsymbol{y}} = \boldsymbol{F}^{(k+1)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k - \boldsymbol{F}^{(k)}(\boldsymbol{y}) \times k(\boldsymbol{\epsilon})^{k-1}, \tag{11}$$

multiplying above Equation (11) by ϵ as

$$\frac{\partial \boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k}{\partial \boldsymbol{y}} \times \boldsymbol{\epsilon} = (\boldsymbol{F}^{(k+1)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k - k\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^{k-1}) \times \boldsymbol{\epsilon} \tag{12}$$

$$= \boldsymbol{F}^{(k+1)}(\boldsymbol{y})(\boldsymbol{\epsilon})^{k+1} - k\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k. \tag{13}$$

To this end, we exploit Derivative function sub-network, named $\boldsymbol{G}$ to take effect as above process. We denote the k order output of network $\boldsymbol{G}$ as $\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k$, simply recorded as g_{out}^k. Referring Equation (13), we can find out the connection between k order output and $k+1$ order one

$$g_{out}^{k+1} = \boldsymbol{G}(g_{out}^k) + k\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k. \tag{14}$$

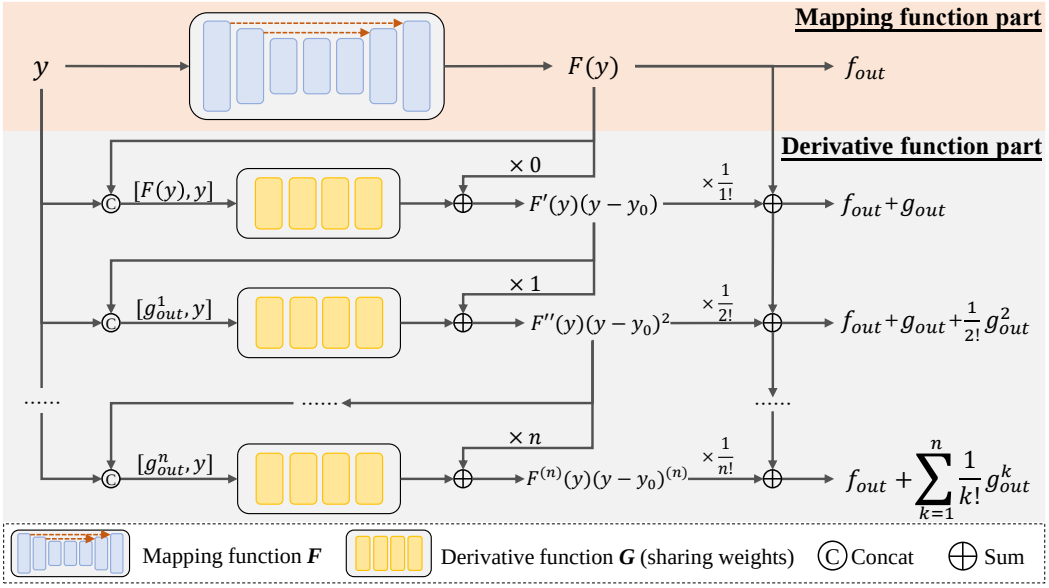

Figure 1: The overview structure of Deep Taylor's Approximations Framework. It consists of two parts, Mapping function part $\boldsymbol{F}$ and Derivative function part $\boldsymbol{G}$. The former learns to map the degraded input image y to the main energy of ground truth, while the later combines them with the degraded input to progressively recover the local high-order details. The parameters of Derivative function part $\boldsymbol{G}$ are shared across the progressive stages. The whole framework is trained in an end-to-end manner.

Replacing $\boldsymbol{F}^{(k)}(\boldsymbol{y})(\boldsymbol{\epsilon})^k$ with g^k_{out} as

$$g^{k+1}_{out} = \boldsymbol{G}(g^k_{out}) + k \cdot g^k_{out}. \tag{15}$$

Implementation Referring above analysis, we can find that the Mapping function part $\boldsymbol{F}$ only requires the degraded image as input as

$$f_{out} = \boldsymbol{F}(\boldsymbol{y}), \tag{16}$$

referring Equation (15), the Derivative function part $\boldsymbol{G}$ needs the g^k_{out} from the Mapping function sub-network $\boldsymbol{F}$ and blurry image y. It is because that the unfolding iteration process of $\boldsymbol{G}$ running involves y. In this regard, concatenating g^k_{out} and y into $\boldsymbol{G}$ as input for inference

$$g^{k+1}_{out} = \boldsymbol{G}(Concat([g^k_{out}, \boldsymbol{y}])). \tag{17}$$

