Proposal for post-PhD Research

Dr. KHODJA AbdErraouf
abderraouf_khodja@zjnu.edu.cn

1 Summary of the Proposal

Image deraining has benefited a lot from the fast development of neural networks in the last few years. However, the requirement of large amounts of rainy-clean image pairs for supervision limits the wide use of these models. Although there have been a few attempts in training an image deraining model with only single rainy images, existing self-supervised deraining approaches suffer from inefficient network training, loss of useful information, or dependence on rain modeling.

In this proposal, we present our strategy to develop an effective method to train an image deraining model with only single rainy images. Firstly, briefly introduce the recent development in field of self supervised learning for single image deraining and other image restoration tasks. Motivated by our previous work on Rain2Rain learning [1]. We venture into developing a self supervised learning method for image deraining. Our proposal first introduces the problem that self supervised learning address in Section 1. In section ??, we then present various learning methodologies in the context of unsupervised and self supervised learning for image restoration that are related to our proposal method. Section 1 dives into the technical details of our proposal. Finaly, Section 1.1 discuss the current experimental results that validates our initial assumptions, challanges and future milestones of this research proposal.

Introduction

Image deraining is a low-level vision task that is fundamental in computer vision, since rain drops degrades the visual quality of collected images and may adversely affect subsequent image analysis and processing tasks, such as classification and semantic segmentation [18].

Traditional image deraining methods such as BM3D [7], NLM [4], and WNNM [11], use local or non-local structures of an input rainy image. These methods are non-learning-based without the need for ground-truth images. Recently, convolutional neural networks (CNNs) provide us with powerful tools for image deraining. Numerous CNN-based image derainers, e.g., DnCNN [34], U-Net [26], RED [20], MemNet [28], and SGN [10], have superior performance over traditional derainers. However, CNN-based derainers depend heavily on a large number of rainy-clean image pairs for training. Unfortunately, collecting large amounts of aligned pairwise rainy-clean training data is extremely challenging and expensive in real-world photography. Additionally, models trained with synthetic rainy-clean image pairs degrade greatly due to the domain gap between synthetic and real rain.

To mitigate this problem, a series of unsupervised and self-supervised methods that do not require any clean images for training are proposed. These methods require

- Training the network with multiple independent rainy observations per scene [17].
- Designing specific blind-spot network structures to learn self-supervised models on only single rainy images [13, 15, 30], and making further improvements by using rain models [15, 30].
- Training the network with noisier-rainy pairs, where the noisier image is derived from the rainy one with synthetic rain added [31, 22].

However, these requirements are not practical in real-world deraining scenarios. Firstly, capturing multiple rainy observations per scene remains very challenging, especially for motion scenarios. Secondly, the relatively low accuracy and heavy computational burden of blind-spot networks greatly limit the application. Moreover, self-supervised methods with rain model assumptions may work well in synthetic experiments when the rain distribution is known as a prior. However, these methods degrade sharply when dealing with real-world rainy images where the rain distribution remains unknown.

Image Deraining With Only Rainy Image

Image deraining methods using only rainy images can be categorized into two groups: traditional derainers and deep derainers. Traditional derainers include BM3D [7], NLM [4], and WNNM [11]. For deep derainers trained with only a single rainy image, Ulyanov et al. [29] proposed deep image prior (DIP), where the image prior is captured from the CNN network rather than specially designed; Self2Self [25] and rainy-as-Clean [31] are recent works. Lehtinen et al. [17] introduced rain2rain to train a deep derainer with multiple rainy observations of the same scenes. Subsequently, self-supervised deraining models, including rain2Void [13] and rain2Self [3], were proposed to train the networks only with one rainy observation per scene. Specifically, the carefully designed blind-spot1 networks are used to avoid learning the identity transformation. Recently, Probabilistic rain2Void [14], Laine19 [15], and Dilated Blind-Spot Network [30] further introduced explicit rain modeling and probabilistic inference for better performance. Masked convolution [15] and stacked dilated convolution layers [30] were introduced for faster training. Different from blind-spot-based self-supervised methods, in Noisier2rain [22], training pairs are prepared by generating synthetic rains from a rain model and adding them to single rainy images. However, the rain model is hard to specify, especially in real-world scenarios. rainy-as-Clean [31] mentioned above shares similar philosophy.

Additionally, Soltanayev and Chun [27] used Stein's unbiased risk estimator (SURE) to train AWGN deraining models on single rainy images, and Zhussip et al. [36] extended it to the case of correlated pairs of rainy images. Cha and Moon [5] used SURE to fine-tune a supervised derainer for each test rainy image. However, SURE-based algorithms are only designed for Gaussian additive rain, and the rain level is required to be known as a prior.