Taking together above two operation steps, the final output of n order Deep Taylor's Approximations framework can be obtained as

$$O = f_{out} + \sum_{k=1}^{n} \frac{1}{k!} g^k_{out}. \tag{18}$$

Loss Function Following the setting in (13), we apply the $\mathcal{L}_1$ loss function to optimize the proposed framework. Suppose $\boldsymbol{x}$ denotes the corresponding ground truth of given degraded image $\boldsymbol{y}$, the final loss can be recorded as

$$L = \mathcal{L}_1(\boldsymbol{x} - O) + \lambda \mathcal{L}_1(\boldsymbol{x} - f_{out}), \tag{19}$$

where λ is a weighted factor.

4 Experiments

To demonstrate the effectiveness and scalability of our proposed framework, we conduct experiments over two image restoration tasks, *i.e.*, image deraining, image deblurring and multi-spectral image

super-resolution. In the following part, we first conduct the ablation experiments over several representative methods with/without integrating into our framework and report results on standard benchmarks. And then, we show the quantitative and qualitative results to explore our Deep Taylor's Approximations Framework with different series orders. In addition, we also testify the effect of our proposed framework by keeping the parameters consistent between the baselines and these equipped with our framework. Finally, we demonstrate the visual results to further validate the interpretability and the underlying mechanism.

4.1 Experimental Settings

For image deraining task, we choose four representative deraining methods, including DDN (11), PReNet (37), DCM (13) and MPRnet (60) to integrate with our proposed Deep Taylor's Approximations Framework for comparison between the integrated and the original. Specifically, all the aforementioned methods are exploited as mapping function part of our framework and then operate together with derivative function part for image deraining. Regarding experiment rainy datasets, the first three methods including DDN, DCM and PReNet are trained and evaluated over three widely-used standard benchmark datasets, including Rain100H, Rain100L, Rain800. In detail, Rain100H, a heavy rainy dataset are composed with 1,800 rainy images for training and 100 rainy samples for testing. However, referring the PReNet work, we find that 546 rainy images from the 1,800 training samples have the same background contents with testing images. Therefore, for fair comparison, we screen out these 546 rainy images from training set, and train all the selected baseline models on the remaining 1,254 training images. And, Rain100L dataset contains 200 training samples and 100 testing images with light level degraded rain streaks. In addition, the Rain800 is proposed in (61), which includes 700 training and 100 testing images. As for the recent method MPRnet, referring original paper, 13,712 clean-rain image pairs gathered from Rain100H, Rain100L, Test100, Test2800 and Test1200 are used for training and two above Rain100H, Rain100L for testing. For fair comparison, we follow the datasets setting as original paper. For the implementation, owing to unreleased open codes of DCM, we insert it into the training framework of PreNet for experiment comparison while the remaining follows the same pipeline.

For image deblurring task, typical methods like DCM (13), MSCNN (32) and RDN (68; 67) are adopted in our experiments. As in (64; 47; 26), we use the GoPro (32) dataset that contains 2,103 image pairs for training and 1,111 pairs for evaluation. Furthermore, to demonstrate generalizability, we take the model trained on the GoPro dataset and directly apply it on the test images of the HIDE (43) and RealBlur (41) datasets. The HIDE dataset is specifically collected for human-aware motion deblurring and its test set contains 2,025 images. While the GoPro and HIDE datasets are synthetically generated, the image pairs of RealBlur dataset are captured in real-world conditions. The RealBlur dataset has two subsets: (1) RealBlur-J is formed with the camera JPEG outputs, and (2) RealBlur-R is generated offline by applying white balance, demosaicking, and denoising operations to the RAW images.

For MS image super-resolution where A specifies the down-sampling and blurring operation and N is the additional noise, typical methods like PNN (31) and PANNET (57) are adopted in our experiments. Both of them fuse the PAN image and MS image to generate the latent pan-sharpened MS image. In our work, three kinds of satellite datasets, recorded as GaoFen2, WorldView-II and WorldView-III are employed for experimental verification. The paired training samples are unavailable in practice. When we construct the training set, the Wald protocol is employed to generate the paired samples. For example, given the MS image $H \in R^{M\times N\times C}$ and the PAN image $P \in R^{rM\times rN\times b}$, both of them are downsampled with ratio r, and the downsampled versions are denoted by $L \in R^{M/r\times N/r\times C}$ and $p \in R^{M\times N\times b}$. In the training set, L and p are regarded as the inputs, while H is the ground truth. For each database, we take hundreds of image pairs, and images are divided into two parts for training and test. In the training set, the PAN images are cropped into patches with the size of 128×128 pixels, and the corresponding MS patches are with the size of 32×32 pixels.

For quantitative comparisons, two popular measurement metrics are employed in the following quantitative comparison, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

4.2 Integrating Existing Methods into Our Framework

We evaluate the integrated models against their baselines (*i.e.*, the models trained without integration) in terms of PSNR/SSIM with 3-order (see Section 4.3). As shown in Table 1 and Table 2, we can

clearly find that, by integrating with our proposed framework, all the baselines obtain performance gain over all the datasets in image deraining, image deblurring task and MS image super-resolution, which validates the effectiveness of our framework. For example, in Table 2, DCM (13) obtains 0.36dB and 0.14dB PSNR gain on GoPro and RealBlur dataset. Taking two representative examples in Figure 3, the result of "Integrated" is cleaner than that of "Original" with fewer blurry effect. Meanwhile, for image deraining, the results of "Original" maintain the better spatial details than that of "Integrated". The quantitative comparison from Table 1 also testifies above analysis.

Table 1: Comparison of quantitative results in terms of PSNR (dB) and SSIM over image deraining. "Original" represents to employ the same architecture as original paper, and "Integrated" indicates to integrate them with our proposed framework. The corresponding experiment setting can be referred as section 4.2.

Model	Methods	Rain100H		Rain100L		Rain800	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DDN (11)	Original	22.26	0.690	34.85	0.951	24.04	0.867
	Integrated	22.34	0.701	35.16	0.953	24.21	0.867
PReNet (37)	Original	26.77	0.858	32.44	0.950	22.03	0.720
	Integrated	26.78	0.858	34.44	0.950	22.04	0.723
DCM (13)	Original	28.66	0.889	37.15	0.980	26.78	0.859
	Integrated	28.77	0.901	37.31	0.981	26.93	0.861
MPRnet (60)	Original	30.44	0.872	36.24	0.959	-	-
	Integrated	30.52	0.873	36.30	0.960	-	-

Table 2: Comparison of quantitative results in terms of PSNR (dB) and SSIM over image deblurring. All the models follow the same settings as original works.

Model	Methods	GoPro		HIDE		RealBlur-J		RealBlur-R	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCM (13)	Original	30.25	0.929	28.00	0.901	35.55	0.927	28.60	0.871
	Integrated	30.39	0.933	28.10	0.926	35.91	0.930	28.69	0.873
MSCNN (32)	Original	29.08	0.914	25.72	0.874	32.51	0.841	27.87	0.827
	Integrated	29.19	0.916	25.80	0.877	32.79	0.852	27.99	0.830
RDN (68)	Original	29.20	0.929	26.44	0.859	28.38	0.899	26.50	0.871
	Integrated	29.31	0.934	26.50	0.866	28.44	0.901	26.56	0.880

Table 3: Comparison of quantitative results in terms of PSNR (dB) and SSIM over MS image super-resolution. All the models follow the same settings as original works.

Model	Methods	WV2		GF2		WV3	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PNN (31)	Original	40.7550	0.9624	43.1208	0.9704	29.9418	0.9121
	Integrated	40.8317	0.9764	43.2201	0.9806	30.122	0.9135
PANNET (57)	Original	40.8176	0.9626	43.0659	0.9685	29.684	0.9072
	Integrated	40.9225	0.9631	43.1127	0.9685	29.712	0.9085

4.3 Deep Taylor's Approximations Framework with Different Orders

In this section, we perform the ablation studies about different orders of our proposed Deep Taylor's Approximations Framework. For simplicity, we take the representative method, *i.e.*, DCM, to validate the underlying mechanism over image deraining and image deblurring. In detail, DCM is used for mapping function part and taken together with different orders derivative function part from 0 to 6. The 0-order represents the original DCM (13) method. And the parameters of derivative function part are shared cross all the derivative stages[2].