Goal and Objectives

In this work, we propose , a novel self-supervised image deraining framework that overcomes 2021 the limitations above. Our approach consists of a training image pairs generation strategy based on sub-sampling and a self-supervised training scheme with a regularization term. Specifically, training input and target are generated by random neighbor sub-samplers, where two sub-sampled paired images are extracted from a single rainy image with each element on the same position of the two images being neighbors in the original rainy image. In this way, if we assume that rain with each pixel is independent conditioned on its pixel value and there is no correlation between rain in different positions, then these two sub-sampled paired rainy images are independent given the ground-truth of the original rainy image. Accordingly, inspired by rain2rain [17], we use the above training pairs to train a deraining network. Besides, we develop a regularization term to address the essential difference of pixel ground-truth values between neighbors on the original rainy image. The proposed self-supervised framework aims at training deraining networks with only single images available, without any modifications to the network structure. Any network that performs well in supervised image deraining tasks can be used in our framework. Moreover, our method does not depend on any rain models either.

To evaluate the proposed Neighbor2Neighbor, a series of experiments on both synthetic and real-world rainy images are conducted. The extensive experiments show that our Neighbor2Neighbor outperforms traditional derainers and existing self-supervised deraining methods learned from only single rainy images. The results demonstrate the effectiveness and superiority of the proposed method.

The main contributions of our paper are as follows:

- We propose a novel self-supervised framework for image deraining, in which any existing deraining networks can be trained without any clean targets, network modifications, or rain model assumptions.
- From the theoretical perspective, we provide a sound motivation for the proposed framework.
- Our method performs very favorably against state-of-the-art self-supervised deraining methods especially on real-world datasets, which shows its potential applications in real-world scenarios.

Motivation

In Section 3, we describe the theoretical framework that motivates our proposed method in Section 4. The summary of Section 3 is as follows: In Section 3.1, we revisit the related theory proposed in rain2rain, which proves that paired rainy images taken from the same scene can also be used to train deraining models. Before we extend this theory to the case when only single rainy observation is available, we discuss the case of paired rainy images with slightly different ground-truths in Section 3.2, which is very useful for the extension to single rainy images. Then in Section 3.3, we mathematically formulate the underlying theory of training deraining networks where training pairs are generated using image pair samplers, and we further propose a regularizer to solve the problem caused by non-zero ϵ which is discussed in Section 3.2.

Downsampling

Downsampling is a basic operation for synthesizing low-resolution images from high resolution images. More generally, we consider both downsamping and upsampling, i.e., the resize operation. There are several resize algorithms nearest-neighbor interpolation, area resize, bilinear interpolation, and bicubic interpolation.

Different resize operations bring in different effects some produce blurry results while some may output over-sharp images with overshoot artifacts.

Rain2Rain Learning

rain2rain [17] is a deraining method trained without the need for ground-truth clean images. This method only requires pairs of independent rainy images of the same scene. Given two independent rainy observations named y and z of the same ground-truth image x, rain2rain tries to minimize the following loss in terms of θ,

$$argmin\theta E_{x,y,z}\|f_\theta(y) - z\|_2^2 \tag{1}$$

where f_θ is the deraining network parameterized by θ. Minimizing Equation (1) yields the same solution as the supervised training with the '2-loss. For detailed discussions, refer to Section 2 of [17] and Section 3.1 of [36].

Paired Images with Similar Ground Truths

rain2rain [17] mitigates the need of clean images. However, capturing multiple rainy observations of a scene remains a very challenging problem. The ground-truths of two rainy observations are difficult to be the same due to occlusion, motion, and lighting variation. Thus, we propose to extend the Equation (1) to the case where the gap between the underlying clean images $\epsilon := Ez|x(z) - Ey|x(y) \neq 0$.

Theorem 1 Let y and z be two independent rainy images conditioned on x, and assume that there exists an ϵ 6= 0 such that $Ey|x(y) = xandEz|x(z) = x + \epsilon$. Let the variance of z be $\sigma 2z$. Then it holds that

$$Ex, y\|f\theta(y) - x\|22 = Ex, y, z\|f\theta(y) - z\|22 - \sigma 2z + 2\epsilon Ex, y(f\theta(y) - x).(2) \tag{2}$$

The proof is given in the supplementary material. Theorem 1 states that when the gap ϵ, since $Ex, y(f\theta(y) - x)6 \equiv 0$, optimizing $Ex, y, z\|f\theta(y) - z\|_2^2$ does not yield the same solution as the supervised training loss $Ex, y\|f\theta(y) - x\|_2^2$.

Fortunately, if $\epsilon \to 0$, which means the gap is sufficiently small, $2\epsilon E_{x,y}(f\theta(y) - x) \to 0$, so the network trained with rainy image pair (y,z) works as a reasonable approximate solution to the supervised training network. Note that when $\epsilon = 0$, since $\sigma 2z$ is a constant, minimizing both sides of Equation (2) results in $argmin\theta Ex, y, z\|f\theta(y) - z\|_2^2$, which is the objective of rain2rain.

1.1 Extension to Single rainy Images

Inspired by rain2rain where training pairs are independent rainy image pairs of the same scene, we go a step further and propose to generate independent training pairs from single rainy images y by sampling. To be specific, an image pair sampler $G = (g1, g2)$ is used to generate a rainy image pair $(g1(y), g2(y))$ from a single rainy image y. The contents of two sampled images $(g1(y), g2(y))$ are closely resembled. Similar to Equation (1), we try to adopt the sampled image pair as two rainy observations, which becomes:

$$argmin\theta Ex, y\|f\theta(g1(y)) - g2(y)\|2.(3) \tag{3}$$

Different from rain2rain, the ground-truths of two sampled rainy images $(g1(y), g2(y))$ differ, i.e., $\epsilon = Ey|x(g2(y)) - Ey|x(g1(y))6 = 0$. According to Theorem 1, directly applying Equation (3) is not appropriate and leads to over-smoothing. Thus, we consider the non-zero gap ϵ.