Training Configuration. One NVIDIA GTX 2080Ti GPU is used for training. In the experiments, all the variant DCM networks share the same training setting. The patch size is 100×100, and the

[2]We use two convolutional layers in the experiments, and this two-layer structure can be readily replaced by other advanced embodiments for better performance.

Table 4: Comparison of quantitative results in terms of PSNR (dB) and SSIM about the different order Taylor's approximation results from 0 to 6 over image deraining.

Model	Methods	Rain100H		Rain100L		Rain800	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCM (13)	order-0	28.66	0.889	37.15	0.980	26.78	0.859
	order-1	28.67	0.889	37.17	0.980	26.78	0.859
	order-2	28.71	0.890	37.20	0.980	26.79	0.859
	order-3	28.77	0.901	37.31	0.981	26.93	0.861
	order-4	28.69	0.889	37.22	0.980	26.76	0.859
	order-5	28.73	0.889	37.24	0.981	26.83	0.860
	order-6	28.77	0.891	37.18	0.980	26.72	0.859

Table 5: Comparison of quantitative results in terms of PSNR (dB) and SSIM about the different order Taylor's approximation results from 0 to 6 over image deblurring.

Model	Methods	GoPro		HIDE		RealBlur-J		RealBlur-R	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCM (13)	order-0	30.32	0.921	28.03	0.913	35.83	0.918	28.62	0.862
	order-1	30.33	0.922	28.05	0.914	35.85	0.918	28.64	0.862
	order-2	30.36	0.928	28.07	0.920	35.88	0.925	28.66	0.867
	order-3	30.39	0.933	28.10	0.926	35.91	0.930	28.69	0.873
	order-4	30.35	0.932	28.06	0.925	35.87	0.929	28.65	0.872
	order-5	30.36	0.925	28.07	0.918	35.88	0.922	28.66	0.865
	order-6	30.36	0.932	28.08	0.924	35.89	0.929	28.66	0.871

batch size is 4. The ADAM algorithm is adopted to train the models with an initial learning rate 1×10^{-3}, and ends after 100 epochs. When reaching 30, 50 and 80 epochs, the learning rate is decayed by multiplying 0.2 and λ is set as 1.0 in loss function.

Experiment Analysis. As shown in Table 4 and Table 5, we can clearly find that, by integrating with our proposed framwork, all the baselines obtain performance gain in image deraining and image deblurring task over standard benchmarks, which validates the effectiveness of our framework. In addition, we also report the visual results in Figure 2 and Figure 3 for image deraining and deblurring. Owing to the 3-order performs the best, it is chose as the baseline framework for implementation. Taking the trade-off 3-order DCM for example, we evaluate it over image deraining task and analyze the different high-order output of derivative function part to testify the underlying mechanism.

4.4 Effect of the framework

Compared to the original baselines, the integrated version brings extra parameters. For fair comparison, to eliminate the influence of parameter, we implement the derivative part as refine module followed by the selective baselines to further investigate the effect of the framework. Taking DCM as an example, we conduct the experiments between (the DCM integrated with Taylor's Framework) and (the DCM concatenated by the derivative part in multiple order times directly) to keep the number of parameter consistent. Moreover, the derivative part is implemented by the two-layer convolution interpolated by Relu activation with/without residual learning. For simplification, we denote some annotations before our presentation:

DCM-C-R: DCM concatenated by the derivative part with residual learning;

DCM-C-no-R: DCM concatenated by the derivative part without residual learning;

DCM-T-no-R: DCM with Taylor's derivative part without residual learning;

DCM-T-R: DCM with Taylor's derivative part with residual learning.

1) without residual learning. To be specific, we adopt the 3-order for evaluation. In terms of PSNR, it can be seen in Table 6 that the former outperforms the latter by 1.57 dB and 1.21 dB over Rain100L and Rain100H, respectively. The corresponding results clearly testify that the performance gain is from the effectiveness of our proposed Taylor's Framework.

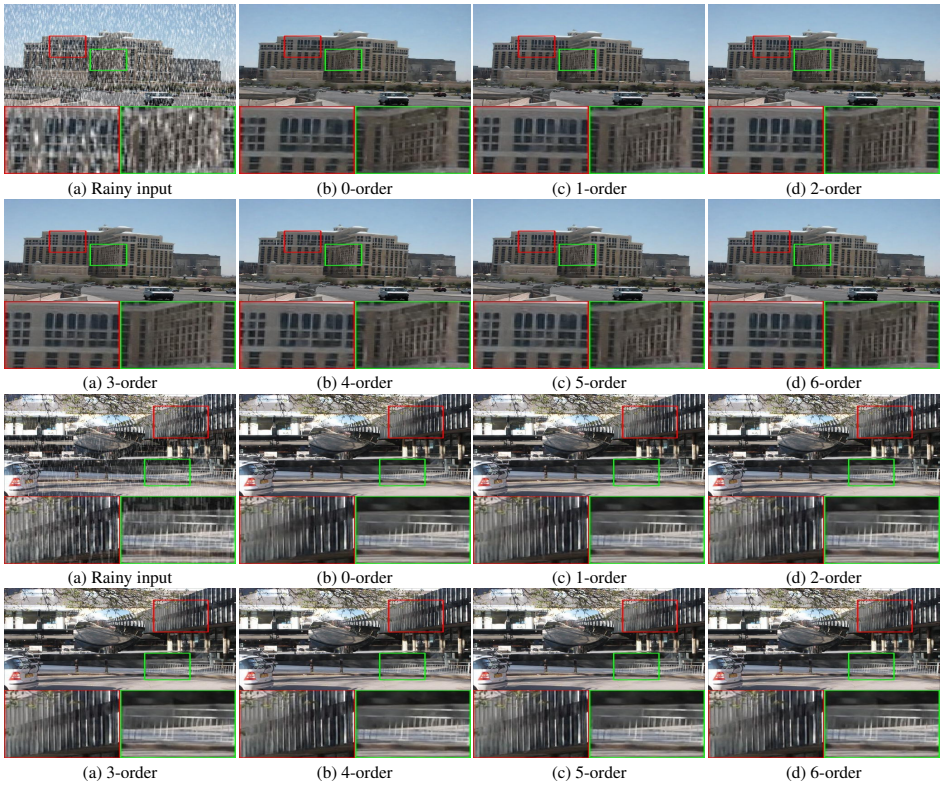

Figure 2: Visual results from the Rain800 rainy dataset for DCM (13) with different-order framework.

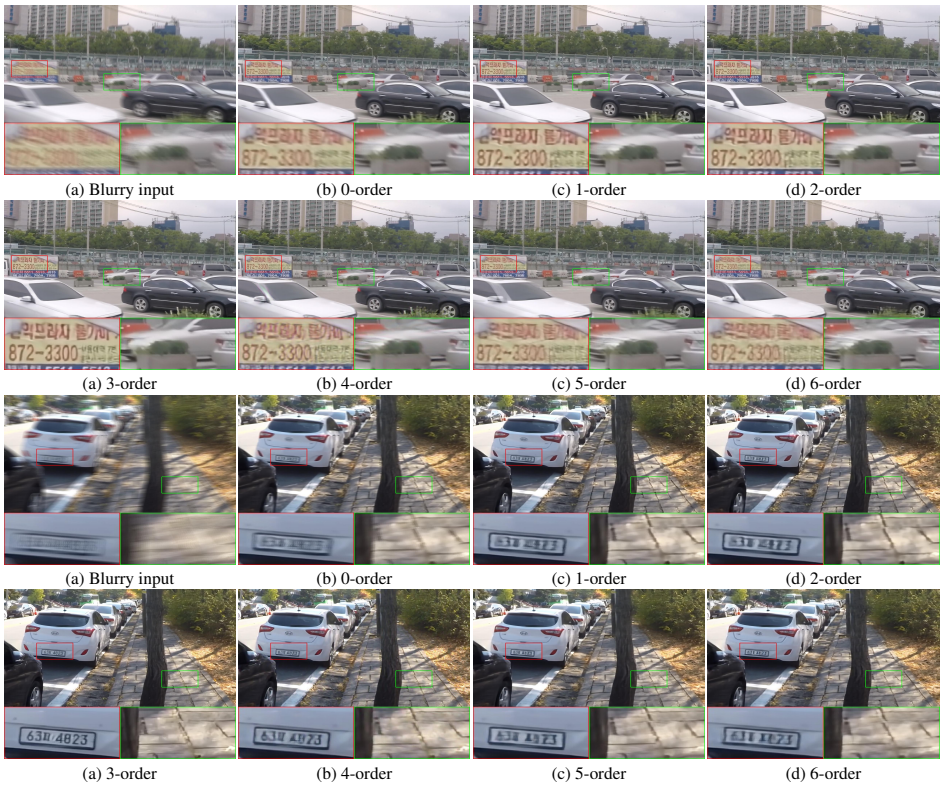

Figure 3: Visual results from the GoPro blurry dataset for DCM (13) with different-order framework.