Considering the optimal (ideal) derainer f_θ^* that is trained with clean images and the '2-loss, given x, it satisfies that $f * \theta(y) = x$ and $f_\theta^*(g`(y)) = g`(x), for` \in 1, 2$. Thus, the following holds with the optimal network $f^*\theta$:

$$E_{y|x}f^*\theta(g1(y)) - g2(y) - (g1(f * \theta(y)) - g_2(f * \theta(y)))(4) = g1(x) - Ey|xg2(y) - (g1(x) - g2(x)) = g2(x) - Ey|xg2(y) = \tag{4}$$

With the last two terms in Equation (4), we consider the gap between the ground truths of the training image pair. If the gap is zero, the subtraction of the last two terms in Equation (4) vanishes, and Equation (4) becomes a special case of rain2rain paired training in Equation (1). However, if the gap is non-zero, these two terms serve as a correction of the ground truth gap between the first two terms in Equation (4), forcing (4) to be zero. Therefore, Equation (4) provides a constraint that is satisfied when a derainer f_θ is the ideal one f_θ^* . To exploit this (ideal) constraint, rather than directly optimizing Equation (3), we consider the following constrained optimization problem:

$$min_\theta Ey|x\|f_\theta(g1(y)) - g2(y)\|_2^2, s.t. Ey|xf\theta(g1(y)) - g2(y) - g1(f_\theta(y)) + g2(f\theta(y)) = 0. \tag{5}$$

With the equation $Ex, y = E_x E_{y|x}$, we further reformulate it as the following regularized optimization problem:

$$min_\theta E_{x,y}\|f_\theta(g1(y)) - g_2(y)_2^2 + \gamma E_{x,y}\|f_\theta(g_1(y)) - g_2(y) - g_1(f_\theta(y)) + g_2(f_\theta(y))\|_2^2 \tag{6}$$

Technical Method and Experimentation

Based on previous works on unsupervised learning for image restoration we propose the following methodologies to address the problem of single image image restoration.

Sub-ampling Based Methods

After the successful application of Sub-sampling in image task like image denoising, Sub-sampling have increasingly gain popularity in image restoration tasks.

- Write about the population that you will study
- If it is an observational epidemiological study, then write about the exposure variable you will study. Hopefully, your background section will already have covered the prevalence of the exposure. If it is an intervention research, you will write about the intervention that you want to test.
- You will describe in details about the comparison group. If your study is one where you will be testing hypotheses, then it is important that you write about the comparison groups. You will write about the prevalence and how you will obtain measurements about the exposure and comparison groups.
- You will write about the outcomes in details, and specifically about how you will measure the exposure/intervention and the health outcome you want to study
- You will describe in details the power and sample size for this study. You can use the OpenEpi webpage to calculate your sample size and power for your study
- You will write in details about how you will eliminate bias in the measurement of the different variables in your study

- You will need to write, once you obtain data from your participants, how you would propose to analyse such data. You need not write too much details here, as you have not yet collected any data but an indicative set of statements as to what you will do should be sufficient.

These are the three compulsory sections that you will need to include in your proposal and then submit using Learn. If you use this template on Overleaf, then you can generate a PDF of your paper by selecting the PDF symbol on the top of this window. Save the PDF in your hard drive and then upload that one copy of PDF to Learn. If you use this template on Word, then convert the Word document to PDF and upload the document through Learn. Your document must contain images, tables, lists. All facts that you write must be accompanied by appropriate citation and referencing. The referencing information must be simple (just a number in square brackets), and an alphabetical order of the references in the bottom of the document should be sufficient. If you want to use APA style of referencing, that is OK too. For example, I have cited here a secondary analysis of data from papers published for about 40 years on statistical inference. It was an interesting paper written by Stang et.al. [2] and published in 2016. In my case the paper is cited in square brackets like this: `[1]`, and a full citation of the paper is mentioned in the references section. If you want to do the same but use APA 6th Edition citation, this is fine as well (and is used at the University here). But do not be discouraged to use your own style, as long as the citation information and the reference is there, that should be OK.

If you want to write using this template on Overleaf, I have written a tutorial that you can use to learn more about how to write on Overleaf. Or watch their several videos on the site to learn more about this tool to write your paper.

References

[1] Abderraouf Khodja, Zhonglong Zheng, Jiashuaizi Mo, Dawei Zhang, and Liyuan Chen. Rain to rain: Learning real rain removal without ground truth. *IEEE Access*, 9:57325–57337, 2021.

[2] Andreas Stang, Markus Deckert, Charles Poole, and Kenneth J Rothman. Statistical inference in abstracts of major medical and epidemiology journals 1975-2014: a systematic review. *European journal of epidemiology*, November 2016.