Table 6: The effect of the proposed framework in terms of PSNR (dB) and SSIM over image de-raining. All the models follow the same settings as original works.

Model	Rain100L		Rain100H	
	PSNR	SSIM	PSNR	SSIM
DCM-C-no-R	35.74	0.975	27.56	0.868
DCM-T-no-R	37.31	0.980	28.77	0.901

2) with residual learning. Residual learning aims to learn the identity mapping and avoid the gradient vanish/explosion issue. Different from it, our proposed framework is customized for the specific image restoration task by revisiting the similarity between the Taylor's Approximation and degradation model. Furthermore, as suggested, we conduct the experiments between residual learning and our proposed framework. It can be seen in Table 7 that the results of the DCM-T-no-R also surpass that of DCM-C-R, and combining the residual learning with our Taylor's framework brings additional improvement. This fact further testifies the effectiveness of our proposed framework.

Table 7: The effect of the proposed framework in terms of PSNR (dB) and SSIM over image de-raining. All the models follow the same settings as original works.

Model	Rain100L		Rain100H	
	PSNR	SSIM	PSNR	SSIM
DCM-C-R	37.24	0.980	28.71	0.901
DCM-T-no-R	37.31	0.980	28.77	0.901
DCM-T-R	30.36	0.980	28.83	0.901

4.5 Limitations and Discussions

First, we evaluate the effectiveness of proposed framework over three typical image restoration tasks (*i.e.*, image deraining, image deblurring and MS image super-resolution) and we will conduct more comprehensive experiments on restoration tasks (*e.g.*, low-light image enhancement, image denoising and image dehazing). Second, the focus of this work is not designing a new image restoration network, so both the mapping function and derivative function steps can be readily replaced by other advanced embodiments for better performance. since we the mapping function part is implemented only by several convolution layers in our work. It should be constructed by various complex neural architectures for comparison in the future. In addition, the parameters of derivative step are shared across different orders. We will explore the situation when the parameters are not sharing weights.

5 Conclusion

In this paper, we propose a novel framework, Deep Taylor's Approximations Framework for image restoration by unfolding Taylor's Formula. Our proposed framework is orthogonal to existing deep learning-based image restoration methods and thus can be easily integrated with these methods for further improvement. Extensive evaluations demonstrate the effectiveness and interpretability of our framework in image restoration task, *i.e.*, image deraining, MS image super-resolution and image deblurring. We believe the Deep Taylor's Approximations Framework also has potential to advance other image/video restoration tasks, *e.g.*, image/video enhancement and image/video denoising.

Acknowledgments and Disclosure of Funding

We would like to thank all the reviewers for their insightful comments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61901433, and in part by the University of Science and Technology of China (USTC) Research Funds of the Double First-Class Initiative under Grant YD2100002003 and YD2100002004.

Broader Impact

Since image degradation is the most common and inevitable phenomena in imaging systems, *e.g.*, from the point spread function of the optical system to the shaking during shooting, the image restoration technology has broad impacts and practical values in various applications. Related fields include remote sensing, medicine, astronomy, military, and civilian imaging equipment. Image restoration technology aims to recover high-quality images from given low-quality counterparts. In daily life, it can help people who cannot afford professional cameras to take photos with low-end devices. Therefore, our image restoration method based on the proposed unfolding Taylor's approximations can provide high-quality clear images to facilitate intelligent data analysis tasks in these fields.

The negative consequences may accompany image restoration technology despite the many benefits it brings. This is mainly associated with certain risks of privacy and consumer experience. For example, in media or criminal cases, the identity of certain persons will be blurred to protect privacy. In this case, image deblurring technology may reveal the personal identity, thereby compromising their privacy. In addition, some users may beautify the image before sharing it. Therefore, the use of image restoration technologies to restore the posted images may arouse users' antipathy. Furthermore, it is important to be cautious of the results of any image restoration algorithms as failures, leading to misjudgments and affecting subsequent use.

References

[1] Babacan, S.D., Molina, R., Katsaggelos, A.K.: Variational bayesian blind deconvolution using a total variation prior. IEEE Transactions on Image Processing **18**(1), 12–26 (2009)

[2] Bai, Y., Cheung, G., Liu, X., Gao, W.: Graph-based blind image deblurring from a single photograph. IEEE Transactions on Image Processing **28**(3), 1404–1418 (2019). https://doi.org/10.1109/TIP.2018.2874290

[3] Bossu, J., Hautiere, N., Tarel, J.P.: Rain or snow detection in image sequences through use of a histogram of orientation of streaks. International journal of computer vision **93**(3), 348–367 (2011)

[4] Chakrabarti, A.: A neural approach to blind motion deblurring. In: European Conference on Computer Vision (2016)

[5] Chan, T., Wong, C.K.: Total variation blind deconvolution. IEEE Transactions on Image Processing **7**(3), 370–375 (1998). https://doi.org/10.1109/83.661187

[6] Dong, J., Pan, J., Sun, D., Su, Z., Yang, M.H.: Learning data terms for non-blind deblurring. In: European Conference on Computer Vision (2018)

[7] Du, S., Liu, Y., Ye, M., Xu, Z., Li, J., Liu, J.: Single image deraining via decorrelating the rain streaks and background scene in gradient domain. Pattern Recognition **79**, 303–317 (2018)

[8] Du, Y., Xu, J., Qiu, Q., Zhen, X., Zhang, L.: Variational image deraining. In: WACV (2020)

[9] Fan, Y., Yu, J., Mei, Y., Zhang, Y., Fu, Y., Liu, D., Huang, T.S.: Neural sparse representation for image restoration. arXiv preprint arXiv:2006.04357 (2020)

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

[11] Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

[12] Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.: Lightweight pyramid networks for image deraining. IEEE transactions on neural networks and learning systems **31**(6), 1794–1807 (2019)

[13] Fu, X., Qi, Q., Zhu, Y., Ding, X., Zha, Z.J.: Rain streak removal via dual graph convolutional network. In: AAAI (2021)

[14] Gao, H., Tao, X., Shen, X., Jia, J.: Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)

[15] Gong, D., Tan, M., Zhang, Y., van den Hengel, A., Shi, Q.: Blind image deconvolution by automatic gradient activation. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)

[16] Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I., Shen, C., Van Den Hengel, A., Shi, Q.: From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

[17] Goodchild, M.F.: A general framework for error analysis in measurement-based gis. Journal of Geographical Systems **6**(4), 323–324 (2004)

[18] Helton, J.C., Davis, F.J.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety **81**(1), 23–69 (2003)

[19] Heuvelink, G.: Error propagation in environmental modelling with gis. University of Liverpool (1998)

[20] Heuvelink, G., Stein, P.: Propagation of errors in spatial modelling with gis. International journal of geographical information systems (1989)

[21] Kang, L.W., Lin, C.W., Fu, Y.H.: Automatic single-image-based rain streaks removal via image decomposition. IEEE Transactions on Image Processing **21**(4), 1742–1755 (2012). https://doi.org/10.1109/TIP.2011.2179057

[22] Kim, J.H., Lee, C., Sim, J.Y., Kim, C.S.: Single-image deraining using an adaptive nonlocal means filter. In: ICIP (2013)

[23] Kobayashi, T., Miller, H., Othman, W.: Analytical methods for error propagation in planar space–time prisms. Journal of Geographical Systems **13**(4), 327–354 (2011)

[24] Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: Neural Information Processing Systems (2009)

[25] Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: DeblurGAN: Blind motion deblurring using conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

[26] Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In: IEEE International Conference on Computer Vision (2019)

[27] Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)

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

[29] Li, S., Ren, W., Wang, F., Araujo, I.B., Tokuda, E.K., Junior, R.H., Cesar-Jr, R.M., Wang, Z., Cao, X.: A comprehensive benchmark analysis of single image deraining: Current challenges and future perspectives. International Journal of Computer Vision **129**(4), 1301–1322 (2021)

[30] Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 3397–3405 (2015). https://doi.org/10.1109/ICCV.2015.388

[31] Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G.: Pansharpening by convolutional neural networks. Remote Sensing **8**(7), 594 (2016)

[32] Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

[33] Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)

[34] Pan, J., Sun, D., Pfister, H., Yang, M.H.: Deblurring images via dark channel prior. IEEE transactions on pattern analysis and machine intelligence **40**(10), 2315–2328 (2017)

[35] Park, D., Kang, D.U., Kim, J., Chun, S.Y.: Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In: European Conference on Computer Vision (2020)

[36] Park, W.J., Lee, K.H.: Rain removal using kalman filter in video. In: ICSMA (2008)

[37] Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)

[38] Ren, W., Cao, X., Pan, J., Guo, X., Zuo, W., Yang, M.H.: Image deblurring via enhanced low-rank prior. IEEE Transactions on Image Processing **25**(7), 3426–3437 (2016). https://doi.org/10.1109/TIP.2016.2571062

[39] Ren, W., Zhang, J., Pan, J., Liu, S., Ren, J., Du, J., Cao, X., Yang, M.H.: Deblurring dynamic scenes via spatially varying recurrent neural networks. IEEE transactions on pattern analysis and machine intelligence (2021)

[40] Richardson, W.H.: Bayesian-based iterative method of image restoration. JoSA **62**(1), 55–59 (1972)

[41] Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: European Conference on Computer Vision (2020)

[42] Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)

[43] Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., Shao, L.: Human-aware motion deblurring. In: IEEE International Conference on Computer Vision (2019)

[44] Shi, F., Cheng, J., Wang, L., Yap, P.T., Shen, D.: Lrtv: Mr image super-resolution with low-rank and total variation regularizations. IEEE transactions on medical imaging **34**(12), 2459–2466 (2015)

[45] Suin, M., Purohit, K., Rajagopalan, A.N.: Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: IEEE Conference on Computer Vision and Pattern Recognition (2020)

[46] Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

[47] Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

[48] Tukey, J.W.: The propagation of errors, fluctuations and tolerances, supplementary formulas (1957)

[49] Wang, C., Xing, X., Wu, Y., Su, Z., Chen, J.: Dcsfn: Deep cross-scale fusion network for single image rain removal. In: ACM MM (2020)

[50] Wang, H., Wu, Y., Li, M., Zhao, Q., Meng, D.: A survey on rain removal from video and single image. arXiv preprint arXiv:1909.08326 (2019)

[51] Weiner, N., Extrapolation, I.: Smoothing of stationary time series: With engineering applications (1949)

[52] Xu, J., Zhao, W., Liu, P., Tang, X.: Removing rain and snow in a single image using guided filter. In: CSAE (2012)

[53] Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: European Conference on Computer Vision (2010)

[54] Xu, L., Ren, J.S.J., Liu, C., Jia, J.: Deep convolutional neural network for image deconvolution. In: Neural Information Processing Systems (2014)

[55] Xue, J., Leung, Y., Ma, J.H.: High-order taylor series expansion methods for error propagation in geographic information systems. Journal of Geographical Systems **17**(2), 187–206 (2015)

[56] Yan, Y., Ren, W., Guo, Y., Wang, R., Cao, X.: Image deblurring via extreme channels prior. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)

[57] Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J.: Pannet: A deep network architecture for pan-sharpening. In: IEEE International Conference on Computer Vision. pp. 5449–5457 (2017)

[58] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

[59] Yang, W., Tan, R.T., Wang, S., Fang, Y., Liu, J.: Single image deraining: From model-based to data-driven and beyond. IEEE Transactions on pattern analysis and machine intelligence (2020)

[60] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (2021)

[61] Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology (2017)

[62] Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

[63] Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. TCSVT (2019)

[64] Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)

[65] Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R.W., Yang, M.H.: Dynamic scene deblurring using spatially variant recurrent neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

[66] Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.: Deblurring by realistic blurring. In: IEEE Conference on Computer Vision and Pattern Recognition (2020)

[67] Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. IEEE transactions on pattern analysis and machine intelligence (2019)

[68] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition (June 2018)

[69] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)

[70] Zhu, H., Wang, C., Zhang, Y., Su, Z., Zhao, G.: Physical model guided deep image deraining. In: ICME (2020)

[71] Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: IEEE International Conference on Computer Vision (2011)